Distributionally robust views on queues and related stochastic models

van Eekelen, Wouter

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

CentER

Distributionally Robust Views on Queues and Related Stochastic Models

WOUTER VAN EEKELEN

TILBURG UNIVERSITY

Distributionally Robust Views on Queues and Related Stochastic Models

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof. dr. W.B.H.J. van de Donk, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de Aula van de Universiteit op

dinsdag 19 december 2023 om 16:00 uur

door

Wouter Johannes Eduard Cornelis van Eekelen,

geboren te Roosendaal en Nispen

PROMOTOR: prof. dr. Johan S.H. van Leeuwaarden (Tilburg University)

COPROMOTOR: prof. dr. ir. Dick den Hertog (University of Amsterdam)

PROMOTIECOMMISSIE: prof. dr. ir. Onno J. Boxma (Eindhoven University of Technology)
prof. dr. ir. Jan C. Fransoo (Tilburg University)
prof. dr. Bertrand Melenberg (Tilburg University)
dr. Yasemin Merzifonluoglu (Tilburg University)
dr. Ward Romeijnders (University of Groningen)

Distributionally Robust Views on Queues and Related Stochastic Models

Acknowledgments

The journey of writing this thesis began in September 2019, and reaching the end would have been much more difficult without the help and encouragement of many along the way.

First and foremost, I want to thank my supervisors, Johan van Leeuwaarden and Dick den Hertog. Johan, your guidance was invaluable throughout my time as a Ph.D. student, and it profoundly shaped my approach to research, writing, teaching and all other facets of academia. The numerous lessons you imparted to me shall also serve as guiding principles during the remainder of my academic career. Dick, our discussions were always insightful, and your endless positivity and enthusiasm never ceased to motivate me.

I would also like to extend my gratitude to John Hasenbein, for offering me the chance to visit the University of Texas at Austin and for our collaborative efforts that culminated in the third chapter of this thesis, and Guus Boonstra, whose outstanding Master's thesis project became the inspiration for the sixth chapter. My sincere thanks as well to all members of the committee for their careful reading of this thesis and their constructive feedback, which helped improve the final version you see before you.

My gratitude also goes out to the academic community to which I have belonged for the past few years, the Ph.D. students from the OR department in particular, for the welcome diversions during the workday. Finally, a big thanks to my friends and family whose support never wavered, despite their impression that I have a talent for making things sound more difficult needlessly—which I will not dispute, as it is probably true.

Wouter van Eekelen
Chicago, September, 2023

Contents

1
Introduction

1.1. Motivation

In this thesis, we discuss distribution-free perspectives on stochastic models for queues, inventory and finance. Traditionally, stochastic models rely on the assumption that the probability distributions of the driving random variables, such as service times, demand and stock prices, are fully known. In contrast, a distribution-free perspective assumes only partial knowledge of these distributions, for example only the mean and variance are known. Distributionally robust analysis seeks to determine the worst-case model performance by optimizing over the set of probability distributions that satisfy this partial information. For the stochastic models in this thesis, identifying this worst-case probability distribution requires solving semi-infinite optimization problems using duality theory.

We will first present a brief overview of related literature strands, in order to position this thesis. In Section 1.2, some concrete examples of the duality techniques are presented, which are closely connected with the theory of generalized moment problems. We will then discuss in Section 1.3 specific methodological challenges for stochastic models with multiple independent and identically distributed (i.i.d.) random variables. We will conclude this introduction in Section 1.4 with an outline of the chapters in this thesis.

1.1.1. Related literature

This thesis connects three common themes in the applied probability and optimization literature: Generalized moment problems, minimax stochastic programming, or distributionally robust optimization (DRO), and extremal queueing models.

Generalized moment problems. The study of moment problems dates back to the late 19th century and has attracted the attention of prominent probabilists of that era [43, 153, 205, 206]. In its most classical form, the problem of moments is essentially a feasibility problem that aims to determine if there exists a probability distribution that satisfies the given distributional information. Chebyshev [43] formulated the problem of determining bounds for tail probabilities in terms of only the mean and variance. A formal proof of the tight bound, now known as Chebyshev's inequality, was later provided by Markov [153]. In the 1950s and 1960s, there was a renewed interest in this type of problem, which gave rise to the vast literature on Chebyshev systems. We refer to the monograph of Karlin and Studden [128] for a comprehensive review. We further highlight the concurrent advent of duality theory as a favored method for solving these problems, as demonstrated in works such as [117, 128, 130]. Marshall and Olkin [154] were the first to generalize Chebyshev's inequality to the multivariate setting, and Mallows [150] used structural properties to sharpen the traditional Chebyshev inequalities. In more recent research, the connections between moment problems, nonnegativity of polynomials and semidefinite programming have been exploited (see, e.g., [29, 139, 172, 173, 177]).

Minimax stochastic programming and DRO. A natural progression from generalized moment problems is the application of this methodology to decision-making problems with uncertain parameters. Minimax stochastic programming involves defining a set of probability distributions to hedge against and then attempting to find the decision that provides the best protection against worst-case outcomes. Such a minimax approach has its roots in the duality theory for games, as introduced by Von Neumann. In his pioneering work, Scarf [191] was arguably the first to bring the concept of minimax optimization into the area of operations research. Scarf considered a single-item newsvendor problem where the demand distribution is not precisely known, but only characterized by limited information. The set of admissible distributions, also known as the ambiguity set, contained all distributions with specified mean and variance. Scarf used duality techniques to find the worst-case distribution that maximizes the total costs of the newsvendor, and then determined the order quantity such that these maximal total costs were minimized. Subsequently, generalized moment problems have been used to solve such minimax optimization problems in a vast amount of literature (see, for instance, [32, 34, 71, 197, 199, 232]). Another work that plays an important role in this thesis is the study conducted by Ben-Tal and Hochman [19], who derived robust bounds for the expected value of a convex function of random variables when in lieu of the variance, the mean absolute deviation from the mean (MAD) is used as dispersion measure. In recent years, the minimax paradigm regained traction in the operations research literature under the acronym DRO, or distributionally robust stochastic programming. The term DRO was first coined by [66], and has since then become the standard terminology in the operations research community. Distributionally robust optimization is advocated as the unifying paradigm between two distinct fields for decision-making under uncertainty: robust optimization and stochastic programming. Robust optimization [17, 22] provides an effective way to deal with problems subject to parameter uncertainty through uncertainty sets. Although encoding uncertainty through these uncertainty sets often results in computationally tractable problems, the resulting solutions might turn out to yield overly conservative outcomes. In contrast, stochastic programming captures parameter uncertainty with full distributional information, but often yields intractable models [198]. Ambiguity sets provide a powerful modeling tool for capturing distributional information about uncertain parameters in terms of their support and descriptive statistics. A significant portion of the literature on DRO is moment-based, with partial information given by means, moments, and dispersion measures [66, 101, 178, 224]. As adequately described in [224], the tractability of moment-based DRO problems relies on the intricate interplay between the objective function and the structure of the ambiguity set. Other popular ambiguity sets are based on statistical-distance measures, which restrict their members to be within a specific distance of a reference distribution. Examples of these statistical-distance measures include ϕ-divergences and the Wasserstein distance (see, e.g., [15, 157]). In the present work, we focus mainly on moment-based information. See [181] for overviews of many more DRO applications and techniques.

Extremal queue problem. Extremal stochastic models have been the subject of study for a considerable period, dating back to the early works of Hoeffding [114], Hoeffding and Shrikhande [115], Kingman [133]. At first, this body of literature primarily focused on simple univariate problems or sums of random variables. However, extremal stochastic models also encompass problems that involve finding bounds for performance metrics of stochastic processes, such as the waiting time in the GI/G/1 queue. One of the most renowned bounds in this context is the one proposed by Kingman [132], which bounds the expected waiting time when one only knows the means and variances of the service and interarrival times and is known to be tight in heavy traffic. Nevertheless, an important unresolved problem in queueing theory is to find the sharpest bound for the expected waiting time in the GI/G/1 queue with this mean-variance information. In search for the tight upper bound, foundational work was done by [73, 187], and by [222] in the context of the GI/M/1 queue. Whitt [222] considered the GI/M/1 queue with given mean and variance of the interarrival time, and showed that the expected steady-state waiting time is maximized when the interarrivals follow a specific two-point distribution. Daley conjectured (see [24]) that the overall worst-case behavior, in terms of expected waiting time, would be caused by two-point distributions for both the interarrival and the service time. That conjecture was proved invalid by counterexamples in [222] when fixing either the distribution of the interarrivals or the distribution of the services, but the conjecture remained standing for the case when both are unspecified, except for their first two moments.

Positioning of the thesis. This thesis unifies these strands of research in the following manner. We employ primarily the techniques presented in the generalized moment problem literature to derive new distribution-free bounds. These novel bounds will be valuable for analyzing distributionally robust stochastic programs, as well as for the extremal analysis of queueing models.

1.2. The weak duality technique

We now introduce the duality framework by means of obtaining the sharpest possible upper bound for the expectation $\mathbb{E}_{\mathbb{P}}[g(X)]$ given limited information on the distribution $\mathbb{P}$ of the random variable X. We model this through the generalized moment problem

$$\max_{\mathbb{P}\in\mathcal{M}_+(\Omega)} \quad \mathbb{E}_{\mathbb{P}}[g(X)] \quad \text{s.t.} \quad \mathbb{E}_{\mathbb{P}}[h_j(X)] = q_j \text{ for } j = 0, \dots, m, \tag{1.1}$$

where $\mathbb{P}$ belongs to the set of nonnegative measures $\mathcal{M}_+(\Omega)$ with support confined to Ω, and $g, h_0, \dots, h_m$ are measurable, real-valued functions with $h_0 \equiv 1$ and $q_0 = 1$. Hereafter, throughout the entire manuscript, we define measurable functions as exclusively real-valued ones with respect to the appropriate Borel algebra. The functions $h_0, \dots, h_m$ describe the available distributional information, and the expectations of these moment functions, $q_0, \dots, q_m$, are assumed to be known and finite. There are typically two common approaches to solving problem (1.1). For the first approach, we need a concept that is analogous to basic solutions in standard linear programming. The Richter-Rogosinski theorem (see, e.g., [186, Theorem 1] and [198, Lemma 3.1]) asserts the existence of an extremal distribution for (1.1) with at most $m + 1$ support points. Leveraging this general theory, one can reduce the semi-infinite linear program (1.1) to a finite-dimensional optimization problem. This method is widely used for addressing problems in the uncertainty quantification literature, as discussed in [98, 170]. Nevertheless, the resulting optimization problems may be nonconvex, making them potentially challenging to solve.

The second approach involves solving the Lagrangian dual of (1.1),

$$\min_{\lambda_0,\dots,\lambda_m} \sum_{j=0}^{m} \lambda_j q_j \quad \text{s.t.} \quad \sum_{j=0}^{m} \lambda_j h_j(x) \geqslant g(x), \quad \forall x \in \Omega, \tag{1.2}$$

which can be used in lieu of the primal problem (1.1). It is worth emphasizing that in (1.2), one optimizes over a set of dual functions characterized by the weighted sum of the moment functions $h_0, \dots, h_m$, whereas in (1.1), one optimizes over probability measures. However, this approach can be somewhat strenuous, as we are still dealing with an infinite number of constraints. Therefore, we will take a different route.

We propose the notion of weak duality as a tool for solving generalized moment problems, as discussed in, for example, [102, Chapter 2] and [191]. It turns out to be advantageous to consider (1.1) directly in conjunction with its dual problem (1.2). By making an educated guess for the dual solution, we can construct a corresponding primal solution. We can then use weak duality to confirm the optimality of these candidate solutions.

We shall first demonstrate this versatile solution strategy for generalized moment problems with generic objective functions. To illustrate the approach, we will use the probably most well-known example from DRO. Then we will demonstrate that there exist combinations of function classes and distributional information for which we can simplify the search for optimal solutions even further. This ultimately leads to optimal distributions that depend exclusively on the distributional information, irrespective of the functional form of g. We shall refer to this property as the "insensitivity" property.

1.2.1. Concerning particular solutions

In his seminal work, Scarf [191] solved the distributionally robust stochastic program

$$\min_{q\in\mathbb{R}_+} \max_{\mathbb{P}\in\mathscr{P}_{(\mu,\sigma)}} \mathbb{E}_{\mathbb{P}}[C(q,X)]. \tag{1.3}$$

Here the ambiguity set $\mathscr{P}_{(\mu,\sigma)}$ is defined as the set of all distributions with given mean μ and variance σ^2, and $C(q,x)$ denotes the total cost incurred by the newsvendor when ordering q units to meet a demand of X. Scarf identified the worst-case distribution that yields the maximal costs for a given order quantity, and then determined the order quantity that minimized the maximal total cost. Particularly pertinent to our interests is the inner maximization stage. Scarf determined this "worst-case" through the use of weak duality, which involves solving the problem

$$\max_{\mathbb{P}\in\mathscr{P}_{(\mu,\sigma)}} \mathbb{E}_{\mathbb{P}}[(X-q)^+], \tag{1.4}$$

where $(x)^+ = \max\{x,0\}$. Scarf essentially sought, given the order quantity decision, the demand distribution that maximized the expected excess demand. This problem is tantamount to solving the semi-infinite linear program (LP)

$$\begin{aligned} &\max_{\mathbb{P}} \int_x (x-q)^+ \, \mathrm{d}\mathbb{P}(x) \\ &\text{s.t.} \int_x \mathrm{d}\mathbb{P}(x) = 1, \int_x x \, \mathrm{d}\mathbb{P}(x) = \mu, \int_x x^2 \, \mathrm{d}\mathbb{P}(x) = \mu^2+\sigma^2, \end{aligned} \tag{1.5}$$

which has a finite number of moment constraints (three in this case), and an infinite number of decision variables, modeled through the probability distribution $\mathbb{P}$ (an infinite-dimensional object essentially). Just as for linear programming, a dual problem can be set up using standard techniques. The dual problem of (1.5) is yet another semi-infinite LP, but this time with an infinite number of constraints, and it can be written as

$$\begin{aligned} &\min_{\lambda_0,\lambda_1,\lambda_2} \lambda_0 + \lambda_1\mu + \lambda_2(\mu^2+\sigma^2) \\ &\text{s.t. } M(x) := \lambda_0 + \lambda_1 x + \lambda_2 x^2 \geqslant (x-q)^+, \ \forall x. \end{aligned} \tag{1.6}$$

It is not hard to see that problem (1.6) is weakly dual to (1.5), i.e., the optimal value of (1.6) provides an upper bound for (1.5). This is easily shown as follows. Suppose that $\mathbb{P}$ is feasible for the primal and $(\lambda_0,\lambda_1,\lambda_2)$ is feasible for the dual. Then, by the constraints of (1.6),

$$\int_x (x-q)^+ \mathrm{d}\mathbb{P}(x) \leqslant \int_x M(x)\mathrm{d}\mathbb{P}(x) = \lambda_0 + \lambda_1\mu + \lambda_2(\mu^2+\sigma^2),$$

so that (1.6) is weakly dual to (1.5). As such, for every feasible $\mathbb{P}$ and $(\lambda_0,\lambda_1,\lambda_2)$, (1.6) provides an upper bound for the objective value of (1.5). Hence,

$$\max_{\mathbb{P}} \int_x (x-q)^+ \mathrm{d}\mathbb{P}(x) \leqslant \min_{\lambda_0,\lambda_1,\lambda_2} \lambda_0 + \lambda_1\mu + \lambda_2(\mu^2+\sigma^2).$$

Now, if one finds a feasible distribution $\mathbb{P}^*$ and a feasible dual vector $(\lambda_0^*, \lambda_1^*, \lambda_2^*)$ such that

$$\int_x (x-q)^+ \mathrm{d}\mathbb{P}^*(x) = \lambda_0^* + \lambda_1^* \mu + \lambda_2^*(\mu^2 + \sigma^2),$$

then these solutions are optimal for the primal problem (1.5) and the dual problem (1.6); moreover, strong duality holds. Strong duality can be established directly under mild conditions on the moment vector $\mathbf{q} = (q_1, \ldots q_m)$ (see, for example, [117] and [177, Theorem 2.1]). However, we will not simply assume this; rather, we will seek optimal primal-dual solutions to demonstrate strong duality in a more constructive manner. To this end, recall the complementary slackness property from linear programming. This result readily extends to semi-infinite LPs. That is, for $\mathbb{P}^*$ and $(\lambda_0^*, \lambda_1^*, \lambda_2^*)$ the optimal solutions to the primal and dual problems respectively,

$$\int_x (M^* - g)\mathrm{d}\mathbb{P}^*(x) = 0.$$

This semi-infinite programming variation of complementary slackness (see, e.g., [201, p. 813]) implies that a sufficient condition for a feasible primal-dual solution pair to be optimal is that the extremal distribution is supported on the points where the optimal dual function $M^*(x)$ coincides with $g(x)$. It is thus sufficient to construct a feasible dual function M^* and check its optimality by constructing a feasible distribution with all of its probability mass on the points where $M^*(x) = g(x)$.

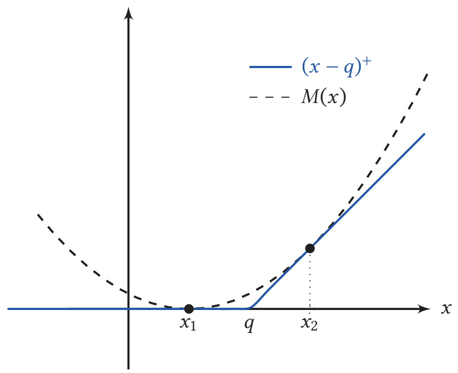


Figure 1.1: Primal-dual solutions to Scarf's distributionally robust stochastic program

These insights combined provide a constructive method for solving a semi-infinite LP, like Scarf's or in general. The solution of the dual problem not only gives an upper bound for the primal problem (by weak duality), but also identifies a candidate worst-case distribution. Indeed, in this way Scarf proved (1.5) is solved by $\lambda_0, \lambda_1, \lambda_2$ yielding a parabola $M(x)$, as in Figure 1.1, that touches $(x-q)^+$ at two points. Since $M(x)$ is convex, it touches $(x-q)^+$ at the points x_1, x_2 where $M(x_1) = 0$, $M(x_2) = (x_2 - q)$, and

$$M'(x_1) = 0 \iff \lambda_1 + 2\lambda_2 x_1 = 0 \implies x_1 = -\frac{\lambda_1}{2\lambda_2}.$$
$$M'(x_2) = 1 \iff \lambda_1 + 2\lambda_2 x_2 = 1 \implies x_2 = \frac{1-\lambda_1}{2\lambda_2}.$$

Substituting $x_2 = \frac{1-\lambda_1}{2\lambda_2}$ into $M(x_2) = (x_2 - q)$ and $x_1 = -\frac{\lambda_1}{2\lambda_2}$ into $M(x_1) = 0$ gives

$$\lambda_0 = \frac{(1-4q\lambda_2)^2}{16\lambda_2}, \quad \lambda_1 = \frac{1}{2} - 2q\lambda_2.$$

Inserting these values into the dual objective function leads us to consider

$$\min_{\lambda_2} \frac{(1-4q\lambda_2)^2}{16\lambda_2} + \left(\frac{1}{2} - 2q\lambda_2\right)\mu + \lambda_2(\mu^2 + \sigma^2).$$

Taking the derivative with respect to λ_2 and setting it equal to zero, we find that

$$\lambda_2^* = \frac{1}{4\sqrt{(\mu-q)^2 + \sigma^2}}. \tag{1.7}$$

Thus, the dual objective value equals

$$\frac{(1-4q\lambda_2^*)^2}{16\lambda_2^*} + \left(\frac{1}{2} - 2q\lambda_2^*\right)\mu + \lambda_2^*(\mu^2 + \sigma^2) = \frac{1}{2}(\mu - q + \sqrt{(\mu-q)^2 + \sigma^2}).$$

Using the proposed support points

$$x_1 = q - \frac{1}{4\lambda_2^*} = q - \sqrt{(\mu-q)^2 + \sigma^2}, \quad x_2 = q + \frac{1}{4\lambda_2^*} = q + \sqrt{(\mu-q)^2 + \sigma^2},$$

we return to the primal problem and solve

$$p_1 + p_2 = 1, \quad p_1 x_1 + p_2 x_2 = \mu, \quad p_1 x_1^2 + p_2 x_2^2 = \mu^2 + \sigma^2. \tag{1.8}$$

Hence,

$$p_1 = \frac{1}{2}\left(1 - \frac{\mu - q}{\sqrt{(\mu-q)^2 + \sigma^2}}\right), \; p_2 = \frac{1}{2}\left(1 + \frac{\mu - q}{\sqrt{(\mu-q)^2 + \sigma^2}}\right)$$

with primal objective value

$$\int_x (x-q)^+ \mathrm{d}\mathbb{P}(x) = (x_2 - q)p_2 = \frac{1}{2}(\mu - q + \sqrt{(\mu-q)^2 + \sigma^2}).$$

Thus, we observe that the objective values of the primal and dual feasible solutions agree. By weak duality, we conclude that this primal-dual solution pair is optimal for the semi-infinite LP (1.5) and its dual problem (1.6). In the end, this yields the tight upper bound

$$\max_{\mathbb{P} \in \mathcal{P}_{(\mu,\sigma)}} \mathbb{E}[(X-q)^+] = \frac{1}{2}\left(\mu - q + \sqrt{(\mu-q)^2 + \sigma^2}\right). \tag{1.9}$$

It is worth noting that the precise form of the objective function plays a crucial role in identifying the optimal solution to $\max_{\mathbb{P}\in\mathscr{P}_{(\mu,\sigma)}} \mathbb{E}[(X-q)^+]$. So, even though a general method of solution (i.e., weak duality) exists, determining the optimal solution vector $(\lambda_0^*, \lambda_1^*, \lambda_2^*)$ typically remains an *ad hoc* procedure. Thus, finding the optimal primal-dual solution pair can still be an arduous task. However, as we will discuss next, specific conditions on the shape of the function g can considerably simplify the search for the optimal solution.

1.2.2. Concerning insensitive solutions

As in the previous subsection, we strive to find the extremal distribution for X that gives the worst-case expectation for the function g. However, here it is assumed that besides being measurable and finite valued, $g(\cdot)$ is a convex function of X. To describe all considered distributions, we define an ambiguity set that consists of all distributions that comply with the limited information available. The partial information consists of X having a known bounded support, $\text{supp}(X) \subseteq \Omega := [a,b]$ with $-\infty < a \leqslant b < \infty$, mean $\mathbb{E}_{\mathbb{P}}[X] = \mu$ and MAD $\mathbb{E}_{\mathbb{P}}[|X-\mu|] = d$. This defines the ambiguity set

$$\mathscr{P}_{(\mu,d)} = \{\mathbb{P} : \ \text{supp}(X) \subseteq [a,b],\ \mathbb{E}_{\mathbb{P}}[X] = \mu,\ \mathbb{E}_{\mathbb{P}}[|X-\mu|] = d\}. \tag{1.10}$$

In what follows, X is a random variable whose distribution belongs to the set $\mathscr{P}_{(\mu,d)}$. The extremal distribution that solves the resulting moment problem is a three-point distribution with support $\{a, \mu, b\}$ and respective probabilities [19]

$$p_1 = \frac{d}{2(\mu-a)}, \quad p_2 = 1 - \frac{d}{2(\mu-a)} - \frac{d}{2(b-\mu)}, \quad p_3 = \frac{d}{2(b-\mu)}. \tag{1.11}$$

Observe that this extremal distribution does not in any way depend on the objective function g. It is quite insightful to see why the moment problem (1.1), with a convex objective function and MAD as the dispersion measure, returns such a simple worst-case distribution. Under mean-MAD ambiguity of one random variable X, (1.1) specializes to the semi-infinite LP

$$\begin{aligned} &\max_{\mathbb{P}} \int_{\Omega} g(x)\,\mathrm{d}\mathbb{P}(x) \\ &\text{s.t.} \int_{\Omega} \mathrm{d}\mathbb{P}(x) = 1, \ \int_{\Omega} x\,\mathrm{d}\mathbb{P}(x) = \mu, \ \int_{\Omega} |x-\mu|\,\mathrm{d}\mathbb{P}(x) = d, \end{aligned} \tag{1.12}$$

to which the problem

$$\begin{aligned} \min_{\lambda_0,\lambda_1,\lambda_2} \quad & \lambda_0 + \lambda_1\mu + \lambda_2 d \\ \text{s.t.} \quad & U(x) := \lambda_0 + \lambda_1 x + \lambda_2|x-\mu| \geqslant g(x),\ \forall x \in [a,b] \end{aligned} \tag{1.13}$$

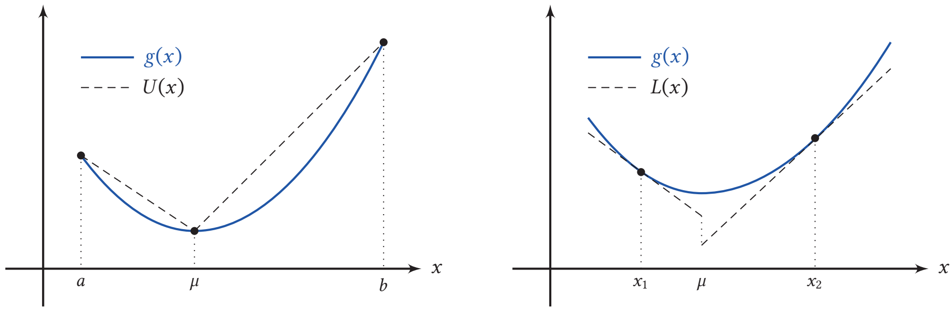
is weakly dual. The function $U(x)$ has a "kink" at $x = \mu$. Since the majorant is piecewise linear and convex, every convex function $g(x)$ is majorized by letting $U(x)$ touch at the boundary points a, b and at the kink point μ, as displayed in Figure 1.2a. For this choice of $U(x)$, the dual variables are

$$\lambda_0 = f(a) - \lambda_1 a - \lambda_2(\mu - a),\ \lambda_1 = \frac{1}{2}\left(\frac{f(b)-f(\mu)}{b-\mu} + \frac{f(\mu)-f(a)}{\mu-a}\right),$$
$$\lambda_2 = \frac{1}{2}\left(\frac{f(b)-f(\mu)}{b-\mu} - \frac{f(\mu)-f(a)}{\mu-a}\right).$$

By the complementary slackness property, we expect the three points at which $g(x)$ and $U(x)$ coincide to constitute the support of the extremal distribution. From the constraints of (1.12), we obtain a linear system of three unknown probabilities and three equations:

$$p_1 + p_2 + p_3 = 1, \quad p_1 a + p_2 \mu + p_3 b = \mu, \quad p_1|a-\mu| + p_3|b-\mu| = d,$$

yielding (1.11). After substituting this primal-dual solution pair in (1.12) and (1.13), it immediately follows that strong duality holds. As a result, this particular category of functions (i.e., convex functions), in conjunction with a specific type of information, renders the extremal distribution "insensitive" to the exact form of the objective function. That is, it only depends on the distributional information specified.



(a) The majorant $U(x)$ **(b)** The minorant $L(x)$

Figure 1.2: Some convex function $g(x)$ and the dual functions for mean-MAD(-β) information

These tight upper bounds correspond to worst-case scenarios. As a second example, we show how MAD information can be used to determine best-case scenarios and hence tight lower bounds. Define a second ambiguity set, which is a subset of $\mathscr{P}_{(\mu,d)}$:

$$\mathscr{P}_{(\mu,d,\beta)} = \left\{\mathbb{P} \ :\ \mathbb{P} \in \mathscr{P}_{(\mu,d)},\ \mathbb{P}(X \geqslant \mu) = \beta\right\}. \tag{1.14}$$

Here, some information regarding the skewness of the distribution, $\mathbb{P}(X \geqslant \mu) = \beta$, is added to the ambiguity set. Ben-Tal and Hochman [19] proved a general lower bound for the expectation of a convex function of a random variable with $\mathbb{P} \in \mathscr{P}_{(\mu,d,\beta)}$.

For the tight lower bound, it can be shown that the best-case distribution is a two-point distribution which is insensitive to the precise functional form of g, except that it must be convex. For $X \sim \mathbb{P}$ with $\mathbb{P} \in \mathscr{P}_{(\mu,d,\beta)}$, the tight lower bound follows from solving

$$\begin{aligned} \min_{\mathbb{P}} \quad & \int_{\Omega} g(x)\mathrm{d}\mathbb{P}(x) \\ \text{s.t.} \quad & \int_{\Omega} \mathrm{d}\mathbb{P}(x) = 1,\ \int_{\Omega} x\mathrm{d}\mathbb{P}(x) = \mu,\ \int_{\Omega} |x-\mu|\mathrm{d}\mathbb{P}(x) = d,\ \int_{\Omega} {}_{\{x \geqslant \mu\}}(x)\mathrm{d}\mathbb{P}(x) = \beta, \end{aligned} \tag{1.15}$$

which is a semi-infinite LP with four equality constraints.

Consider the dual of (1.15),

$$\begin{aligned} \max_{\lambda_0,\lambda_1,\lambda_2,\lambda_3} \quad & \lambda_0 + \lambda_1\mu + \lambda_2 d + \lambda_3\beta \\ \text{s.t.} \quad & g(x) \geqslant \lambda_0 + \lambda_1 x + \lambda_2|x-\mu| + \lambda_{3\{x\geqslant\mu\}}(x),\ \forall x \in \Omega. \end{aligned} \tag{1.16}$$

Define $L(x) := \lambda_0 + \lambda_1 x + \lambda_2|x-\mu| + \lambda_{3\{x\geqslant\mu\}}$. Then the inequality in (1.16) can be written as $g(x) \geqslant L(x)$, $\forall x$, i.e. $L(x)$ minorizes $g(x)$. In this case, $L(x)$ has both a "kink" and a jump discontinuity at $x = \mu$, as depicted in Figure 1.2b.

Thus, the goal is to find the tightest minorant that maximizes the objective value of the dual problem. By the supporting hyperplane theorem, $L(x)$ touches the epigraph of $g(x)$ at two points on opposite sides of μ (i.e., at $x_1 \leqslant \mu \leqslant x_2$). Using this insight, it follows from solving the moment conditions

$$p_1 + p_2 = 1, \quad p_1x_1 + p_2x_2 = \mu, \quad p_1|x_1-\mu| + p_2|x_2-\mu| = d, \quad p_2 = \beta,$$

that

$$x_1 = \mu - \frac{d}{2(1-\beta)}, \quad x_2 = \mu + \frac{d}{2\beta}$$

and $\mathbb{P}(X = x_1) = 1-\beta$ and $\mathbb{P}(X = x_2) = \beta$. Solving for $\lambda_0, \lambda_1, \lambda_2$ and λ_3, such that we obtain $L(x)$, yields

$$\lambda_0 = g(x_1) + \frac{(\lambda_1-\lambda_2)d}{2(1-\beta)} - \lambda_1\mu, \quad \lambda_3 = g(x_2) - g(x_1) + \frac{\lambda_2 d}{(1-\beta)} - \frac{(\lambda_1+\lambda_2)d}{2\beta(1-\beta)}.$$

To ensure the solution is dual feasible, the free dual variables λ_1, λ_2 are chosen such that $\lambda_2 - \lambda_1$ and $\lambda_2 + \lambda_1$ equal the slope of $g(x)$ at $x = x_1$ and x_2, respectively. It is then easily verified that the optimal value of (1.15) and (1.16) is given by $(1-\beta)g(x_1) + \beta g(x_2)$, attained by the two-point distribution with support $\{x_1, x_2\}$ and respective probabilities $\{(1-\beta), \beta\}$. This extremal distribution is again insensitive to the objective function. This insensitivity will prove to be rather helpful in the distribution-free analysis of stochastic models.

1.3. The multivariate problem with i.i.d. random variables

Now consider $g_n(\mathbf{X})$ as function of the random vector $\mathbf{X} = (X_1, \dots, X_n)$ with support $\Omega \subseteq \mathbb{R}^n$. This function might reflect the realization of a stochastic process or some cost in a decision problem, for which the vector $\mathbf{X}$ then represents the driving sequence of the stochastic process or the uncertain parameters in the decision problem. We then seek a solution to the problem

$$\max_{\mathbb{P}\in\mathscr{P}} \mathbb{E}_{\mathbb{P}}[g_n(\mathbf{X})] \tag{1.17}$$

with $\mathbb{P}$ some joint distribution from the set $\mathscr{P}$ of distributions compatible with the partial information. The partial information available here pertains to the joint measure and might contain information about both marginal and cross moments, among other types of information. It turns out most of the duality techniques discussed in the previous section still apply to the multivariate setting, albeit under some additional assumptions (see, e.g., [117, 196, 201]).

We will first introduce a particular instance of (1.17), and then we discuss how these multivariate problems are typically solved using the duality theory of generalized moment problems. Finally, we will elaborate on why this becomes significantly more complicated for the setting with i.i.d. random variables.

1.3.1. The extremal queue problem

When i.i.d. random variables are involved, problem (1.17) becomes substantially more difficult to solve. The following function is a classic example in applied probability that brings this difficult i.i.d. setting:

$$f_n(x_1, \dots, x_n) = \max\{0, x_1, \dots, x_1 + \dots + x_n\},$$

for which $\mathbb{E}_{\mathbb{P}}[f_n(\mathbf{X})]$ expresses the expected running maximum of a random walk $S_n := X_1 + \dots + X_n$ ($S_0 := 0$) with the i.i.d. driving sequence $\{X_n, n \geqslant 1\}$ distributed as X. The random walk $\{S_n, n \geqslant 0\}$ and its running maximum $M_n := \max\{S_0, S_1, \dots, S_n\}$ are extensively studied in standard texts on probability theory [7, 56, 57, 79] with methods that require full distributional information of $X_1, \dots, X_n$. However, in this thesis, we concentrate on the setting in which the distribution is only partially known. Denote by W_n the waiting time of customer n in the GI/G/1 queue, where $W_0 = 0$. Let $X_n = V_n - U_n$ be the difference between service and interarrival time. The waiting-time process satisfies Lindley's recursion

$$W_{n+1} = (W_n + X_n)^+,\ n \geqslant 0,$$

which, by the i.i.d. assumption, is equivalent to

$$W_{n+1} \stackrel{d}{=} \max\{0, X_1, \dots, X_1 + X_n\} = \max\{S_0, S_1, \dots, S_n\} = M_n,$$

where $\stackrel{d}{=}$ denotes equality in distribution. The sequence $\{W_n, n \geqslant 0\}$ is thus generated by the random walk $\{S_n, n \geqslant 0\}$, through its running maximum $\{M_n, n \geqslant 0\}$. Then, if only the first two moments of the interarrival and service times are known, it is of interest to solve the problem

$$\max_{\mathbb{P} \in \mathscr{P}_{V,U}} \mathbb{E}[f_n(\mathbf{X})]$$

where $\mathscr{P}_{V,U}$ is based on the available moment information and $\mathbf{X}$ is a random vector with elements $\{V_i, U_i\}_{i=1}^n$. As $n \to \infty$, this yields the extremal queue problem, one of the longest-standing open problems in queueing theory [50, 60, 222], which aims to determine tight bounds for the moments of the stationary GI/G/1 waiting time $W := \lim_{n\to\infty} W_n$.

1.3.2. Strong duality for generalized moment problems

The concept of duality theory to solve generalized moment problems originated in the 1960s, most notably in the works of [117, 118, 128, 130, 133]. Smith [201] provided a modern account of this duality theory with applications in operations research and decision theory, while Shapiro [196] discussed more rigorously the necessary topological conditions for strong duality, relating the duality results for generalized moment problems to the more general theory on conic duality. More recently, a stream of research exploits the connections between moment problems, nonnegativity of polynomials and semidefinite programming (see, e.g., [29, 139]). In his now-classical work, Stieltjes [205, 206] introduced the problem of moments in its most basic form. It concerns itself with identifying whether the set

$$\mathscr{P}(\mathbf{q}) := \left\{ \mathbb{P} \in \mathscr{M}_+(\Omega) : \int h_j(x)\mathrm{d}\mathbb{P}(x) = q_j,\, j = 0, \dots, m \right\}, \tag{1.18}$$

contains any elements, and it is thus recognized as a feasibility problem. As a closely related concept that we will leverage, denote with $\mathscr{C}$ the cone of moments $\mathbf{q} \in \mathbb{R}^m$ that yield a nonempty ambiguity set $\mathscr{P}(\mathbf{q})$, i.e.,

$$\mathscr{C} := \{\mathbf{q} \in \mathbb{R}^m : \exists \mathbb{P} \in \mathscr{M}_+(\Omega) \text{ such that } \mathbb{P} \in \mathscr{P}(\mathbf{q})\}.$$

This set identifies the moment vectors for which the moment problem admits a solution.

We now turn to the generalized moment problem, in which we aim to find the tight upper bounds on some objective $\mathbb{E}[g_n(\mathbf{X})]$, transforming the feasibility problem into an optimization problem. Like (1.1), problem (1.17) can be stated as a semi-infinite linear programming problem,

$$\max_{\mathbb{P} \in \mathscr{M}_+(\Omega)} \int_\Omega g_n(\mathbf{x})\mathrm{d}\mathbb{P}(\mathbf{x}) \quad \text{s.t.} \quad \int_\Omega h_j(\mathbf{x})\mathrm{d}\mathbb{P}(\mathbf{x}) = q_j \text{ for } j = 0, \dots, m, \tag{1.19}$$

but now considering the joint probability distribution of the components of $\mathbf{X}$, defined on the support Ω, instead of the probability distribution of a single random variable X. The dual problem of (1.19) can be expressed as

$$\min_{\lambda_0, \dots, \lambda_m} \sum_{j=0}^m \lambda_j q_j \quad \text{s.t.} \quad \sum_{j=0}^m \lambda_j h_j(\mathbf{x}) \geqslant g(\mathbf{x}), \quad \forall \mathbf{x} \in \Omega, \tag{1.20}$$

which is similar to (1.2), but involves multivariate functions. The duality results discussed in Section 1.2 can be readily extended to the multivariate setting (see, e.g., [117, 133]). To achieve this, we only need the additional assumption that the moment vector $\mathbf{q}$ lies in the interior of the moment cone $\mathscr{C}$. If this is the case, strong duality is ensured to hold. However, even though solving the dual is sufficient, it involves checking nonnegativity of a multivariate function, which can be a daunting task in general. When we only consider marginal moment information, the dual problem usually admits a computationally more tractable reformulation at the cost of disregarding the dependency structure between the components of $\mathbf{X}$. On the other hand, cross-moment ambiguity sets incorporate this information, resulting in tighter bounds but often leading to hard optimization problems [25, 163, 165]. The computational tractability of the dual problem is a widely studied problem in the DRO literature; see, for example, [29, 66, 100, 224].

1.3.3. Computational intractability of i.i.d. driving sequences

Consider a stochastic model with independent random variables $\{X_i\}_{i=1}^n$ of which the distributions $\{\mathbb{P}_i\}_{i=1}^n$ belong to some ambiguity set $\{\mathcal{P}_i\}_{i=1}^n$. In the previous subsection, we considered the joint probability distribution $\mathbb{P}$ defined on the support of $\mathbf{X}$. Instead, we shall now optimize over the (constrained) joint measure $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2 \otimes \cdots \otimes \mathbb{P}_n$, where the marginal distributions $\{\mathbb{P}_i\}_{i=1}^n$ all lie in their respective ambiguity sets, and $\otimes$ denotes the product measure operator. The objective function then takes the form

$$\int_{x_n} \cdots \int_{x_2} \int_{x_1} g_n(x_1, \ldots, x_n)\, \mathrm{d}\mathbb{P}_1(x_1) \cdot \mathrm{d}\mathbb{P}_2(x_2) \cdots \mathrm{d}\mathbb{P}_n(x_n), \tag{1.21}$$

which is no longer linear, nor concave, with respect to the optimization variables (i.e., the probability measures). As a consequence, (1.21) does not yield a convex optimization problem, and therefore, strong duality might not hold. Yet another way to look at this is as follows: Independence is a structural property imposed on all (joint) probability distributions in the set $\mathcal{P}$. As an example, denote by $\mathcal{P}_0$ the set of all probability distributions with support $\Omega = \{0, 1\}^2$. We thus consider two binary-valued random variables X, Y. Define $\delta_{(x,y)}$ as the Dirac measure which puts a probability mass of one on the support point (x, y). The set $\mathcal{P}_0$ can be characterized as the closure (with respect to the weak topology) of the convex hull of its extreme points, which are the Dirac measures, and is thus a convex set. However, notice that all mixture distributions (convex combinations) of the two extreme points $\delta_{(0,0)}$ and $\delta_{(1,1)}$ have perfectly correlated components. As a result, these convex combinations cannot be elements of the ambiguity set that has the independence property. The structural property of independence degrades the structure of the set of probability distributions $\mathcal{P}_0$, transforming it into a nonconvex set that is less ideal to work with since it makes finding a global optimum considerably more difficult. To alleviate these difficulties, one could of course relax—or, in a sense, convexify—problem (1.21) by using (1.17) instead, but this still allows (higher-order) correlations between the random variables, even when we impose the cross-moment constraints $\mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] = 0,\ \forall i \neq j$.

By means of an example, we will explain why optimization problem (1.21) is challenging, as already pointed out in the works by [114] and [133]. There do exist methods to solve (1.21) under the assumption of independent (not necessarily identical) random variables, although they can become computationally cumbersome. To illustrate this point, we will use the random walk example described at the beginning of this section. Formally, finding the tight bound for the running maximum of a random walk can be achieved by solving the following optimization problem over probability distributions:

$$\begin{aligned} \max_{\mathbb{P}, \{\mathbb{P}_i\}_{i=1}^n}\ & \mathbb{E}_{\mathbb{P}}\left[\max\{0, X_1, \ldots, X_1 + \cdots + X_n\}\right] \\ \text{s.t. } & \mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2 \otimes \cdots \otimes \mathbb{P}_n, \\ & \mathbb{E}_{\mathbb{P}_i}[h_j(X_i)] = q_{i,j}, \quad i = 1, 2, \ldots, n,\ \ j = 1, \ldots, m. \end{aligned} \tag{1.22}$$

The measure $\mathbb{P}$, defined on $\mathbb{R}^n$, is the joint probability measure of the random vector $\mathbf{X}$, which is constructed from the individual probability distributions $\{\mathbb{P}_i\}_{i=1}^n$ through the product measure operator. By writing $\mathbb{P}$ in this particular form, we impose independence among the random variables in $\mathbf{X}$. For the sake of exposition, let us concentrate on the setting with mean-variance information constraints. Problem (1.22) then becomes

$$\begin{aligned} \max_{\mathbb{P},\{\mathbb{P}_i\}_{i=1}^n} \ & \mathbb{E}_{\mathbb{P}}\left[\max\{0, X_1, \dots, X_1 + \cdots + X_n\}\right] \\ \text{s.t. } & \mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2 \otimes \cdots \otimes \mathbb{P}_n, \\ & \mathbb{E}_{\mathbb{P}_i}[X_i] = \mu_i, \quad i = 1, 2, \dots, n, \\ & \mathbb{E}_{\mathbb{P}_i}[X_i^2] = \mu_i^2 + \sigma_i^2, \quad i = 1, 2, \dots, n. \end{aligned} \tag{1.23}$$

Note that the objective function is linear in the joint probability distribution $\mathbb{P}$, and the moment constraints are linear in the marginal distributions $\{\mathbb{P}_i\}_{i=1}^n$. If we would ignore the independence constraint, (1.23) becomes a marginal-moment problem with mean-variance constraints for each marginal distribution $\mathbb{P}_i$. It is worth recalling that such infinite-dimensional problems can be reduced to finite-dimensional ones, of which the optimal objective values are preserved, since the optimal distribution is always achieved by discrete point distributions. In fact, from the independence constraint, it follows that every marginal distribution is optimized by a discrete distribution $\mathbb{P}_i^*$ with support on at most $m + 1$ points, where m is the number of moment constraints [170]. In other words, the Richter-Rogosinski theorem still holds even if we impose independence. For an elegant proof using Fubini's theorem, we refer the interested reader to the work of Kingman [133, Theorem 3], who also provides an excellent discussion on some of the topics discussed in this introduction.

Since our example involves two information constraints (mean and variance), each $\mathbb{P}_i^*$ consists of at most three point masses. We denote the support and probabilities for each X_i as $(x_1^{(i)}, x_2^{(i)}, x_3^{(i)})$ and $(p_1^{(i)}, p_2^{(i)}, p_3^{(i)})$, respectively. After applying the Richter-Rogosinski theorem, (1.23) can be reduced to the following optimization problem:

$$\begin{aligned} \max_{\mathbf{p},\mathbf{x}} \ & \sum_{\boldsymbol{\alpha}\in\{1,2,3\}^n} \left(\prod_{i=1}^n p_{\alpha_i}^{(i)}\right) \max\{x_{\alpha_1}^{(1)}, \dots, x_{\alpha_1}^{(1)} + \cdots x_{\alpha_n}^{(n)}\} \\ \text{s.t. } & \sum_{j=1}^3 p_j^{(i)} x_j^{(i)} = \mu_i, \quad i = 1, 2, \dots, n, \\ & \sum_{j=1}^3 p_j^{(i)} \left(x_j^{(i)}\right)^2 = \mu_i^2 + \sigma_i^2, \quad i = 1, 2, \dots, n, \\ & 0 \leqslant \mathbf{p} \leqslant 1, \ \mathbf{x} \in \Omega. \end{aligned} \tag{1.24}$$

Clearly, this problem can be computationally challenging due to the decision variables being both the support points and the probabilities. While numerical optimization can be used to determine the extremal distribution, it may not provide the necessary insights into the structural properties of the extremal distribution that are needed to solve the problem for arbitrary n.

The example presented above does not assume that the independent variables $\{X_i\}_{i=1}^n$ are identically distributed. Imposing identical distributions would lead to losing the ability to reduce the semi-infinite LP to its finite-dimensional counterpart. To overcome this limitation, one could replace the identicality constraint with the constraint of identical moments (up to a certain order), as proposed in [27]. Unfortunately, an exact analysis of the i.i.d. setting resulting in explicit solutions remains elusive. Although a general framework for distribution-free analysis under the i.i.d. assumption is still desired, this thesis does offer some results that leverage a simple insight to analyze i.i.d. stochastic models by exploiting the insensitivity property introduced earlier. However, we acknowledge that this approach may not be applicable in more general settings. For further details, please refer to Chapters 2, 7 and 8.

1.4. Contributions and outline

This thesis presents seven chapters on various stochastic models assessed with techniques from distribution-free analysis. Each chapter has its own notation that is tailored to its content. We now highlight the main contributions of each chapter and explain how all themes throughout the thesis fit together through the principles outlined in this introduction. We emphasize two key contributions, which serve as a unifying theme across all chapters:

1. We utilize semi-infinite linear programs and primal-dual techniques (as discussed in Section 1.2) to establish sharp bounds for the distribution-free analysis and optimization of stochastic models. These tight bounds provide an effective approach for handling distribution-free decision making problems.

2. We propose specific combinations of objective functions and ambiguity sets that result in insensitive extremal distributions. Such insensitivity not only allows us to evaluate stochastic models driven by i.i.d. sequences, but it also simplifies DRO problems by reducing them to stochastic programs with a worst-case distribution that is independent of the decision variables.

In Chapter 2, we address the extremal queue problem by determining the worst possible performance of the GI/G/1 queue under mean-dispersion constraints for the interarrival- and service-time distributions. To address this problem, we use as dispersion measure the MAD rather than the more commonly used variance. We then make the crucial observation that the expected waiting time can be written in the form

$$\mathbb{E}_{\mathbb{P}}[W_n] = \mathbb{E}_{\mathbb{P}}[\max\{0, X_1, \dots, X_1 + \dots + X_n\}],$$

which is componentwise convex in the random variables $X_1, \dots, X_n$. As a consequence, we can utilize the tight mean-MAD bounds, as derived in Section 1.2.2, for the distribution-free analysis of the GI/G/1 queue. Our analysis leverages the insensitivity property of the extremal distribution to derive the sharpest possible upper bounds for all moments of the waiting time, circumventing the computational issues that arise with the i.i.d. assumption. Furthermore, by utilizing the mean-MAD lower bound, we are able to produce tight lower bounds that, in conjunction with the tight upper bounds, provide distribution-free performance intervals.

In Chapter 3, we examine an M/M/s queue in which the arrival rate is a random variable Λ with only known mean, variance and range. Through the use of semi-infinite programming duality, we establish tight bounds for the expected wait. These bounds are based on an arrival rate that takes only two values. Unlike the distribution that maximizes Scarf's bound (1.4), this two-point distribution possesses the insensitivity property, as with the mean-MAD ambiguity set. The proofs rely heavily on the fact that the expected wait $W(\cdot)$, as a function of the arrival rate λ, has a convex derivative. Specifically, we demonstrate that the third derivative $W'''(\lambda)$ is nonnegative, which leads to the insensitivity property and substantially simplifies the distribution-free analysis of $\mathbb{E}_{\mathbb{P}}[W(\Lambda)]$.

In Chapter 4, we derive novel bounds on the tail probability of a random variable given support, mean and dispersion information. As alternatives for the renowned Chebyshev inequality, we present tight lower and upper bounds on the tail probability for random variables with known bounded support, mean and mean absolute deviation. We derive these bounds as exact solutions to semi-infinite linear programs, employing the weak duality framework set forth in Section 1.2.1. To this end, we consider as objective function $\mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\{X\geqslant t\}}(X)]$, where $\mathbb{1}_{\Xi}$ denotes the (discontinuous) indicator function for the event Ξ. We then apply these bounds for distribution-free analysis of the newsvendor model, stop-loss reinsurance and a chance-constrained optimization problem in radiotherapy.

In Chapter 5, we consider the problem of bounding conditional expectations based on information about the moments of the underlying distribution and the observed random event. For this, we consider objective functions that take the following semi-infinite programming form:

$$\mathbb{E}_{\mathbb{P}}[g(X)\,|\,X\geqslant t]=\frac{\mathbb{E}_{\mathbb{P}}[g(X)\mathbb{1}_{\{X\geqslant t\}}(X)]}{\mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\{X\geqslant t\}}(X)]}=\frac{\int g(x)\mathbb{1}_{\{X\geqslant t\}}(x)\mathrm{d}\mathbb{P}(x)}{\int \mathbb{1}_{\{X\geqslant t\}}(x)\mathrm{d}\mathbb{P}(x)},$$

which is nonlinear with respect to the probability measure $\mathbb{P}$; that is, the objective is a linear-fractional function of $\mathbb{P}$. However, through a simple transformation, we can reformulate this problem as a semi-infinite LP. This enables us to use the techniques described in Section 1.2.1 to obtain tight bounds for conditional expectations.

Chapter 6 focuses on the multi-item newsvendor problem with a constrained budget, where demand information is limited to its range, mean and mean absolute deviation. We determine optimal order quantities by solving a minimax problem that minimizes costs for the worst-case demand distributions. Using the mean-MAD bounds and recognizing that the newsvendor cost function is separable, i.e.,

$$\mathbb{E}_{\mathbb{P}}\Big[\sum_{i=1}^{n}(X_i-q_i)^+\Big]=\sum_{i=1}^{n}\mathbb{E}_{\mathbb{P}}[(X_i-q_i)^+],$$

we reduce the problem to a stochastic program with a particularly simple structure. This optimization problem can be solved through a greedy approach, reminiscent of the algorithm that resolves the continuous knapsack problem. The proposed method prescribes a policy that first ranks items based on their marginal effect on total cost and then orders them until the budget is exhausted.

Chapter 7 focuses on the Appointment Scheduling Problem (ASP), which involves scheduling planned appointments for a single server serving a given number of customers within a fixed period. The objective is to minimize a cost function that takes into account both the costs of waiting and overtime costs for the server. Even with fully specified service-time distributions, the ASP presents a computationally challenging stochastic program. When there is limited distributional information available, one can apply distribution-free analysis techniques to find the schedule that minimizes costs in worst-case scenarios. However, determining the extremal distributions under the independence assumption is difficult. Consequently, existing methods consider relaxations that allow dependence between service times, which may result in highly correlated extremal distributions. Despite the challenges posed by the independence assumption, the ASP can be analyzed using similar distribution-free techniques as those used in Chapter 2 for the GI/G/1 queue.

Finally, Chapter 8 concludes this thesis by examining classes of functions and distributional information that exhibit the insensitivity property. We also suggest possible directions for future research to develop a more general framework for distribution-free analysis of i.i.d. stochastic models.

2

MAD dispersion measure makes extremal queue analysis simple

2.1. Introduction

Queueing theory has existed for more than a century, with the GI/G/1 queue playing a central role in this theory as a model for a single server and independent generally distributed interarrival and service times. In this model, the waiting times of consecutive customers can be expressed as maxima of a random walk with step size equal in distribution to the difference of the service and interarrival times. This random walk and its maxima can be studied with mathematical techniques for sums of random variables, covered in many standard texts on probability theory: [7, 56, 79]. For all moments of the maxima (i.e., waiting times), general expressions are available that involve convolutions of the distribution of the step size. To use these general expressions, one thus needs to specify the precise distribution of the increments, and in the case of the GI/G/1 queue the distributions of both the interarrival and service times.

Special cases of the GI/G/1 queue can be studied with dedicated techniques for Markov chains. For instance, the M/G/1 queue with Poisson arrivals and the GI/M/1 queue with exponential services have explicit solutions that are more insightful than the general random walk results; see, for example, [7] and [57]. Another large, somewhat opposite branch of queueing theory concerns finding approximations and bounds. For the steady-state waiting time in the

This chapter is based on the research paper [210].

GI/G/1 queue, a classical bound for its expected value was obtained by Kingman [132] in terms of the first two moments of both the interarrival and the service time. While Kingman's bound is sharp in situations of heavy traffic, when the server's utilization approaches 1, it leaves room for improvement for all other values of the utilization.

In search for that sharpest possible (tight) upper bound when the first two moments of the interarrival and service times are given, significant progress was made by [24, 73, 187, 222], and it was conjectured by Daley in [24] that the overall worst-case behavior of the waiting time in the GI/G/1 queue would be caused by two-point distributions for both the interarrival and the service time. After this initial thrust of research, it remained silent for a while, until recently Chen and Whitt [49] derived sufficient conditions for the tight bounds to be attained by the conjectured two-point distributions. Without these additional conditions, the exact form of the extremal distributions can only be determined numerically, as the solution of a hard non-convex nonlinear optimization problem. Extensive numerical experiments led Chen and Whitt to conjecture that for the general problem, with undetermined service- and interarrival-time distributions, the worst case is formed by two-point distributions, in line with the conjecture postulated several decades ago. Finding the extremal queue for given mean-variance information is therefore one of the longest-standing problems in the field. That problem remains open.

We consider the same problem of finding the sharpest possible bounds for GI/G/1 queue performance measures, but we choose to quantify dispersion in terms of mean absolute deviation (MAD) instead of variance. MAD is hardly used in queueing theory, or random walk theory for that matter. The random walk and GI/G/1 queue are intrinsically linked with independent and identically distributed (i.i.d.) sums of random variables, and the variance thus naturally emerges as the quantity of interest (e.g., variance of the sum, central limit theorem). The variance and MAD, however, are equally adequate descriptors of dispersion, and are both easily calibrated on data using basic statistical estimators. Using the MAD instead of the variance as dispersion measure has several important advantages for, e.g., analyzing the waiting times in GI/G/1 queues. First, not only simple explicit expressions for the worst-case distributions can be obtained, but also for the best-case ones. Hence, a sharp upper bound and a sharp lower bound for the expected waiting time can be obtained. Second, our approach is for i.i.d. sums of random variables, while existing DRO approaches have to tolerate possible dependence structures between the random variables. Third, our approach is suitable for analyzing both transient behavior and the steady state. Fourth, because of its computational tractability our approach can also be extended to many optimization variants.

We organize the theoretical results in terms of three key theorems that build towards our main goal, finding the tightest possible upper bounds for all moments of the steady-state waiting time of the GI/G/1 queue with MAD as dispersion measure. We first provide a tight bound for the expectation of a general convex function of finitely many random variables with mean, MAD and range information. This is, in fact, a known result due to [19]. We present a new proof that finds the extremal distribution, the distribution that attains the tight bound, as the solution to a semi-infinite linear program (LP). The novel proof clearly illustrates why the MAD constraint leads to a tractable LP in the univariate setting. Since the solution of this LP does not depend on the objective function, we can recursively apply the univariate result to obtain independent extremal distributions that resolve the multivariate problem, thus making the latter amenable to distribution-free analysis. The traditional moment constraints for the mean and variance, although a popular choice, may not necessarily yield tractable counterparts because the worst-case distribution depends on the (multivariate) objective function. After showing the general result, we prove tight bounds for moments of random walk maxima and customer waiting times, all characteristics that can be brought into the form of the expectation of a convex function. An additional hurdle is to extend the transient setting to the case of infinitely many random variables, which is required for the all-time maximum, but also in that case, the extremal distribution for the step size remains the same three-point distribution. Finally, we resolve the extremal queue problem and present tight bounds for all moments of the waiting time. For this, we apply the random walk results combined with the additional reasoning that the step size is now replaced by the difference between the service and interarrival times. Consequently, the tights bounds for the stationary moments are attained by three-point distributions for both the interarrival and service time.

In addition to the theoretical results, we also present various results related to the application and computation of the tight bounds. The computational complexity of the bound for the expected finite-time maximum grows exponentially in the number of steps. For the random walk, we present an alternative expression that considers a number of terms that grows roughly as a cubic, rather than exponential, function of the number of partial sums, a substantial reduction in computational complexity for random walks with many steps. We also express, as easy to compute complex contour integrals, the tight moment bounds for the all-time maximum and the steady-state waiting time. In addition, we derive tight lower bounds that, together with the tight upper bounds, provide sharp performance intervals for distribution-free analysis. Further, for choosing the range of the step sizes, we propose a heuristic approach reminiscent of the construction of confidence intervals in statistical estimation. Finally, we illustrate the use of our approach when the mean and MAD are not known precisely and need to be estimated from data. The bounds remain effective also in these more realistic settings.

2.1.1. Contributions and outline

The contributions of this chapter can be summarized as follows:

1. We suggest to use MAD instead of variance, and obtain by concise mathematical proof the worst-case three-point distribution for a rich class of extremal problems. This proof for MAD gives insight into why the traditional moment constraints, although a popular choice, may not necessarily yield tractable counterparts.

2. We leverage this result to obtain tight upper and lower bounds for performance measures, including transient and steady-state queue length moments. Under mean-MAD constraints, these bounds are the sharpest possible (and thus cannot be improved). The mean-MAD approach is a new quantitative method applicable to random walks, queues and related stochastic processes. This generic approach is a computationally tractable

way to analyze key performance measures of such processes.

3. We present guidelines that describe how to compute the novel tight bounds efficiently. Moreover, we demonstrate our approach when the mean and MAD are not known precisely and need to be estimated from data. Also in these more realistic settings, the bounds remain sharp.

The remainder of this chapter is organized as follows. Section 2.2 presents the novel proof of the tight bounds for the expectation of general convex functions of random variables. Section 2.3 then derives sharp bounds for the moments of random walk maxima, and subsequently conveys both the transient and stationary results for the random walk to the GI/G/1 setting. In Section 2.4 we discuss the results that are related to the application and computation of the tight bounds, including the derivation of the matching lower bound. We conclude in Section 2.5, also mentioning various possibilities for further research.

2.2. Bounding a convex function using mean-MAD-range information

In this section, we explain how information about the mean, MAD and range can be used for establishing tight bounds for the expectation of a general convex function of random variables that match this information. In Section 2.2.1 we present, in Theorem 2.1, a general upper bound for the expectation of a convex function of independent random variables. We then provide a novel primal-dual proof for Theorem 2.1 in Section 2.2.2 and explain in Section 2.2.3 why MAD as dispersion measure is computationally tractable, and why variance is not.

2.2.1. Tight bound for convex functions

The moments of random walk maxima and GI/G/1 waiting times can all be expressed as the expectation of $h_n(X_1, \dots, X_n)$ with $h_n(\cdot)$ some function convex in the components of $(X_1, \dots, X_n)$ and $X_1, \dots, X_n$ independent random variables, which represent the interarrival and service times in the queueing model (or, alternatively, step sizes in the random walk model). Let $\mathbf{X} = (X_1, \dots, X_n)$ follow some distribution $\mathbb{P}$. We use the notation $\mathbf{X} \sim \mathbb{P} \in \mathscr{P}$ to say that $\mathbf{X}$ has a probability distribution $\mathbb{P}$ from the set of probability distributions $\mathscr{P}$. Let $\mathbb{E}_{\mathbb{P}}[\cdot]$ denote the expectation over the probability distribution $\mathbb{P}$.

Assuming we only have partial information consisting of the mean, dispersion measure and range of the random variables in $\mathbf{X}$, the first question we ask and answer in this chapter is: What *extremal* distributions of $\mathbf{X}$ result in the worst-case expectation of the convex function $h_n(X_1, \dots, X_n)$? This question thus requires solving the optimization problem

$$\max_{\mathbb{P} \in \mathscr{P}} \mathbb{E}_{\mathbb{P}}[h_n(\mathbf{X})] \tag{2.1}$$

with $\mathscr{P}$ the set of all considered distributions. To describe all considered distributions, we define an *ambiguity set* that consists of all distributions of componentwise independent $\mathbf{X}$ with known supports, means and MADs. The partial information for $(X_1, \dots, X_n)$ consists of (i) X_i

has bounded support, $\text{supp}(X_i) = [a_i, b_i]$ with $-\infty < a_i \leqslant b_i < \infty, i = 1, \ldots, n$, (ii) $\mathbb{E}_{\mathbb{P}}(X_i) = \mu_i$ and (iii) $\mathbb{E}_{\mathbb{P}}|X_i - \mu_i| = d_i$. This defines the ambiguity set

$$\mathscr{P}_{(\mu,d)} = \left\{ \mathbb{P} : \text{ supp}(X_i) \subseteq [a_i, b_i], \ \mathbb{E}_{\mathbb{P}}(X_i) = \mu_i, \ \mathbb{E}_{\mathbb{P}} |X_i - \mu_i| = d_i, \ \forall i, \ X_i \perp\!\!\!\perp X_j, \ \forall i \neq j \right\}, \quad (2.2)$$

where $X_i \perp\!\!\!\perp X_j$, $\forall i \neq j$, denotes stochastic independence of the components $X_1, \ldots, X_n$. To avoid trivialities we assume $\mu_i \in (a_i, b_i)$ and $d_i \in (0, \frac{2(b_i-\mu_i)(\mu_i-a_i)}{(b_i-a_i)})$. The latter interval follows from basic properties of MAD, which we discuss in Appendix A of this thesis. In what follows, $\mathbf{X}$ is a vector of random variables whose distribution belongs to the set $\mathscr{P}_{(\mu,d)}$.

The extremal distribution that solves (2.1) with $\mathscr{P} = \mathscr{P}_{(\mu,d)}$ can be shown to be a three-point distribution. This classical result follows from the general upper bound in [19] on the expectation of a convex function of independent random variables with mean-MAD ambiguity:

Theorem 2.1. *The extremal distribution that solves*

$$\max_{\mathbb{P} \in \mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}[h_n(\mathbf{X})] \quad (2.3)$$

consists for each X_i of a three-point distribution with values $\tau_1^{(i)} = a_i$, $\tau_2^{(i)} = \mu_i$, $\tau_3^{(i)} = b_i$ and probabilities

$$p_1^{(i)} = \frac{d_i}{2(\mu_i - a_i)}, \quad p_2^{(i)} = 1 - \frac{d_i}{2(\mu_i - a_i)} - \frac{d_i}{2(b_i - \mu_i)}, \quad p_3^{(i)} = \frac{d_i}{2(b_i - \mu_i)}. \quad (2.4)$$

Ben-Tal and Hochman [19] proved Theorem 2.1 by introducing a piecewise linear function on the interval $[a, b]$ that intersects the convex function in a, μ and b, and then applying the classic Edmundson-Madansky bound to the subintervals $[a, \mu]$ and $[\mu, b]$. In the next subsection, we give another proof of Theorem 2.1 that also elucidates why using as dispersion measure MAD instead of variance makes the analysis so simple.

A direct consequence of Theorem 2.1 is that the worst-case expectation of $h_n(\mathbf{X})$ is obtained by enumerating over all 3^n permutations of outcomes a_i, μ_i, b_i of components X_i, as formulated in the next result.

Corollary 2.2. *It holds that*

$$\max_{\mathbb{P} \in \mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}[h_n(\mathbf{X})] = \sum_{\boldsymbol{\alpha} \in \{1,2,3\}^n} h_n(\tau_{\alpha_1}^{(1)}, \ldots, \tau_{\alpha_n}^{(n)}) \prod_{i=1}^{n} p_{\alpha_i}^{(i)}. \quad (2.5)$$

2.2.2. Novel primal-dual proof of Theorem 2.1

Our proof of Theorem 2.1 will crucially rely on the fact that the solution of the univariate case can be straightforwardly extended to the multivariate case. We thus start by considering some univariate measurable function $f(x)$ (with the univariate function $h_1(x_1)$ as an example) that has finite values on $[a, b]$, the support of the distribution $\mathbb{P}$. Under mean-MAD ambiguity of one random variable X, we then need to solve

$$\begin{aligned} \max_{\mathbb{P}(x)\geqslant 0} \quad & \int_x f(x)\mathrm{d}\mathbb{P}(x) \\ \text{s.t.} \quad & \int_x \mathrm{d}\mathbb{P}(x) = 1, \ \int_x x\,\mathrm{d}\mathbb{P}(x) = \mu, \ \int_x |x-\mu|\,\mathrm{d}\mathbb{P}(x) = d, \end{aligned} \tag{2.6}$$

a semi-infinite linear program with three equality constraints. A perhaps surprising, yet classical fact, is that the semi-infinite LP (2.6) can be reduced to an equivalent finite-dimensional optimization problem that yields the same optimal value. Indeed, the Richter-Rogosinski theorem [98, 186, 198] states that there exists an extremal distribution for problem (2.6) with at most three support points. While finding these points in closed form is typically not possible (for general semi-infinite problems), we have shown in Chapter 1 that this is possible for the problem at hand, by resorting to its dual problem,

$$\begin{aligned} \min_{\lambda_0,\lambda_1,\lambda_2} \quad & \lambda_0 + \lambda_1\mu + \lambda_2 d \\ \text{s.t.} \quad & M(x) := \lambda_0 + \lambda_1\mu + \lambda_2|x-\mu| \geqslant f(x), \ \forall x \in [a,b], \end{aligned} \tag{2.7}$$

and exploiting the specific structure of $M(x)$ induced by the MAD constraint $\int_x |x-\mu|\mathrm{d}\mathbb{P}(x) = d$. As explained in Chapter 1, the tightest majorant $M(x)$, with λ_1, λ_2 and λ_3 chosen optimally, touches $f(x)$ at three points: $x = a$, μ and b. For illustrative purposes, Figure 2.1 is presented once more. Because the majorant is piecewise linear and convex, we can majorize every convex function $f(x)$, choosing $M(x)$ such that it touches at the boundary points a and b, as well as the kink point $x = \mu$. The optimal probabilities of these support points can now easily be obtained by solving the linear system resulting from the equations of (2.6), and primal-dual optimality can be demonstrated as outlined in Section 1.2.2.

To deal with the multivariate case, we successively apply the univariate result. The multivariate problem can be expressed as

$$\max_{\mathbb{P}_n\in\mathscr{P}^n_{(\mu,d)}} \cdots \max_{\mathbb{P}_2\in\mathscr{P}^2_{(\mu,d)}} \max_{\mathbb{P}_1\in\mathscr{P}^1_{(\mu,d)}} \mathbb{E}_{\mathbb{P}_\otimes}[h_n(X_1, X_2, \dots, X_n)]. \tag{2.8}$$

Here, $\mathbb{P}_i$ and $\mathscr{P}^i_{(\mu,d)}$, for all i, represent the marginal distributions and ambiguity sets of $X_1, \dots, X_n$, respectively, and $\mathbb{P}_\otimes$ denotes the product measure. Let $h_n(X_1, \dots, X_n)$ be componentwise convex in the vector $\mathbf{X}$. Suppose we first apply the univariate result to X_1, then we need to consider as objective functional $\mathbb{E}_{\mathbb{P}_1}[h_n(X_1, x_2, \dots, x_n)]$, with $x_2, \dots, x_n$ all being fixed. Notice that $h(\cdot, x_2, \dots, x_n)$ is a convex function of X_1 for all possible realizations of $X_2, \dots, X_n$. Hence, for all realizations $x_2, \dots, x_n$, $\mathbb{E}[h_n(X_1, x_2, \dots, x_n)]$ can be bounded using the univariate result. As stated in Theorem 2.1, the worst-case distribution is independent of the values for $x_2, \dots, x_n$. Then, by substituting the extremal distribution for X_1, we obtain

$$\max_{\mathbb{P}_n\in\mathscr{P}^n_{(\mu,d)}} \cdots \max_{\mathbb{P}_2\in\mathscr{P}^2_{(\mu,d)}} \mathbb{E}_{\mathbb{P}_\otimes}\Big[\sum_{\alpha_1\in\{1,2,3\}} p^{(1)}_{\alpha_1} h_n(\tau^{(1)}_{\alpha_1}, X_2, \dots, X_n)\Big]. \tag{2.9}$$

Since the worst-case probabilities for x_1 are nonnegative, the worst-case expectation becomes a convex function in $x_2, \dots, x_n$. Consequently, we can apply the univariate result to x_2 in a likewise manner. Successively repeating this reasoning n times completes the proof.

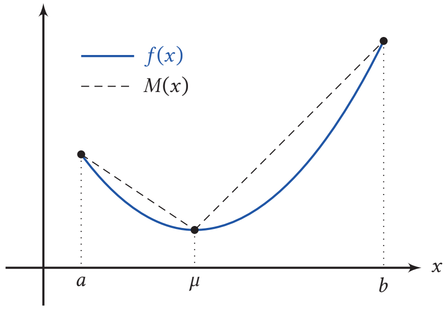


Figure 2.1: Some convex function $f(x)$ and its piecewise linear majorant $M(x)$

To the best of our knowledge, our proof is the first to exploit the specific shape of the kink-majorant to find an analytic solution for the semi-infinite LP. While the dual problems are often solvable as semidefinite or second-order conic programs, analytic solutions as in our case are typically hard to attain, and require special structural properties of the LP's objective function and its constraints. In the univariate case, this proof method does not require convexity of $f(x)$ and works for arbitrary measurable functions, as demonstrated in (1.2.1). Convexity is needed, however, in the proof of Theorem 2.1 to extend the univariate case to the multivariate case. The proof method is of independent interest, and can for instance be applied to study the mean-MAD counterparts of the mean-variance analyses in, for example, [61, 163, 166, 174, 226].

2.2.3. Why is MAD computationally easier than variance?

Now that we fully grasp why and how the proof of Theorem 2.1 relies on the specific structural properties of the mean-MAD constraints, and in particular the univariate result seamlessly passes into the multivariate counterpart, we can also explain why the comparable challenge with mean-variance constraints becomes much more difficult if not impossible. For the univariate case, the same proof argument works when σ^2 is given instead of d, i.e., when $|x-\mu|$ in (2.6) is replaced by $(x-\mu)^2$. Hence, irrespective of whether MAD or variance is used as dispersion measure, for determining the tight upper bound of $f(x)$, it suffices to consider distributions with support on at most three points. There is, however, a crucial complication when extending to the multivariate case.

To see this, observe that when σ^2 is used as dispersion measure, the end points and kink point do *not* necessarily span the support of the extremal distribution. That is, upon replacing $|x-\mu|$ with $(x-\mu)^2$, the tightest majorant $F(x)$ does not necessarily touch $f(x)$ in a, b and μ. Hence, if the variance is used as dispersion measure, then the worst-case distribution might depend on the function $f(x)$. This has severe consequences for the multivariate case, i.e., when we consider $h_n(x_1,\ldots,x_n)$. In that case, the worst-case distribution depends on the values of $x_2,\ldots,x_n$, and calculating (in closed form) the worst-case distribution as a function of $x_2,\ldots,x_n$ is not straightforward. Even if we were able to derive such a worst-case distribution, substituting this distribution in the worst-case expectation might result in a complicated function of $x_2,\ldots,x_n$

that is likely nonconvex, and hence applying the univariate result to x_2 is no longer possible.

2.3. Tight bounds for random walks and queues

We now leverage the general tight bounds in Theorem 2.1 for proving tight bounds for moments of random walk maxima in Section 2.3.1 (Theorem 2.3) and the GI/G/1 queue waiting time in Section 2.3.2 (Theorem 2.4). In Section 2.3.3 we draw a technical comparison between the extremal queue problem with MAD information and variance information. Section 2.3.4 provides some further insights regarding the mean-variance ambiguity set.

2.3.1. Extremal random walk maxima

Consider the partial sums $S_n := X_1 + \cdots + X_n$ ($S_0 := 0$) of i.i.d. random variables $X_1, X_2, \ldots$ distributed as X. The random walk $(S_n, n \geqslant 0)$ arises in many application domains, including queueing theory, inventory management and risk theory; see, e.g., Chapter XIV of [7]. If $(S_n, n \geqslant 0)$ indeed models congestion, shortfall or capital position, large values of S_n are of particular interest, and it is natural to consider the maxima sequence $M_n := \max\{S_0, S_1, \ldots, S_n\}$.

Notice that M_n can be expressed as $h_n(X_1, \ldots, X_n)$, with

$$h_n(x_1, \ldots, x_n) = \max\{0, x_1, \ldots, x_1 + \cdots + x_n\}, \tag{2.10}$$

and the expected maximum can be expressed as $\mathbb{E}[M_n] = \mathbb{E}[h_n(\mathbf{X})]$, with $h_n(\mathbf{X})$ jointly convex in $\mathbf{X} = (X_1, \ldots, X_n)$. Thus, under the partial information contained in $\mathscr{P}_{(\mu,d)}$, (2.5) is an upper bound on $\mathbb{E}[M_n]$ that cannot be improved. Let $X_{(3)}$ denote the random variable with the extremal three-point distribution, identified in Theorem 2.1 but for the special case when $X_1, X_2, \ldots$ are i.i.d., that attains this bound. To state our results and for later reference, let $\Omega(\mu, d, a, b)$ denote a three-point distribution on the values $\{a, \mu, b\}$ with probabilities

$$p_1 = \frac{d}{2(\mu - a)}, \quad p_2 = 1 - \frac{d}{2(\mu - a)} - \frac{d}{2(b - \mu)}, \quad p_3 = \frac{d}{2(b - \mu)}. \tag{2.11}$$

Then, $X_{(3)} \sim \Omega(\mu, d, a, b)$.

For $\mathbb{E}[X] < 0$ the all-time maximum $M := \lim_{n\to\infty} M_n$ is a proper random variable (M_n converges in distribution to M, which will be finite with probability one if $\mathbb{E}[X] < 0$). Let $c_m(M)$ denote the m-th cumulant of M. Recall that $c_1(M)$ is the mean, $c_2(M)$ is the variance, and $c_3(M)$ is the central moment $\mathbb{E}[(M - \mathbb{E}[M])^3]$. From general random walk theory we know that (see e.g., [1])

$$c_m(M) = \sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E}[(S_k^+)^m]. \tag{2.12}$$

We can now prove results similar as for $\mathbb{E}[M_n]$, regarding the worst-case distribution and tight upper bound. The following theorem states the extremal solution for the all-time maximum of the random walk.

Theorem 2.3. *Consider the random walk with generic step size X contained in the ambiguity set $\mathscr{P}_{(\mu,d)}$. The tight upper bounds for all cumulants $c_m(M)$ of the all-time maximum M are the cumulants of the random walk with extremal step size $X_{(3)}$.*

Proof. Consider the function

$$f_n^m(x_1, \ldots, x_n) = \sum_{k=1}^{n} \frac{1}{k} \left(\max\{0, x_1 + \ldots + x_k\}\right)^m, \tag{2.13}$$

which is convex in the vector $(x_1, \ldots, x_n)$. Hence, for i.i.d. increments with generic X,

$$\max_{\mathbb{P}\in\mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}[f_n^m(\mathbf{X})] \tag{2.14}$$

is solved by the extremal random variable $X_{(3)}$. This gives the bound, with $X_1^*, X_2^*, \ldots$ i.i.d. as $X_{(3)}$,

$$l_n := \sum_{k=1}^{n} \frac{1}{k} \mathbb{E}[(S_k^+)^m] \leqslant \mathbb{E}[f_n^m(X_1^*, \ldots, X_n^*)] =: u_n. \tag{2.15}$$

The result follows by observing that the sequences $\{l_n\}$ and $\{u_n\}$ are both monotone, and converging to well-defined limits. □

We conclude that the extremal three-point distribution for $\mathbb{E}[M_n]$ in Theorem 2.1 is also the extremal distribution for all cumulants of M. When calculating the associated tight upper bounds for $c_m(M)$, (2.12) shows that we are confronted with an infinite summation of increasingly complex summands. Here, another line of classical random walk theory can help, which transforms such infinite sums into complex contour integrals. We discuss these methods in Section 2.4.1.

2.3.2. Extremal GI/G/1 queue

We now turn to the extremal GI/G/1 queue problem. Consider a single-server queue with i.i.d. interarrival times $\{U_n\}$ distributed as U, i.i.d. service times $\{V_n\}$ distributed as V, and server utilization $\rho = \mathbb{E}[V]/\mathbb{E}[U] < 1$. Let W_n be the waiting time of customer n. The sequence $(W_n, n \geqslant 0)$ with $W_0 = 0$ satisfies the Lindley recursion

$$W_{n+1} = (W_n + V_n - U_n)^+, \quad n \geqslant 0. \tag{2.16}$$

Let W be the steady-state waiting time. Since $W_n \stackrel{d}{=} M_n$ and $W \stackrel{d}{=} M$ the results for the random walk maxima will carry over to the waiting times. The main difference is that the step size X is now interpreted as the difference $V - U$ between the generic service time and generic interarrival time. If one has mean-MAD information about both V and U this is more informative than mean-MAD information about $V - U$, and this additional information should lead to even sharper bounds. Let us consider the steady-state queue length W, which satisfies $W \stackrel{d}{=} (W + V - U)^+$. Denote by σ_U^2 and σ_V^2 the variances of U and V, respectively. Let $\rho = \mathbb{E}[V]/\mathbb{E}[U] < 1$. Then, the following bounds for $\mathbb{E}[W]$ are known if one possesses information about the first two moments of U and V:

- Kingman's upper bound:

$$\mathbb{E}[W] \leqslant \frac{\sigma_V^2 + \sigma_U^2}{2(\mathbb{E}[U] - \mathbb{E}[V])}. \tag{2.17}$$

- Daley's upper bound:

$$\mathbb{E}[W] \leqslant \frac{\sigma_V^2 + \rho(2-\rho)\sigma_U^2}{2(\mathbb{E}[U] - \mathbb{E}[V])}. \tag{2.18}$$

- Upper bound of Chen and Whitt [49] based on the two-point conjecture:

$$\mathbb{E}[W] \leqslant \frac{\sigma_V^2 + \kappa(\rho)\sigma_U^2}{2(\mathbb{E}[U] - \mathbb{E}[V])}, \tag{2.19}$$

with $\kappa(\rho) = 2\rho(1-\rho)/(1-\delta)$ and $\delta \in (0,1)$ the solution of $\delta = \exp(-(1-\delta)/\rho)$.

We shall compare these bounds with the tight bounds for mean-MAD-range information that we derive next.

The GI/G/1 queue assumes that interarrival times and service times are independent, so it is natural to assume that V has ambiguity set $\mathscr{P}_{(\mu_V,d_V)}$ and U has ambiguity set $\mathscr{P}_{(\mu_U,d_U)}$. We then consider all distributions $\mathbb{P}$ that lie in $\mathscr{P}_{(\mu_V,d_V)} \times \mathscr{P}_{(\mu_U,d_U)}$, the set containing all product measures of feasible marginal distributions for V and U. The extremal queue problem with mean-MAD dispersion information can now be phrased as

$$\max_{\mathbb{P}\in\mathscr{P}_{(\mu_V,d_V)}\times\mathscr{P}_{(\mu_U,d_U)}} \mathbb{E}[f(\mathbf{X})], \tag{2.20}$$

where $\mathbb{E}[f(\mathbf{X})]$ describes $\mathbb{E}[W_n]$ or $c_m(W)$ and $\mathbf{X}$ is the random vector with elements U_1, V_1, U_2, $V_2, \ldots$. This is the classical setting of the extremal GI/G/1 queue treated in [49, 73, 187, 222], but with MADs instead of variances describing the ambiguity set. Let the random variables $V_{(3)}$ and $U_{(3)}$ follow the extremal three-point distributions $\Omega(\mu_V, d_V, a_V, b_V)$ and $\Omega(\mu_U, d_U, a_U, b_U)$, respectively. The next result presents the solution to the extremal queue problem in (2.20).

Theorem 2.4. *Consider the GI/G/1 queue with generic interarrival time U with ambiguity set $\mathscr{P}_{(\mu_U,d_U)}$ and generic service times V with ambiguity set $\mathscr{P}_{(\mu_V,d_V)}$. Consider the tight upper bounds for the transient mean waiting time $\mathbb{E}[W_n]$ and all cumulants of the steady-state waiting time W.*

(i) *For given interarrival time U, the tight upper bounds follow from the service time $V_{(3)}$.*

(ii) *For given service time V, the tight upper bounds follow from the interarrival time $U_{(3)}$.*

(iii) *The overall tight upper bounds follow from interarrival time $U_{(3)}$ and service time $V_{(3)}$.*

Proof. As in Theorem 2.1, the tight bounds for $\mathbb{E}[W_n]$ follow from the general upper bound in [19] on the expectation of a convex function of the random vector $(X_1, \ldots, X_n)$ with mean-MAD ambiguity, but now with X_i replaced by $V_i - U_i$. The function describing $\mathbb{E}[W_n]$, see expression (2.10), is indeed convex in *both* V_i and U_i, and hence the result follows. Similarly, the expression for $c_m(W)$ (see Theorem 2.3) is also convex in both V_i and U_i, and hence the tight bounds for $c_m(W)$ follow from our proof of Theorem 2.3. □

Table 2.1 shows an example of the tight bound for $\mathbb{E}[W]$ associated with $(U_{(3)}, V_{(3)})$, and compares it with the other known bounds that require variance information.The variance of the extremal three-point distribution $\Omega(\mu, d, a, b)$ is $\frac{d}{2}(b-a)$, the maximal variance for distributions in the ambiguity set $\mathscr{P}_{(\mu,d)}$. We thus know the variances of $U_{(3)}$ and $V_{(3)}$, and can calculate the other three bounds. In heavy traffic, Kingman's bound is known to be asymptotically correct, and hence the other three (sharper) bounds also converge to the heavy-traffic limit as $\rho \uparrow 1$. Furthermore, we remark that the mean-MAD bounds in Theorem 2.4 are crucially influenced by the choice of range, in this example set to $[0, 10]$ for both the interarrival- and service-time distributions. Notice that Table 2.1 is not meant to compare mean-MAD with mean-variance bounds. The displayed differences merely express different ways of dealing with ambiguity. We make a connection between mean-MAD and mean-variance information in Section 2.3.4.

Table 2.1: Bounds for $(1-\rho)\mathbb{E}[W]/\rho$ for $(\mu_U, d_U, a_U, b_U) = (1, 1, 0, 10)$ and $(\mu_V, d_V, a_V, b_V) = (\rho, 0.1, 0, 10)$

ρ	Thm. 2.4	C & W (2.19)	Daley (2.18)	Kingman (2.17)
0.1	4.06613	7.00020	7.25000	27.50000
0.2	2.52306	5.27810	5.75000	13.75000
0.5	2.03141	3.63750	4.25000	5.50000
0.7	2.49160	3.17138	3.60714	3.92857
0.8	2.61932	3.00523	3.31250	3.43750
0.9	2.69802	2.86711	3.02778	3.05556
0.95	2.72609	2.80627	2.88816	2.89474
0.99	2.74547	2.76091	2.77753	2.77778

2.3.3. Comparison with classical extremal queue problem

For the variance counterpart, Chen and Whitt also formulate a semi-infinite linear optimization problem. The crucial difference is that they cannot use the univariate function extension (as explained in Section 2.2), and hence should work directly with the multivariate function. This in turn implies that the dual problem cannot be solved explicitly (like in the univariate case), let alone that there is a zero duality gap. Another complication is that the expectation of the multivariate function (2.13) cannot be expressed directly in V and U, but rather in terms of convolutions of the distributions of V and U. Chen and Whitt [49] try to circumvent these challenges by applying stochastic comparison techniques to obtain results for the mean steady-state waiting time. In this way, Chen and Whitt [49] prove a similar but weaker result than Theorem 2.4 for the steady-state setting, but with variance as dispersion measure. They provide sufficient conditions that guarantee the extremal distributions of U and V to be the widely conjectured two-point distributions.

An important message of this chapter is that with MAD the extremal distribution remains unaltered going from the univariate to the multivariate setting, and that with variance this reasoning fails. In fact, one intuitively expects formidable challenges when seeking for extremal distributions under variance constraints. This intuition is confirmed by Chen and Whitt's formulation of the extremal distribution as the solution of a nonconvex nonlinear optimization problem. While this optimization problem can be solved numerically, a closed-form solution and hence identification of the extremal distribution remains out of reach.

Under variance constraints, it is thus conjectured that the tight bound comes from specific two-point distributions for both U and V. In fact, the bound (2.19) in Table 2.1 holds under the assumption that this conjecture is true, and was shown by Chen and Whitt [50] to be very close to the tight upper bound. Theorem 2.4 rules out a similar conjecture in the MAD setting. The tight bounds in Theorem 2.4 always involve three-point distributions.

2.3.4. Further comparison with mean-variance ambiguity

As already explained in Section 2.2.3, mean-variance ambiguity appears less computationally tractable than mean-MAD ambiguity. Let $\mathscr{P}^*_{(\mu,\sigma)}$ denote the ambiguity set that contains all distributions with known range, mean and variance, i.e.,

$$\mathscr{P}^*_{(\mu,\sigma)} = \left\{\mathbb{P} \,:\, \operatorname{supp}(X_i) \subseteq [a_i, b_i],\ \mathbb{E}_{\mathbb{P}}(X_i) = \mu,\ \mathbb{E}_{\mathbb{P}}(X_i - \mu)^2 = \sigma^2,\ \forall i,\ X_i \perp\!\!\!\perp X_j,\ \forall i \neq j\right\}. \tag{2.21}$$

We now show how the key result for mean-MAD ambiguity, Theorem 2.1, can be used to obtain results for mean-variance ambiguity, using the following property:

Proposition 2.5. *Let* $d_{\min} = 2\sigma^2/(b-a)$ *and* $d_{\max} = \sigma$. *Then,*

$$\max_{\mathbb{P}\in\mathscr{P}_{(\mu,d_{\min})}} \mathbb{E}_{\mathbb{P}}[h_n(\mathbf{X})] \leqslant \max_{\mathbb{P}\in\mathscr{P}^*_{(\mu,\sigma)}} \mathbb{E}_{\mathbb{P}}[h_n(\mathbf{X})] \leqslant \max_{\mathbb{P}\in\mathscr{P}_{(\mu,d_{\max})}} \mathbb{E}_{\mathbb{P}}[h_n(\mathbf{X})]. \tag{2.22}$$

Proof. From [21], we know that

$$\frac{2\sigma^2}{b-a} \leqslant d \leqslant \sigma.$$

Hence, $\max_{\mathbb{P}\in\mathscr{P}^*_{(\mu,\sigma)}} \mathbb{E}_{\mathbb{P}}[h_n(\mathbf{X})] = \max_{\mathbb{P}\in\mathscr{P}_{(\mu,d^*)}} \mathbb{E}_{\mathbb{P}}[h_n(\mathbf{X})]$ for some $d^* \in [2\sigma^2/(b-a), \sigma]$. Since $\max_{\mathbb{P}\in\mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}[h_n(\mathbf{X})]$ is nondecreasing in d, see [179], the result follows. □

This result presents a way to delimit the upper bounds of all stationary cumulants $c_m(M)$ and the transient mean $\mathbb{E}[M_n]$ under mean-variance ambiguity. The mean-MAD bounds are specified in terms of specific three-point distributions.

We next show that the lower bound in Proposition 2.5 can lead to a result for infinite-support distributions. Set $b = a + \xi(\mu - a)$ with $\xi \geqslant 1$, and observe that the lower bound in (2.22) comes with the extremal three-point distribution

$$X^{\xi}_{(3)} = \begin{cases} a & \text{w.p. } \frac{\sigma^2}{(\mu-a)^2\xi}, \\ \mu & \text{w.p. } 1 - \frac{\sigma^2}{(\mu-a)^2\xi} - \frac{\sigma^2}{(\mu-a)^2\xi(\xi-1)}, \\ a + \xi(\mu - a) & \text{w.p. } \frac{\sigma^2}{(\mu-a)^2\xi(\xi-1)}. \end{cases}$$

This distribution has mean μ and variance σ^2, irrespective of the range $[a, b]$. We can thus let ξ grow to infinity to investigate what happens for infinite-support distributions.

For the expected all-time maximum, we can exploit an argument very similar to [50], Theorem 5.1. A classic result from regenerative analysis says that the expected all-time maximum is the expected sum of the random walk position over one cycle, denoted by $\mathbb{E}$[integral], divided by the expected length of one cycle, i.e. $\mathbb{E}$[cycle length]; see Sections 3.6 and 3.7 of [189]. This cycle will consist of a period during which the queue remains empty, corresponding to consecutive (negative) steps of size a or μ. As ξ increases, the three-point distribution places probabilities of order $O(1/\xi^2)$ on a and $a+\xi(\mu-a)$, and the rest of the mass on point μ. As ξ grows large, only rarely with probability $O(1/\xi^2)$, a large positive step occurs. The impact of the very large step of size $a+\xi(\mu-a)$ is roughly the area of the triangle with height $a+\xi(\mu-a)$ and width $(a+\xi(\mu-a))/(-\mu)$, and hence $\mathbb{E}[\text{integral}] = (a+\xi(\mu-a))^2/(-2\mu) \sim (\xi(\mu-a))^2/(-2\mu)$ as $\xi \to \infty$. The cycle then consists of an empty period of expected length $(1-p_b)/p_b \sim (\xi(\mu-a))^2/\sigma^2$ and the positive period due to the large step of expected length $(a+\xi(\mu-a))/(-\mu)$, so that $\mathbb{E}[\text{cycle length}] \sim (\xi(\mu-a))^2/\sigma^2$, and the expected all-time maximum converges to $\sigma^2/(-2\mu)$ as $\xi \to \infty$. Since this is a lower bound for $\max_{\mathbb{P}\in\mathcal{P}^*_{(\mu,\sigma)}} \mathbb{E}[M]$, we know that for the random walk with generic step size X it holds that $\max_{\mathbb{P}\in\mathcal{P}^*_{(\mu,\sigma)}} \mathbb{E}[M] \geqslant \sigma^2/(-2\mu)$. This lower bound matches Kingman's upper bound $\mathbb{E}[M] \leqslant \sigma^2/(-2\mu)$, which proves that Kingman's upper bound is tight. Tightness of Kingman's bound was already proven in [60] by identifying a two-point distribution with mean μ, variance σ^2 such that $\mathbb{E}[M]$ approaches the upper limit as one of the two points goes to infinity.

2.4. Computational guidelines

This section presents various guidelines for using and calculating the tight bounds in Theorems 2.1–2.4. Section 2.4.1 shows how the computational complexity that comes with convolutions of three-point distributions can be reduced considerably. Section 2.4.2 discusses some of the numerical aspects of the contour integrals used to determine the tight bounds in the stationary setting. Section 2.4.3 shows how tight lower bounds can be derived, using a similar proof method as employed for the tight upper bounds. Section 2.4.4 extends the computational results for the random walk to the GI/G/1 queue. Section 2.4.5 investigates the impact of the range on the bounds, and Section 2.4.6 demonstrates what happens if this range is unbounded from above. Finally, Section 2.4.7 presents a data-driven approach for estimating the mean and MAD.

2.4.1. Spitzer's formula and Pollaczek's integral for three-point distributions

We recall that Spitzer [202] used combinatorial arguments to establish for $\mathbb{E}[M_n]$ the alternative expression (which strictly requires i.i.d. increments)

$$\mathbb{E}[M_n] = \sum_{k=1}^{n} \frac{1}{k}\mathbb{E}[S_k^+], \tag{2.23}$$

with $x^+ = \max\{0, x\}$. This can be written as $\mathbb{E}[M_n] = \mathbb{E}[f_n(\mathbf{X})]$ with

$$f_n(x_1, \dots, x_n) = \sum_{k=1}^{n} \frac{1}{k} \max\{0, x_1 + \dots + x_k\}. \tag{2.24}$$

A first usage of Spitzer's formula (2.23) is a considerable improvement, in terms of computational complexity, of the tight bound for $\mathbb{E}[M_n]$ in (2.5). By applying Theorem 2.1 to (2.23), we get the following result.

Corollary 2.6.

$$\max_{\mathbb{P}\in\mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}[f_n(\mathbf{X})] = \sum_{k=1}^{n} \frac{1}{k} \sum_{\sum_i k_i = k} \max\{0, k_1 a + k_2 \mu + k_3 b\} \cdot \frac{k!}{k_1! k_2! k_3!} p_1^{k_1} p_2^{k_2} p_3^{k_3}. \tag{2.25}$$

Note that for each fixed k, (2.25) contains a multinomial distribution with support set $\{(k_1, k_2, k_3) \in \mathbb{N}^3 : k_1 + k_2 + k_3 = k\}$ with cardinality $\binom{k+2}{2}$. This implies that the sum over k in (2.25) is over roughly n^3 terms, which is way better than the 3^n terms in (2.5).

Consider the random walk with generic step size X. It is known that formal solutions of the distribution of M_n and M can be expressed in terms of complex contour integrals (see [1, 121] for the algorithmic aspects of these contour integrals). Assume that $\phi_X(s) = \mathbb{E}[\mathrm{e}^{sX}]$ is analytic for complex s in the strip $|\mathrm{Re}(s)| < \delta$ for some $\delta > 0$. A sufficient condition is that the moment generating function $\phi_X(s)$ is finite in a neighborhood of the origin, and hence all moments of X exist. Then

$$\mathbb{E}[\mathrm{e}^{-sM}] = \exp\left\{ \frac{-1}{2\pi i} \int_{\mathscr{C}} \frac{s}{u(s-u)} \log(1 - \phi_X(-u)) \mathrm{d}u \right\}, \tag{2.26}$$

where s is a complex number with $\mathrm{Re}(s) \geqslant 0$, $\mathscr{C}$ is a contour to the left of, and parallel to, the imaginary axis, and to the right of any singularities of $\log(1 - \phi_X(-u))$ in the left half plane. From (2.26) contour integral expressions for the cumulants follow by differentiation:

$$c_m(M) = \frac{(-1)^m\, m!}{2\pi i} \int_{\mathscr{C}} \frac{\log(1 - \phi_X(-u))}{u^{m+1}} \mathrm{d}u. \tag{2.27}$$

Consider $X = X_{(3)}$ with a three-point distribution on values $\{a, \mu, b\}$ with probabilities p_1, p_2, p_3 and moment generating function

$$\phi_{X_{(3)}}(s) = p_1 \mathrm{e}^{sa} + p_2 \mathrm{e}^{s\mu} + p_3 \mathrm{e}^{sb}. \tag{2.28}$$

All moments of $X_{(3)}$ exist, and hence $\phi_{X_{(3)}}(s)$ satisfies the assumption required for representation (2.26) to hold. Since $X_{(3)}$ follows the extremal three-point distribution associated with the tight upper bounds for $c_m(M)$, we obtain the following result:

Corollary 2.7. *Let* $\phi_{X_{(3)}}(s) := \mathbb{E}[\mathrm{e}^{sX_{(3)}}] = p_1 \mathrm{e}^{sa} + p_2 \mathrm{e}^{s\mu} + p_3 \mathrm{e}^{sb}$. *The tight upper bounds on* $c_m(M)$ *identified in* Theorem 2.3 *are given by*

$$\frac{(-1)^m\, m!}{2\pi i} \int_{\mathscr{C}} \frac{\log(1 - \phi_{X_{(3)}}(-u))}{u^{m+1}} \mathrm{d}u, \quad m = 1, 2, \dots, \tag{2.29}$$

where $\mathscr{C}$ *is a contour to the left of, and parallel to, the imaginary axis, and to the right of any singularities of* $\log(1 - \phi_{X_{(3)}}(-u))$ *in the left half plane.*

Expression (2.29) bypasses the cumbersome calculations with convolutions in (2.12). In the next subsection, we demonstrate that this is a numerically efficient way of computing the tight bounds.

2.4.2. Numerical experiments with contour integrals

Numerical aspects of integrals of the type (2.27) have been discussed in e.g., [1, 50, 121]. For distributions with support on a finite set of points, potential numerical problems can arise, because $|\mathrm{Re}(\phi_X(u))|$ does not converge to zero as $|u| \to \infty$; see [2, 3, 50]. For the three-point distributions required in this chapter we have performed extensive numerical experiments with (2.29). These experiments confirmed that the integrals can be calculated up to high accuracy with standard integration routines in Mathematica. For many parameter values a, b, μ, d such that (A.1) holds, we have calculated $\mathbb{E}[M]$ for generic increment $X_{(3)}$ using (2.29), and compared this with results from extensive stochastic simulations. We also compared the results with a third numerical procedure, known to be extremely stable and accurate. Let us explain the third procedure, which might be of independent interest: Choose the boundaries of the support as multiples of $\beta = |\mu|$ by writing that $a = -s\beta$ and $b = m\beta$ with s, m positive integers. Denote by $M_\beta = M/\beta$ the normalized steady-state waiting time. We then get

$$M_\beta \overset{d}{=} (M_\beta + X_\beta)^+,$$

with $X_\beta = X/\beta$ a discrete random variable with support $\{-s, -1, m\}$ and MAD

$$d_\beta := \mathbb{E}[|X_\beta - \mathbb{E}[X_\beta]|] = \frac{1}{\beta}\mathbb{E}[|X - \mathbb{E}[X]|] = \frac{d}{\beta}.$$

Define $X_\beta = A_\beta - s$, so that

$$M_\beta \overset{d}{=} (M_\beta + A_\beta - s)^+$$

for a discrete random variable A_β with support $\{0, s-1, s+m\}$ and probability generating function

$$\mathbb{E}[z^{A_\beta}] = p_a + p_\mu z^{s-1} + p_b z^{m+s},$$

with

$$p_a = \frac{d_\beta}{2(s-1)}, \quad p_\mu = 1 - \frac{d_\beta}{2(s-1)} - \frac{d_\beta}{2(m+1)}, \quad p_b = \frac{d_\beta}{2(m+1)}.$$

Notice that $\mathbb{E}[A_\beta] = s - 1$. The resulting discrete queueing system is sometimes referred to as a bulk service queue. Let r_0 be the unique zero of $z^s - \mathbb{E}[z^{A_\beta}]$ with real $z > 1$. For any $\varepsilon > 0$ with $1 + \varepsilon < r_0$,

$$\mathbb{E}[w^{M_\beta}] = \exp\Big(\frac{1}{2\pi i} \oint_{|z|=1+\varepsilon} \ln\Big(\frac{w-z}{1-z}\Big) \frac{(z^s - \mathbb{E}[z^{A_\beta}])'}{z^s - \mathbb{E}[z^{A_\beta}]} \, dz\Big) \tag{2.30}$$

holds when $|w| < 1 + \varepsilon$; see, e.g., [121]. Alternatively,

$$\mathbb{E}[w^{M_\beta}] = \frac{(s - \mathbb{E}[A_\beta])(w-1)}{w^s - A(w)} \prod_{k=1}^{s-1} \frac{w - z_k}{1 - z_k} \tag{2.31}$$

holds for all w, $|w| < r_0$, in which $z_1, \dots, z_{s-1}$ are the $s-1$ zeros of $z^s - \mathbb{E}[z^{A_\beta}]$ in $|z| < 1$. Upon differentiation, (2.30) and (2.31) provide expressions for all cumulants of M_β that are known to allow for accurate numerical evaluation, see [121]. We have then performed, for a wide range of parameters, the following experiment:

1. Fix β, and then choose integers s and m. In this way we create a standard bulk service queue with discrete-valued generic increment A_β.
2. For ranging d_β, calculate $\mathbb{E}[M_\beta]$ using root-finding procedures and (2.31) or using the contour integral (2.30).
3. Calculate
$$\mathbb{E}[M] = \frac{-1}{2\pi i}\int_{\mathscr{C}} \frac{\log(1-(p_a e^{-ua} + p_b e^{-ub} + p_c e^{-uc}))}{u^2} du.$$
4. Check whether $\mathbb{E}[M] = \beta\mathbb{E}[M_\beta]$.

After thoroughly conducting a range of these experiments, it has become clear that the contour integral (2.29) can be computed with remarkable precision.

2.4.3. Random walk lower bounds

The tight upper bounds correspond to worst-case scenarios. We next show how the same MAD approach can identify best-case scenarios and hence tight lower bounds. For each X_i, define a second ambiguity set, which is a subset of $\mathscr{P}_{(\mu,d)}$:

$$\mathscr{P}_{(\mu,d,\beta)} = \left\{\mathbb{P} : \mathbb{P} \in \mathscr{P}_{(\mu,d)},\ \mathbb{P}(X_i \geqslant \mu_i) = \beta_i,\ \forall i\right\}, \tag{2.32}$$

where $\beta_i \in (\frac{d_i}{2(b_i-\mu_i)}, 1 - \frac{d_i}{2(\mu_i-a_i)})$. Hence, for obtaining a lower bound, we include the additional information $\mathbb{P}(X_i \geqslant \mu_i) = \beta_i$ in the ambiguity set. Now, instead of finding the worst-case distribution, we want to identify the best-case distribution and corresponding tight lower bound. The following result is a direct consequence of the general lower bound in [19] on the expectation of a convex function of independent random variables with $\mathscr{P}_{(\mu,d,\beta)}$ ambiguity.

Theorem 2.8. *It holds that*

$$\min_{\mathbb{P}\in\mathscr{P}_{(\mu,d,\beta)}} \mathbb{E}_{\mathbb{P}}[h_n(\mathbf{X})] = \sum_{\boldsymbol{\alpha}\in\{1,2\}^n} h_n(v_{\alpha_1}^{(1)}, \dots, v_{\alpha_n}^{(n)}) \prod_{i=1}^n q_{\alpha_i}^{(i)}, \tag{2.33}$$

where
$$q_1^{(i)} = \beta_i, \quad q_2^{(i)} = 1-\beta_i, \quad v_1^{(i)} = \mu_i + d_i/2\beta_i, \quad v_2^{(i)} = \mu_i - d_i/2(1-\beta_i). \tag{2.34}$$

In Chapter 1, we have presented a novel proof with the primal-dual method. The extension to the multivariate setting follows from the argument used earlier for proving Theorem 2.1.

Again specialize to the i.i.d. setting, and denote by Y the random variable with two-point distribution on values
$$v_1 = \mu + \frac{d}{2\beta}, \quad v_2 = \mu - \frac{d}{2(1-\beta)},$$

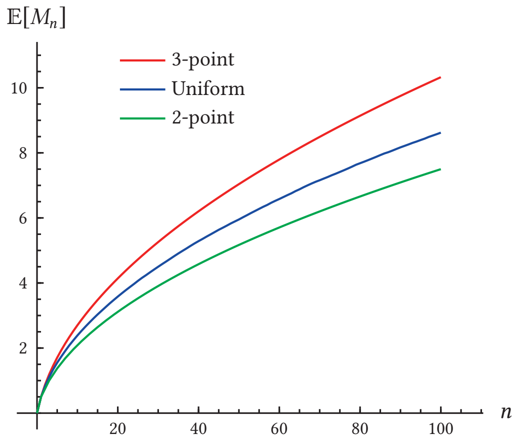


Figure 2.2: Expected random walk maximum $\mathbb{E}[M_n]$ for $U(-b, b)$, where $b = 2$, distributed step sizes with MAD $b/2$ (middle curve, obtained by simulation)

with probabilities β and $1 - \beta$, respectively. Using a similar reasoning as for the upper bound, we obtain for the tight lower bound of $\mathbb{E}[M_n]$ an expression that sums over $O(n^2)$ terms:

$$\sum_{k=1}^{n} \frac{1}{k} \sum_{k_1+k_2=k} \frac{k!}{k_1!k_2!} \beta^{k_1}(1-\beta)^{k_2} \max\{0, k_1 v_1 + k_2 v_2\}. \tag{2.35}$$

The tight lower bound for $c_m(M)$ can be expressed in terms of the integral

$$\frac{(-1)^m\, m!}{2\pi i} \int_{\mathscr{C}} \frac{\log(1-\phi_Y(-u))}{u^{m+1}} \mathrm{d}u, \tag{2.36}$$

where $\phi_Y(s) = \beta \mathrm{e}^{s v_1} + (1-\beta)\mathrm{e}^{s v_2}$, $\mathscr{C}$ is a contour to the left of, and parallel to, the imaginary axis, and to the right of any singularities of $\log(1 - \phi_Y(-u))$ in the left half plane.

We illustrate the lower bound (2.33) (calculated using (2.35)) in Figure 2.2 for the random walk with step size X having a uniform distribution on $[a, b]$. Here we assume a specific distribution just for illustration purposes. The MAD of X can be shown to be $(b - a)/4$. In Figure 2.2 we choose $b = -a = 2$ so that $\mu = 0$ and $d = 1$. The upper and lower bound together provide a tight interval for all possible distributions in the ambiguity set $\mathscr{P}_{(0,1,1/2)}$. Figure 2.3 shows the tight upper bound (2.29) and the lower bound (2.36) for $\mathbb{E}[W]$ with ambiguity set with $\mu = -1$, $d = b/2$ and range $[-b - 2, b]$. The bounds increase with the range and the MAD (which can be shown to hold in general). For a point of reference, we also plot the exact results for one member of the ambiguity set, with the generic increment having a uniform distribution on $[-b - 2, b]$.

2.4.4. Numerical procedures for the GI/G/1 queue

Calculations for $\mathbb{E}[W_n]$ and $c_n(W)$ in the GI/G/1 queue can be performed using similar expressions as for the random walk. Let the random variable $V_{(3)}$ follow a three-point distribution on

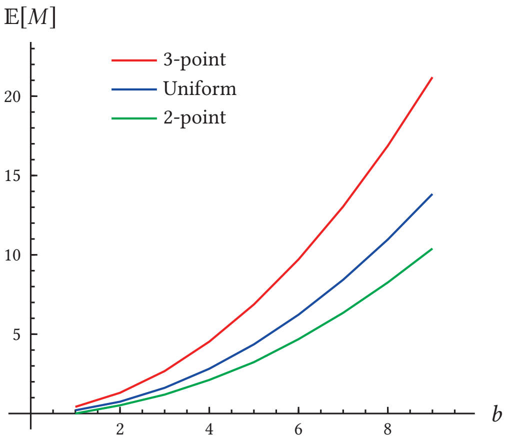


Figure 2.3: Expected all-time maximum $\mathbb{E}[M]$ for $U(-b-2,b)$ and $b \in (1,10)$ (middle curve, obtained by simulation)

values $\{s_1, s_2, s_3\}$ with probabilities

$$p_1 = \frac{d_V}{2(\mu_V - a_V)}, \quad p_2 = 1 - \frac{d_V}{2(\mu_V - a_V)} - \frac{d_V}{2(b_V - \mu_V)}, \quad p_3 = \frac{d_V}{2(b_V - \mu_V)}, \tag{2.37}$$

with $0 \leqslant a_V < \mu_V < b_V$, so that $V_{(3)}$ has mean μ_V and MAD d_V. Similarly, let $U_{(3)}$ have a three-point distribution on values $\{t_1, t_2, t_3\}$ with probabilities

$$r_1 = \frac{d_U}{2(\mu_U - a_U)}, \quad r_2 = 1 - \frac{d_U}{2(\mu_U - a_U)} - \frac{d_U}{2(b_U - \mu_U)}, \quad r_3 = \frac{d_U}{2(b_U - \mu_U)} \tag{2.38}$$

and $0 \leqslant a_U < \mu_U < b_U$, so that $U_{(3)}$ has mean μ_U and MAD d_U.

We then have the representation, see also [49],

$$\mathbb{E}[W_n] = \sum_{k=1}^{n} \frac{1}{k} \sum_{\sum_i k_i = k, \sum_j l_j = k} \max\{0, \sum_{i=1}^{3} k_i s_i - \sum_{j=1}^{3} l_j t_j\} \cdot P(k_1, k_2, k_3) \cdot R(l_1, l_2, l_3) \tag{2.39}$$

with

$$P(k_1, k_2, k_3) = \frac{k!}{k_1! k_2! k_3!} p_1^{k_1} p_2^{k_2} p_3^{k_3}, \quad R(l_1, l_2, l_3) = \frac{k!}{l_1! l_2! l_3!} r_1^{l_1} r_2^{l_2} r_3^{l_3},$$

which requires summing $O(n^5)$ terms.

Let $\phi_{V_{(3)}}(s)$ and $\phi_{U_{(3)}}(s)$ denote the moment generating functions of $V_{(3)}$ and $U_{(3)}$. The tight upper bounds on $c_m(W)$ are given by

$$c_m(W) \leqslant \frac{(-1)^m \, m!}{2\pi i} \int_{\mathscr{C}} \frac{\log(1 - \phi_{V_{(3)}}(-u)\phi_{U_{(3)}}(u))}{u^{m+1}} \mathrm{d}u, \tag{2.40}$$

where $\mathscr{C}$ is a contour to the left of, and parallel to, the imaginary axis, and to the right of any singularities of $\log(1 - \phi_{V_{(3)}}(-u)\phi_{U_{(3)}}(u))$ in the left half plane. Again comparing with extensive simulation, we have found the expression (2.40) accurate and hence suitable for calculating the tight bounds.

2.4.5. Setting the range

Compared to variance, MAD may be more appropriate in case of real-life empirical data that display non-Gaussian features and outliers. Indeed, unlike standard deviation, MAD does not require existence of second moments, and is not so much affected by large deviations from the mean. This feature, however, has major consequences when we let the range $[a,b]$ grow large in which case conditioning on the MAD being d thus allows for distributions with relatively heavy tails. In particular, in the limit $b \to \infty$, this will lead to overly pessimistic scenarios as heavy-tailed distributions with infinite second moments can still have a finite d and hence be member of the ambiguity set. See Section 2.4.6 for more details. While for large but finite b a truly heavy-tailed distribution with infinite second moment is ruled out, the dispersion allowed by the ambiguity set might become too loose for practical purposes.

We now present some guidelines for setting the range, based on the observation that many distributions come with a MAD and standard deviation of comparable size. For the Pearson family of distributions (which includes the gamma and normal distribution) with mean μ and variance σ^2, the MAD d and variance are related as

$$d = 2\alpha\sigma^2 p(\mu) \tag{2.41}$$

with α a constant depending on skewness and kurtosis and $p(\mu)$ the density in μ. For the exponential distribution this relation gives $d = (2/\mathrm{e})\sigma$ and for the normal distribution $d = (\sqrt{2/\pi})\sigma$. Other distributions for which the ratio d/σ is constant include the uniform distribution and discrete distributions such as the Poisson, binomial and negative binomial distribution. In a way similar to constructing confidence intervals in statistical estimation, we then choose to set the range as the mean plus or minus a constant times the MAD:

$$a = \mu - k \cdot d, \quad b = \mu + k \cdot d. \tag{2.42}$$

Here we regard d as the natural scale of deviation, and k as a free parameter that sets the robustness level. So we take the mean and MAD as given, and regard the range as tunable (using common sense or statistical evidence) by the decision maker. We assume that μ and d can be estimated accurately with existing statistical procedures; see e.g. [179]. We should stress that, while intuitive from a probabilistic perspective, the rule (2.42) is only one of many ways to choose the parameters a, b.

Table 2.2 illustrates the results associated with using (2.42) for a setting where we take the M/M/1 queue as the "true" model. The increment X now becomes the difference of two exponential random variables for which we have a closed-form MAD expression in terms of the mean value of X. We consider an instance with unit mean exponential interarrival times and exponential service times with mean ρ, so that the increment X has mean $\mu = \rho - 1$ and MAD $d = \frac{2e^{\rho-1}}{\rho+1}$. We thus have reference values for μ and d, and can investigate the impact of k. The bound grows almost linearly with k, in particular in heavy-traffic scenarios, and this underlines the need for careful selection of the range. While the actual range of the M/M/1 queue spans all real numbers, we see that restricting deviations to twice the MAD ($k = 2$) gives comparable model performance. When reading Table 2.2, keep in mind that the overall goal in this chapter is not to approximate specific models, but rather to come with conservative, robust estimates for an entire class of models that share the same mean-MAD-range properties. In that sense, $k = 2$ is not better than $k = 1.5$ or $k = 2.5$, but rather expresses a different ambiguity assessment or robustness level. The M/M/1 queue serves as a point of reference but is not contained in any of the underlying ambiguity sets, as the range of the step size is unbounded.

Table 2.2: The actual values and bounds of the expected steady-state waiting time $\mathbb{E}[W]$ of the M/M/1 queue with the range $[a, b]$ set through the rule (2.42)

		k					
ρ	$\mathbb{E}[W]$	1.5	1.75	2	2.25	2.5	3
0.1	0.01111	0.10497	0.16434	0.21535	0.25915	0.30116	0.40782
0.5	0.50000	0.56329	0.67919	0.79663	0.91459	1.02840	1.26462
0.6	0.90000	0.86690	1.03323	1.19770	1.36332	1.52804	1.85818
0.7	1.63333	1.41436	1.66589	1.91885	2.17142	2.42373	2.92850
0.8	3.20000	2.57273	3.01339	3.45454	3.89573	4.33672	5.21866
0.9	8.10000	6.21057	7.25250	8.29428	9.33642	10.37811	12.46184
0.99	98.01000	73.55537	85.81540	98.07540	110.33542	122.59543	147.11548

2.4.6. Degenerate behavior for infinite range

The variance of $X_{(3)}$ is $\frac{d}{2}(b - a)$, the maximal variance for distributions in the ambiguity set $\mathscr{P}_{(\mu,d)}$. Hence, for fixed d, the variance becomes unbounded when $b \to \infty$. As a consequence, this results in fairly crude bounds:

Proposition 2.9. *As $b \to \infty$, the bound $\max_{\mathbb{P} \in \mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}[f_n(\mathbf{X})]$ converges to*

$$n \cdot \frac{d}{2} + \sum_{k=1}^{n} \frac{1}{k} \sum_{k_1+k_2=k} \max\{0, k_1 a + k_2 \mu\} \cdot \frac{k!}{k_1! k_2!} p_1^{k_1} p_2^{k_2} \tag{2.43}$$

with $p_1 = \frac{d}{2(\mu-a)}$ and $p_2 = 1 - \frac{d}{2(\mu-a)}$.

Proof. Split the inner summation in (2.25) into three parts. First consider the summation over $\sum_i k_i = k : k_3 \geqslant 2$, i.e. those outcomes where the value b occurs multiple times. Taking the limit $b \to \infty$ inside of the summation and recognizing that the probability mass on this point is of order $O(1/b^{k_3})$ gives

$$\lim_{b\to\infty} \sum_{\sum_i k_i = k\,:\,k_3\geqslant 2} \frac{d^{k_3}\max\{0, k_1 a + k_2\mu + k_3 b\}}{2^{k_3}(b-\mu)^{k_3}} \cdot \frac{k!}{k_1!k_2!k_3!} p_1^{k_1} p_2^{k_2} = 0. \tag{2.44}$$

Next consider $\sum_i k_i = k \,:\, k_3 = 1$, denoting the instances for which b is attained precisely once. Taking the limit $b \to \infty$ inside the sum and using that b occurs with probability $O(1/b)$ results in

$$\lim_{b\to\infty} \sum_{\sum_i k_i = k\,:\,k_3=1} \frac{d\max\{0, k_1 a + k_2\mu + b\}}{2(b-\mu)} \cdot \frac{k!}{k_1!k_2!} p_1^{k_1} p_2^{k_2} = k\cdot\frac{d}{2} \sum_{\sum_i k_1+k_2=k-1} \frac{(k-1)!}{k_1!k_2!} p_1^{k_1} p_2^{k_2} = k\cdot\frac{d}{2}. \tag{2.45}$$

The third part is then $\sum_i k_i = k \,:\, k_3 = 0$, representing the realizations without the point b. As $b \to \infty$, we get

$$\lim_{b\to\infty} \sum_{\sum_i k_i = k\,:\,k_3=0} \max\{0, k_1 a + k_2\mu\} \cdot \frac{k!}{k_1!k_2!} p_1^{k_1} p_2^{k_2} = \sum_{k_1+k_2=k} \max\{0, k_1 a + k_2\mu\} \cdot \frac{k!}{k_1!k_2!} p_1^{k_1} p_2^{k_2}. \tag{2.46}$$

This completes the proof. □

Proposition 2.9 suggests that large running maxima are likely due to a single large step. The feature is caused by heavy-tailed distributions, and in queueing theory dubbed the single big jump principle (see, e.g., [81]). This dominance of one step sharply contrasts intuition for light-tailed distributions, where typically all steps together lead to large sums or maxima. The bound (2.43) for $\mathbb{E}[M_n]$ grows to infinity as $n \to \infty$, rendering the bound useless for the expected all-time maximum $\mathbb{E}[M]$. This is indeed anticipated, and can be understood as follows. Define a sequence of random walks indexed by b with the extremal three-point distribution. Consider the limiting all-time maximum M as $b \to \infty$. Assume that the random walk has negative drift (i.e., $\mathbb{E}[X] < 0$). Then the associated sequence of distributions of $M = M_{(b)}$ will converge to a proper limit $M_{(\infty)}$. However, as $\lim_{b\to\infty} \mathscr{P}_{(\mu,d)}$ contains distributions with infinite second moment, [7, Theorem X.2.1] says that $\mathbb{E}[M_{(\infty)}]$ will be infinite.

2.4.7. Data-driven setting

In applications, one may only have a limited number n of observed interarrival and service times. We consider this realistic setting where knowledge of the stochastic nature is restricted to a set of samples generated independently and randomly according to an unknown distribution $\mathbb{P}$. To apply the mean-MAD framework in this context, we need to construct the ambiguity set that is supposed to contain this unknown $\mathbb{P}$. We will show that we can efficiently estimate the mean, MAD and β, and hence compute robust bounds that are useful in realistic settings.

Let $\mu_n^{(V)}$, $d_n^{(V)}$ and $\beta_n^{(V)}$ denote the consistent estimators of μ_V, d_V and $\beta_V = \mathbb{P}(V \geqslant \mu_V)$, respectively, based on n observed service times $v_1, ..., v_n$, and defined as $\mu_n^{(V)} = \bar{v} = \frac{1}{n}\sum_{i=1}^n v_i$, $d_n^{(V)} = \frac{1}{n}\sum_{i=1}^n |v_i - \bar{v}|$ and $\beta_n^{(V)} = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{[\bar{v},\infty)}(v_i)$. We define similar estimators based on n observed interarrival times. Next, we demonstrate the mean-MAD bounds in this data-driven setting. Since statistical accuracy of the estimators increases with the number of samples, we expect the bounds to converge as n increases.

We have performed extensive simulations to investigate the error between the estimated and true bounds for several values of the sample size n. We generate 1,000 sample paths of sample size 10,000 and compute the corresponding mean relative error. Table 2.3 displays the mean absolute percentage error (MAPE) for both the upper and lower bound estimates, where the interarrival time is $U(0, 10)$ distributed and we differentiate between a 50% and 90% utilization level. Sample paths resulting in instable systems were removed. The lower bound is slightly harder to estimate than the upper bound. Indeed, the lower bound requires estimating the additional parameters β_V and β_U. Also observe that the relative error increases with the system utilization.

Table 2.3: MAPE of the bound estimates for $n = 150, 200, 500, 1000, 2000, 5000, 10000$

		MAPE with sample size n						
Service times	Bound	150	200	500	1000	2000	5000	10000
$U(0, 5)$	UB	15.44%	13.22%	8.31%	5.84%	4.28%	2.72%	1.89%
	LB	25.51%	22.30%	13.86%	9.75%	7.08%	4.53%	3.16%
$U(0, 9)$	UB	33.35%	30.93%	21.93%	16.35%	13.29%	8.92%	6.41%
	LB	36.27%	35.01%	28.72%	22.30%	17.35%	10.77%	7.58%

To further highlight the role of system utilization, we perform a similar data-driven experiment, but now with ground truth a single trace of n customers in an M/M/1 queue. The U_i are sampled from a unit mean exponential distribution, and the V_i are exponentially distributed with mean ρ. The results are shown in Figure 2.4. Indeed, as ρ increases, more observations are required for accurate parameter estimates and hence accurate bounds. Observe that convergence settles in quickly for low-utilization regimes. Taken together, we conclude that the robust bounds are useful for realistic data-driven settings that require statistical estimation of the summary statistics such as the mean and MAD.

2.5. Conclusions and outlook

This chapter explains why MAD simplifies comparable variance-based optimization problems, in a way that is almost unreasonably effective, resulting in a full solution to the extremal queue problem with mean-MAD-range information. When full distributional information is not available, or simply judged too detailed, the mean, MAD and range together form partial yet sufficient information for obtaining robust bounds on the steady-state waiting-time moments in the GI/G/1 queue. Through basic statistical estimation of this partial information the GI/G/1 queue becomes a data-driven model that adjusts to available training data, for which we present tight performance guarantees. While the bounds we have obtained for mean-MAD information are

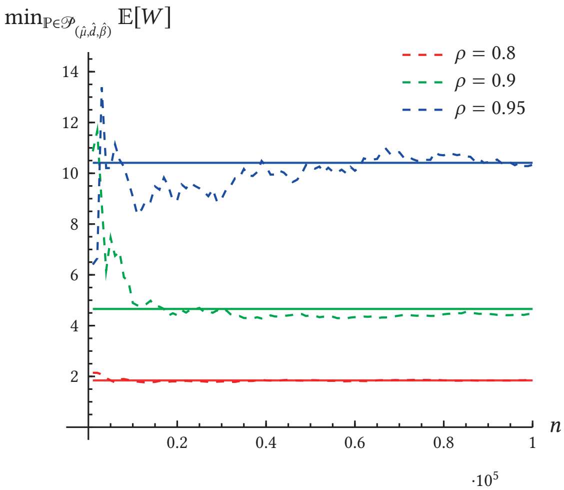


Figure 2.4: Estimation of the mean-MAD ambiguity lower bound for the M/M/1 queue, with the dashed lines corresponding to estimates of the tight bounds

the best possible, it is worth exploring the potential improvements that can be made using more distributional information, such as higher-order moments and other summary statistics. However, it is currently unclear how our approach can be adapted to incorporate this information, so further research in this direction will be necessary.

Compared to the classical extremal queue problem, the key idea in this chapter was to use MAD instead of variance as dispersion measure. This idea is likely applicable to other queueing systems. Examples are the multi-server GI/G/c queue, with a cyclic allocation rule, and tandem queues. Indeed, some of the key performance measures for such systems are expectations of functions that are convex in the random variables (see, e.g., [192]), and therefore the mean-MAD approach can be used. The MAD approach is of interest beyond queueing theory, because the search for extremal distributions of convex functions is relevant in many other settings. Moreover, whenever a performance measure can be viewed as a convex function of i.i.d. random variables with mean-MAD ambiguity (e.g., nested max-operators in production systems; see [37, 89]), our approach will identify the extremal distribution and tight bounds.

The crux of the MAD approach consists of the explicitly solvable dual LP described in Section 2. A simple argument then showed that this solution is independent of the precise objective function (in this chapter describing waiting-time moments of the GI/G/1 queue). Hence, the MAD approach is a generic, computationally tractable way to analyze stochastic processes, such as random walks and queues.

3

Second-order bounds for the M/M/s queue with random arrival rate

3.1. Introduction

The majority of stochastic models in queueing theory assume known arrival processes, which facilitates exact analysis. In fact, the dominant assumption is that potential customers arrive according to a Poisson process with a known intensity. In contrast, we interpret the arrival rate as an unknown parameter of which partial information is available. The arrival rate parameter is then replaced by a random variable with some distribution representing what is known about the market size. We primarily focus on the situation where we know the mean and variance of this market size, with the connection to classical queueing theory when the variance is zero and, hence, the random variable is known to be identical to the arrival rate.

The technical challenge is to solve the maximization problem, which requires to determine tight bounds for the expected wait for all distributions of the random arrival rate that comply with the partial information. A tight upper bound on the expected wait presents for the rational choice model the worst-possible scenario, and yields the corresponding tight lower bound on the market share. For mean-variance information, we will show that these tight bounds are attained by two-point distributions, essentially saying that the worst-case market is one where individuals assume the actual market size attains its maximum size or is well below the expected

This chapter is based on the research paper [212].

market size. To establish such tight bounds, we invoke primal-dual techniques for solving semi-infinite linear programs. Successfully applying these techniques will prove to depend crucially on the properties of the expected wait viewed as function of the arrival rate. For the main queueing system in this chapter, the M/M/s queue, we shall leverage that the expected wait is (i) increasing and (ii) convex in the arrival rate, with (iii) the first derivative with respect to the arrival rate being a convex function. All three properties turn out to be crucial for solving the optimization problems in this chapter, and the convex derivative will be particularly useful for the setting with mean-variance information.

This research connects four themes in the queueing literature: random arrival rates, second-order bounds, parametric convexity and rational queueing.

Queues with random arrival rate. Due to its mathematical tractability, the M/M/s queue is among the cornerstones of queueing theory and is widely applied in service operations management. To enhance practical relevance, several studies proposed to relax the assumption of Poisson arrivals, and instead model the arrival rate as a random variable [9, 44, 47, 107, 111, 204, 223, 233]. This gives rise to a Poisson mixture model for the arrival process, which can deal with forecast errors, overdispersion (compared with the natural fluctuations of a Poisson process) and unknown market size. The mathematical formulae of the M/M/s model, such as the Erlang-C formula for the delay probability, then still apply once the mixing distribution can be characterized. Jongbloed and Koole [125] showed how to estimate this mixing distribution based on available data of arrivals.

Second-order bounds. For this M/M/s queue with random arrival rate, we derive performance bounds that do not require the full mixture distribution, but only utilize the first two moments of the arrival rate. To derive such second-order bounds, we need to solve a semi-infinite linear program. To do so, we first show that the expected stationary waiting time can be written as the expectation of a function of the random arrival rate that has a convex derivative. We then establish a general result, which provides tight second-order bounds as the solutions to semi-infinite linear programs for the expectation of functions with convex derivatives. Such semi-infinite linear programs arise naturally in studying second-order bounds, or more generally, moment problems. For example, [32, 66, 70] use second-moment information to compute tight bounds for distributionally robust stochastic programming. Second-order bounds also play a predominant role in the theory on performance bounds for queues, which started with the work of Kingman [132]. Kingman derived bounds for the mean waiting time in the GI/G/1 queue, expressed in terms of the first two moments of the interarrival- and service-time distributions; however, these bounds are not in general tight.

Convexity and beyond. Numerous studies have taken advantage of the insight that moments of the waiting time can be viewed as the expected value of a convex function. Vasicek [216] showed that the variance of waiting time under a general queueing discipline does not exceed that under the LCFS discipline, and Hajek [96] established that among all arrival processes for an exponential server queue with specified arrival and service rates, the arrival process which minimizes the expected waiting time is the process with constant interarrival times. Weber [219] revealed that the mean queueing time in the G/GI/m queue is nonincreasing and convex in the number of servers, m, indicating that marginal analysis is optimal for determining server allocation among various service facilities so as to minimize the total expected queueing time. The proofs in [96, 216, 219] rely on convexity properties. Convexity of queueing performance metrics is a highly desired property in the parametric optimization of stochastic systems. A substantial body of research employs direct algebraic methods to establish convexity of the steady-state performance metrics as functions of the design parameters; see, for example, [91, 103, 104, 106, 119, 142, 220]. For a comprehensive discussion on the related notion of stochastic convexity, we refer the interested reader to [192, 193, 195]. However, as mentioned earlier, for the present study we require convexity of the objective function's derivative. Several closely related concepts emerge in the extremal analysis of queueing systems. In this context, the need for a convex derivative is replaced by closely related sufficient conditions that warrant the application of the theory of complete Tchebychev systems [128]. For a detailed discussion of the relevance of these systems to queues, please refer to [73, 94].

Rational queueing with limited information. A mature literature exists [108] that seeks to elucidate the rational decision-making processes of users who decide whether or not to access delay-sensitive services based on utility. Naor [162] initiated this line of work for the observable M/M/1 queue with a customer utility that depends linearly on the price, the value of service and the expected waiting time. Edelson and Hilderbrand [74] investigated similar questions for the unobservable M/M/1 queue. These works explore the environment with a deterministic arrival rate. Liu and Hasenbein [143] expanded upon Naor's model by assuming that the arrival rate is drawn from a known probability distribution accessible to the decision maker. This concept was further extended to the unobservable setting in [48]. Their research demonstrated that a socially optimal pricing strategy results in a lower expected arrival rate compared to a price set by a revenue-maximizing decision maker. Hassin et al. [109] examined the unobservable model with a random arrival rate from a distinct angle, where strategic customers base their decisions on a rate-biased time-average property when estimating their expected waiting time. Wang et al. [218] extended the traditional observable model to a distributionally robust setting by taking into account an uncertain arrival rate governed by an unknown underlying probability distribution.

3.1.1. Contributions and outline

We summarize the main contributions of this chapter as follows:

1. For the M/M/s queue with partially known arrival rate we establish novel tight bounds. When mean-variance-support is known, we show that the tight bound on the expected wait is attained by a two-point distribution. We also present similar results for the environment in which the service rate is partially known, or when the dispersion measure is changed from variance to mean absolute deviation or semi-variance.

2. Our proof of these tight bounds combines two results from disparate areas. The first result stems from optimization and says the semi-infinite linear program that describes the tight bound for the expectation of a convex function with a convex derivative is attained by a

two-point distribution. The second result stems from queueing theory and shows that the expected wait as a function of the arrival rate is convex, and we show that its derivative is also a convex function of the arrival rate. Combining these two results provides an effective strategy for finding distributionally robust bounds for queueing systems subject to parametric uncertainty.

3. The bounds are leveraged for analysis of non-observable M/M/s queues that cater to rational users who decide to join or balk based on expected utility. We use the wait bounds to bound the equilibrium arrival rate, which in turn leads to tractable maximin analyses for setting the revenue-maximizing price. In this way, we extend the classical strategic queueing literature with known model parameters as in [162] and [74] to a setting where the arrival rate is only partially characterized in terms of range, mean and variance.

4. We further introduce a broader framework for the methods presented in this chapter, enabling their application to other queueing systems that are subject to parametric uncertainty. We additionally propose to incorporate additional types of information which, in the context of steady-state performance metrics, yield tight bounds under parametric uncertainty. In all cases, the mathematical tractability of the distributionally robust queueing models hinges on the "curvature" properties of the steady-state performance metrics with respect to the uncertain parameters governing these models.

The chapter is structured as follows. Section 3.2 contains the proof of our main result, which provides bounds for the expected wait under uncertain arrival rates. Section 3.3 applies these bounds to a rational queueing model. In Section 3.4, we explore several information sets that complement mean-variance information. In Section 3.5, we examine other types of queueing problems. Finally, in Section 3.6, we conclude and propose future research directions.

3.2. Tight bounds for expected wait with limited market knowledge

In this section, we derive bounds for the expected wait with an uncertain arrival rate. Section 3.2.1 introduces the central problem. Section 3.2.2 derives the second-order bounds for general convex functions with a convex derivative. In Section 3.2.3, we apply these second-order bounds to the expected wait in the M/M/s queue.

3.2.1. The model

Consider an M/M/s queue with Poisson arrivals with rate λ, unit mean exponential service requirements and s parallel servers, each operating with service rate μ. Let $W(s,\lambda)$ denote the total system time experienced by a customer in this queueing system, which depends on the number of servers s and the arrival rate λ. We shall refer to $W(s,\lambda)$ as the expected wait. For notational convenience, we omit at first the functional dependence on the service rate parameter μ, but our findings remain valid for any given value of this parameter. Suppose that $\lambda/\mu < s$.

Then the single-server M/M/1 queue has expected wait $W(1,\lambda) = (\mu - \lambda)^{-1}$. Observe that

$$\frac{\partial}{\partial\lambda}W(1,\lambda) = \frac{1}{(\mu-\lambda)^2} > 0, \quad \frac{\partial^2}{\partial\lambda^2}W(1,\lambda) = \frac{2}{(\mu-\lambda)^3} > 0$$

and

$$\frac{\partial^3}{\partial\lambda^3}W(1,\lambda) = \frac{6}{(\mu-\lambda)^4} > 0, \quad \text{for } \lambda < \mu.$$

Hence, $\lambda \mapsto W(1,\lambda)$ is an increasing, convex function of λ with a convex derivative. We now present some general results for functions ϕ that have the same properties as $W(1,\cdot)$. We present the results in general form, because we will apply these results later to $W(s,\lambda)$ with $s \geqslant 2$ as well, to obtain the solutions for

$$\max_{\mathbb{P}\in\mathscr{P}_{(m,\sigma)}} \mathbb{E}_{\mathbb{P}}[W(s,\Lambda)] \tag{3.1}$$

in which we focus on the case $\mathscr{P} = \mathscr{P}_{(m,\sigma)}$ with the latter shorthand notation for $\mathscr{P}(m,\sigma,\underline{x},\overline{x})$, the set of all distributions with mean m, standard deviation σ and the support contained in the interval $[\underline{\lambda},\overline{\lambda}]$. That is,

$$\mathscr{P}(m,\sigma,\underline{\lambda},\overline{\lambda}) := \left\{\mathbb{P} \in \mathscr{P}_0([\underline{\lambda},\overline{\lambda}]) \,:\, \mathbb{E}_{\mathbb{P}}[\Lambda] = m,\ \mathbb{E}_{\mathbb{P}}[\Lambda^2] = \sigma^2 + m^2\right\}$$

where $\mathscr{P}_0([\underline{\lambda},\overline{\lambda}])$ comprises all probability distributions with support contained in $[\underline{\lambda},\overline{\lambda}]$. We shall refer to such a set as an ambiguity set. Henceforth, the market size is viewed as a random variable with a distribution contained in $\mathscr{P}(m,\sigma,\underline{\lambda},\overline{\lambda})$. We further assume from this point onward that the system under consideration is stable for each realization of the random arrival rate; that is, $\overline{\lambda} < s\mu$.

3.2.2. Bounding the expectation of a convex function with convex derivative

We first demonstrate how (3.1) can be written in terms of a moment problem. To emphasize the generality of these results, we develop bounds for $\int_x \phi(x)\mathrm{d}\mathbb{P}(x)$, where $\phi(\cdot)$ represents a function of a random variable X possessing specific "curvature" properties, shared by $W(s,\cdot)$. We next solve $\max_{\mathbb{P}\in\mathscr{P}_{(m,\sigma)}} \mathbb{E}_{\mathbb{P}}[\phi(X)]$ where the constraints that define $\mathscr{P}_{(m,\sigma)}$ correspond to known first and second moments and support contained in $[\underline{x},\overline{x}]$. In other words, we wish to find

$$\begin{aligned} \max_{\mathbb{P}\in\mathscr{P}_0([\underline{x},\overline{x}])} \quad & \int_x \phi(x)\mathrm{d}\mathbb{P}(x) \\ \text{s.t.} \quad & \int_x x\mathrm{d}\mathbb{P}(x) = m,\ \int_x x^2\mathrm{d}\mathbb{P}(x) = (\sigma^2 + m^2), \end{aligned}$$

which constitutes a semi-infinite linear program with an (infinite-dimensional) decision variable, the probability distribution $\mathbb{P}$, which must reside within $\mathscr{P}_{(m,\sigma)}$. In this formulation, it often proves more manageable to solve the corresponding dual problem, which is also a semi-infinite linear program, but with an infinite number of constraints instead of an infinite-dimensional decision object. For (3.1), this dual problem involves finding π_0, π_1, π_2 such that

$$\pi_0 + \pi_1 x + \pi_2 x^2 \geqslant \phi(x), \quad \forall x \in [\underline{x}, \overline{x}],$$

and $\pi_0 + \pi_1 m + \pi_2(\sigma^2 + m^2)$ is minimized. The following result, attributable to Birge and Dulá [32], offers a solution to (3.1) under these curvature conditions for ϕ. The proof relies heavily on duality theory for moment problems:

LEMMA 3.1 **(Theorem 5.1, [32]).** *Let $x \mapsto \phi(x)$ be a continuously differentiable and convex function on the interval $[\underline{x}, \overline{x}]$ with its derivative $\phi'(x)$ being convex on $[\underline{x}, c]$ and concave on $[c, \overline{x}]$ for $\underline{x} \leqslant c \leqslant \overline{x}$. For a random variable X with distribution $\mathbb{P} \in \mathscr{P}(m, \sigma, \underline{x}, \overline{x})$, the tight upper bound for $\mathbb{E}_{\mathbb{P}}[\phi(X)]$ is attained by a discrete distribution with at most two support points.*

As a consequence of Lemma 3.1, we can restrict attention to two-point distributions in the search for the extremal distribution that attains the tight bound $\mathbb{E}_{\mathbb{P}}[\phi(X)]$. A similar result was already applied by Popescu [178] to the setting in which the support of the uncertain parameter is unbounded, i.e., $\mathrm{supp}(X) = \mathbb{R}$.

LEMMA 3.2 **(Extremal two-point distributions).** *Consider a function $x \mapsto \phi(x)$ that is continuously differentiable and convex on $[\underline{x}, \overline{x}]$, and let X be a random variable with distribution $\mathbb{P} \in \mathscr{P}(m, \sigma, \underline{x}, \overline{x})$.*
(i) *If the derivative $\phi'(x)$ is convex on $[\underline{x}, \overline{x}]$, the tight upper bound for $\mathbb{E}_{\mathbb{P}}[\phi(X)]$ is attained by a two-point distribution with values $\{m - \frac{\sigma^2}{\overline{x}-m}, \overline{x}\}$ and corresponding probabilities $\{\frac{(\overline{x}-m)^2}{(\overline{x}-m)^2+\sigma^2}, \frac{\sigma^2}{(\overline{x}-m)^2+\sigma^2}\}$.*
(ii) *If the derivative $\phi'(x)$ is concave on $[\underline{x}, \overline{x}]$, the tight upper bound for $\mathbb{E}_{\mathbb{P}}[\phi(X)]$ is attained by a two-point distribution with values $\{\underline{x}, m + \frac{\sigma^2}{m-\underline{x}}\}$ and corresponding probabilities $\{\frac{\sigma^2}{(m-\underline{x})^2+\sigma^2}, \frac{(m-\underline{x})^2}{(m-\underline{x})^2+\sigma^2}\}$.*

Proof. A two-point distribution with mean m and variance σ^2 has support

$$x_1 = m - \sqrt{\frac{\alpha}{1-\alpha}}\sigma, \quad x_2 = m + \sqrt{\frac{1-\alpha}{\alpha}}\sigma \tag{3.2}$$

with probabilities $\alpha,\ 1-\alpha$, respectively. We thus need to solve $\max_\alpha \Phi(\alpha)$ with

$$\Phi(\alpha) := \alpha\phi\left(m + \sqrt{\frac{1-\alpha}{\alpha}}\sigma\right) + (1-\alpha)\phi\left(m - \sqrt{\frac{\alpha}{1-\alpha}}\sigma\right)$$

and

$$\alpha \in \left[\frac{\sigma^2}{(\overline{x}-m)^2+\sigma^2}, \frac{(m-\underline{x})^2}{(m-\underline{x})^2+\sigma^2}\right]$$

due to the support being contained in $[\underline{x}, \overline{x}]$. To show that $\alpha^* = \frac{\sigma^2}{(\overline{x}-m)^2+\sigma^2}$ is a maximizer, we will prove that $\Phi(\alpha)$ is nonincreasing in α. For differentiable ϕ, we have that

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\Phi(\alpha) = \phi(x_2) - \phi(x_1) - (x_2 - x_1)\frac{\phi'(x_1) + \phi'(x_2)}{2}. \tag{3.3}$$

In order for $\Phi(\alpha)$ to be nonincreasing, we need

$$\phi(x_2) - \phi(x_1) - (x_2 - x_1)\frac{\phi'(x_2) + \phi'(x_1)}{2} \leqslant 0,$$

or

$$\frac{\phi'(x_2) + \phi'(x_1)}{2} \geqslant \frac{\phi(x_2) - \phi(x_1)}{x_2 - x_1}.$$

This inequality holds by convexity of ϕ'. Thus, as $\Phi(\alpha)$ is nonincreasing in α, the maximum occurs at the lower bound of the feasible set (i.e., $\alpha^* = \frac{\sigma^2}{(\bar{x}-m)^2+\sigma^2}$). Substituting the optimal value α^* into our parameterized expression for the two-point distribution then yields assertion (i).

For assertion (ii), note that for concave ϕ', $\Phi(\alpha)$ is nondecreasing in α and therefore maximized by $\alpha^* = \frac{(m-\underline{x})^2}{(m-\underline{x})^2+\sigma^2}$. □

Returning to the expected wait in the M/M/1 queue, we see that the tight bound for $\mathbb{E}_{\mathbb{P}}[W(1,\Lambda)]$ is attained by the two-point distribution in Lemma 3.2(ii). We shall prove this result for the multi-server system in the next subsection.

3.2.3. Bounding the expected wait as function of the arrival rate

To apply the results in the previous section to the multi-server M/M/s queue with random arrival rate, we must show that the expected wait, as a function of the arrival rate, has a convex derivative. Therefore, we need to study the third derivative of $W(s,\lambda)$ with respect to λ. The expected wait is a classic formula in the queueing literature, for which a host of properties have been discovered, including convexity with respect to λ; see, e.g., [91] and [142]. But to the best of our knowledge, the property of a convex derivative with respect to λ has not been reported. However, this property was reported in [182] for the expected queue size defined as

$$L(s,\lambda) := s\rho + \frac{\rho}{1-\rho}C(s,\lambda), \tag{3.4}$$

where $C(s,\lambda)$ is the Erlang-C formula, the probability that an arriving customer experiences delay in the M/M/s queue and $\rho = \lambda/(s\mu)$. [182] proves that $L(s,\lambda)$ as function of the arrival rate λ, indeed has a convex derivative. By Little's law, $W(s,\lambda) = L(s,\lambda)/\lambda$, and one could try to see whether the convex derivative of $L(s,\lambda)$ implies the same property for $W(s,\lambda)$. We do not see a direct proof for this argument, in which we can use convexity of the derivative of the expected queue length L to demonstrate convexity of the derivative of the expected wait W. We therefore choose to provide a standalone proof for the fact that $W(s,\lambda)$ has a convex derivative, just like $L(s,\lambda)$. The proof method we apply is largely inspired by that in [182]. In the following, we provide a proof sketch. The details are relegated to the Appendix. First, we determine an expression for the third derivative of the expected wait as a function of the server utilization ρ. Then, we write this expression as a function of the Erlang-C formula C and the number of servers s. It then suffices to show that this expression is nonnegative for all values of s and $\forall \rho \in (0,1)$, which implies convexity of $W(s,\lambda)$ as a function of λ with the number of servers s fixed.

LEMMA 3.3 **(Derivative w.r.t. λ is convex).** *The expected wait $W(s,\lambda)$ as function of $\lambda \in (0,\mu)$ has a convex derivative.*

Proof. See Appendix B.1. □

Convexity of $W(s,\cdot)$ as a function of λ implies that the extremal distribution of Λ is given by the extremal distribution in Lemma 3.2(i). This then yields our main result.

THEOREM 3.4 **(Second-order bound for random arrival rate).** *Consider an M/M/s queue with random arrival rate Λ that follows a distribution $\mathbb{P}$ belonging to the ambiguity set $\mathscr{P}_{(m,\sigma)}$. The tight upper bound for the expected wait $\mathbb{E}_{\mathbb{P}}[W(s,\Lambda)]$ is attained by the two-point distribution with support*

$$\lambda_1 = m - \frac{\sigma^2}{\overline{\lambda} - m}, \quad \lambda_2 = \overline{\lambda},$$

and corresponding probabilities

$$p_1 = \frac{(\overline{\lambda} - m)^2}{(\overline{\lambda} - m)^2 + \sigma^2}, \quad p_2 = \frac{\sigma^2}{(\overline{\lambda} - m)^2 + \sigma^2}.$$

We demonstrate the effectiveness of our novel second-order bounds through a numerical experiment. The results are presented in Table 3.1. We assume that the "true" distribution of the random arrival rate follows a beta distribution with support $[0,\overline{\lambda}]$ and standard deviation fixed to $\sigma = 0.1$. We consider the scenario with $s = 2$ servers and vary the mean m to obtain results for multiple utilization levels. The average utilization level is denoted by $\rho = m/(s\mu)$. As shown in Table 3.1, the second-order bounds proposed in Theorem 3.4 perform well for low- and medium-utilization regimes. However, for high utilization, the bounds deviate significantly from the true value when assuming a beta distribution for Λ. Furthermore, the bounds widen as the upper bound of the support $\overline{\lambda}$ increases. The second column corresponds to the second-order lower bound, which can be obtained using the extremal distribution in Lemma 3.2(ii). This holds because determining the lower bound is tantamount to maximizing $\mathbb{E}_{\mathbb{P}}[-\phi(X)]$. Notice that this bound depends on $\underline{\lambda}$, the lower bound of the support, rather than $\overline{\lambda}$.

Table 3.1: Numerical bounds for the expected wait in the M/M/2 queue with random arrival rate following a beta distribution with support $[0,\overline{\lambda}]$ and $\sigma = 0.1$

		$\overline{\lambda} = 1.925$		$\overline{\lambda} = 1.95$		$\overline{\lambda} = 1.99$	
ρ	LB	$\mathbb{E}[W(2,\Lambda)]$	UB	$\mathbb{E}[W(2,\Lambda)]$	UB	$\mathbb{E}[W(2,\Lambda)]$	UB
0.2	1.0445	1.0449	1.0940	1.0449	1.1199	1.0449	1.4312
0.5	1.3389	1.3440	1.4655	1.3440	1.5315	1.3440	2.3236
0.7	1.9753	2.0094	2.3200	2.0096	2.5002	2.0098	4.6625
0.8	2.8099	2.9442	3.5518	2.9465	3.9446	2.9497	8.6515
0.9	5.3921	6.4285	7.6443	6.6019	9.0133	6.9387	25.0550

3.3. Rational queueing model

We apply in this section the results derived in the previous section to rational queueing models. Section 3.3.1 introduces the rational queueing model with a known market size. In Section 3.3.2, we extend this model to the setting in which customers are ambiguity-averse about the total market size.

3.3.1. Model with known market size

Consider a firm that sells delay-prone services to a market of rational delay-sensitive individuals. Individuals value service, but dislike waiting, and will only join when the net service value exceeds the wait costs. They cannot observe real-time queues, and instead estimate their expected wait costs based on beliefs or information about the total arrival rate of all potential customers. The arrival rate or market size is measured in terms of the scale parameter λ, so that the expected time between arrivals of consecutive individuals is $1/\lambda$. All individual joining decisions together give an equilibrium arrival rate or market share λ_e, which could be viewed as the product of λ times the probability that an individual decides to join. We then consider the firm as a price-setting monopolist seeking to maximize revenue. With p the price for service, the firm can influence the net service value of individuals, and hence the joining probability. A low price yields small margins but high joining probability, while a high price increases the profit per customer but suppresses the market share. The firm should thus strike the optimal balance between profit per customer and market share. Such challenges have been thoroughly addressed in the rational queueing literature [108].

Assume that the firm operates according to an M/M/s queue, but that the arrival process consists of individuals that decide to join-or-not based on a rational-choice model. An individual's utility in acquiring service is defined as

$$U = r - p - cW(s, \lambda_e)$$

with r the value of service, p the price for service, c the wait costs per time unit, and λ_e the effective arrival rate of joining individuals. Here we assume that individuals cannot observe real-time queue lengths, but instead know how the expected wait varies with overall demand. Hence, the rational choices are based on a linear relation between net service value and expected wait costs, and an individual will only join when $U \geqslant 0$. This is a common set-up in the rational queueing literature with self-interested individuals that tend to overcrowd systems, because they ignore externalities—inconvenient side effects—that their decisions have on others. When an individual decides to join, this will increase congestion and wait for all customers. Standard game theory then predicts that all individual decisions together will lead to an equilibrium, expressed in the equilibrium arrival rate, the probability $q(p)$ that an arbitrary individual joins the system multiplied by the market size λ. Since not joining the queue gives zero utility, the equilibrium joining probability $q(p)$ then solves $U = r - p - cW(s, q(p)\lambda) = 0$. The arrival process of those who decide to join is then a Poisson process with rate $q(p)\lambda$. To optimize revenue, the firm needs to set the price that strikes the best balance between margin p and market share $\lambda_e(p) = q(p)\lambda$. Notice that the joining probability $q(p)$ is an implicit function that decreases with p and can only be determined as the solution to $U = 0$ when individuals know the arrival rate parameter λ. The firm then effectively solves the monopoly pricing problem, $\max_p p \cdot q(p) \cdot \lambda$, to attain maximal revenue. The optimal solution to this problem for a given market size is already well established [108]. Instead, in the next subsection, we choose to view the market size as a random variable of which only limited information is available.

3.3.2. Model with ambiguous market size

We interpret the market size as an unknown parameter about which the individuals that constitute the market form beliefs. The parameter λ is then replaced by a random variable Λ with a distribution representing what individuals believe or know about the market size. We primarily focus on the situation where individuals only know the support, mean and variance of Λ. From the individual's perspective, variance of Λ expresses uncertainty about the market size, and hence uncertainty about the expected waiting. For this setting with partially-informed customers, or an ambiguously specified arrival process, we consider the revenue-maximizing firm as a maximin decision maker, first determining the worst-case market (a minimization problem), and then maximizing the revenue by selecting the best price for this worst-case market. The firm thus attempts to solve the maximin problem

$$\max_p \min_{\mathbb{P}\in\mathscr{P}_{(m,\sigma)}} \mathbb{E}_{\mathbb{P}}[p \cdot \lambda_e(p)] \tag{3.5}$$

with $\mathbb{P}$ the distribution of the market size Λ and $\mathscr{P}_{(m,\sigma)}$ the ambiguity set that contains all distributions that satisfy the partial information, given by the mean, variance and support. Minimax and maximin optimization problems such as (3.5) arise naturally in decision making under uncertainty. Our strategy for solving such problems will be to solve first the minimization problem $\min_{\mathbb{P}\in\mathscr{P}_{(m,\sigma)}} \mathbb{E}_{\mathbb{P}}[\lambda_e(p)]$, and then the maximization problem for this worst-case market. The technical challenge is to solve the minimization problem, which requires determining tight bounds for the expected wait for all distributions of Λ that comply with the partial information. However, this part was already resolved in the previous section, where we derived an upper bound on the expected wait. A tight upper bound on expected wait presents the worst-possible scenario in the rational choice model and yields the corresponding tight lower bound on the market share. Observe from the expression for the utility $U \geqslant 0$ that solving $\min_{\mathbb{P}\in\mathscr{P}_{(m,\sigma)}} \mathbb{E}_{\mathbb{P}}[\lambda_e(p)]$ is equivalent to solving $\max_{\mathbb{P}\in\mathscr{P}_{(m,\sigma)}} \mathbb{E}_{\mathbb{P}}[W(s, q(p)\Lambda)]$. That is, the worst-possible expected market share arises from the worst-possible utility and hence worst-possible expected wait. To make this more precise, notice that the maximin problem (3.5) can also be written as an optimization problem in terms of the equilibrium joining strategy $q(p)$; that is,

$$\begin{aligned} &\max_{q(p)\in[0,1]\cap[0,s\mu/\bar{\lambda})} \min_{\mathbb{P}\in\mathscr{P}_{(m,\sigma)}} \mathbb{E}_{\mathbb{P}}[q(p)\Lambda] \cdot \mathbb{E}_{\mathbb{P}}[r - cW(s, q(p)\Lambda)] \\ &\quad \equiv \max_{q(p)\in[0,1]\cap[0,s\mu/\bar{\lambda})} \min_{\mathbb{P}\in\mathscr{P}_{(m,\sigma)}} \mathbb{E}_{\mathbb{P}}[q(p)m\,(r - cW(s, q(p)\Lambda))], \end{aligned} \tag{3.6}$$

where the equivalence follows from the fact that the expected value $\mathbb{E}[\Lambda] = m$ is a known constant as it is part of the information contained in the ambiguity set. To see that (3.6) is equivalent to (3.5), we carefully go over the decision processes of the firm and the customers. First, the firm sets a price p, and then the customers settle on a joining strategy with expected utility $U = 0$, while considering all possible distributions that might govern the market size Λ. That is, the customers solve the equation $r = p + \max_{\mathbb{P}\in\mathcal{P}_{(m,\sigma)}} c\mathbb{E}_{\mathbb{P}}[W(s, q(p)\Lambda)]$, with the expected value operator appearing here because of the customers' internal calculations. In equilibrium, the utility equation can then be rewritten as $p = \min_{\mathbb{P}\in\mathcal{P}_{(m,\sigma)}} \mathbb{E}_{\mathbb{P}}[r - cW(s, q(p)\Lambda)]$. Now, if the firm sets the price to p, the expected revenue will equal this price times the market share, $p \cdot \lambda_e(p) = p \cdot \mathbb{E}_{\mathbb{P}}[q(p)\Lambda]$, yielding (3.6). In order to solve the utility equation, we must make additional assumptions regarding how customers experience waiting times in a mixed Poisson model.

The next question then is a rather philosophical one: What to assume for the wait expected by an arbitrary arriving customer? One natural choice is $\mathbb{E}_{\mathbb{P}}[W(s, q\Lambda)]$; see, e.g., [47] and [48]. Here, we assume nature picks a realization $\Lambda = \lambda$, and customers from time 0 onward arrive according to a Poisson process with this rate. Customers remain unaware of the specific universe (i.e., event $\Lambda = \lambda$) they live in but hold a (Bayesian) prior belief about the probability of residing in one universe compared to another. Another option is to base the rational decision made upon arrival on the posterior distribution of Λ, the updated version of the prior distribution of Λ that accounts for size bias. That is, the customer uses the conditional distribution of Λ given the realization of its own arrival event. [109] calls this phenomenon RASTA (Rate-biased ASTA), as a counterpart of PASTA. For any function $g(\cdot)$, the posterior expectation of $g(\Lambda)$ at arrival instants is $\mathbb{E}[\Lambda g(\Lambda)]/\mathbb{E}[\Lambda]$, and when g is nondecreasing and convex, then $\mathbb{E}[\Lambda g(\Lambda)]/\mathbb{E}[\Lambda] \geqslant \mathbb{E}[g(\Lambda)] \geqslant g(\mathbb{E}[\Lambda])$. As an application, consider the M/M/s queue. Then, the expected wait with a random arrival rate $q\Lambda$ equals

$$\overline{W}(s,q) := \frac{\mathbb{E}[\Lambda W(s,q\Lambda)]}{\mathbb{E}[\Lambda]} = \frac{\mathbb{E}[q\Lambda W(s,q\Lambda)]}{\mathbb{E}[q\Lambda]} = \frac{\mathbb{E}[L(s,q\Lambda)]}{\mathbb{E}[q\Lambda]}.$$

The utility of a customer who joins the queue when all the other customers use strategy q equals

$$U(q) = r - p - c\overline{W}(s,q)$$

and the best response of an individual customer is to join if and only if $U(q) \geqslant 0$. When accounting for size bias, we thus need to solve

$$\max_{\mathbb{P}\in\mathcal{P}_{(m,\sigma)}} \overline{W}(s,q) \tag{3.7}$$

to find the worst-case market share. Alternatively, assuming (P)ASTA instead of RASTA, we should consider

$$\max_{\mathbb{P}\in\mathcal{P}_{(m,\sigma)}} \mathbb{E}_{\mathbb{P}}[W(s,q\Lambda)]. \tag{3.8}$$

It is easy to show that (3.7) is also solved by the distribution in Lemma 3.2(ii), as it is known that $\lambda W(s,\lambda)$ is increasingly convex in λ (see [182]) and $\mathbb{E}[\Lambda]$ is a known constant since the mean m is contained in the ambiguity set. For the RASTA arrival assumption, we then obtain the following result.

Corollary 3.5 **(Market share with partially known market size, RASTA).** *Let the market size* Λ *follow a distribution* $\mathbb{P}$ *that belongs to the ambiguity set* $\mathscr{P}(m,\sigma,\underline{\lambda},\overline{\lambda})$*. Suppose the market consists of individuals that base their decision on the posterior distribution of* Λ*. Then, the worst-case joining probability* $q = q^*$ *solves the equation*

$$\frac{r-p}{c} = p_1 \frac{\lambda_1 W(s, q\lambda_1)}{m} + p_2 \frac{\lambda_2 W(s, q\lambda_2)}{m},$$

with $\lambda_1,\ \lambda_2,\ p_1$ *and* p_2 *as defined in* Theorem 3.4.

When we work with (3.8) instead, Theorem 3.4 implies the result for PASTA arrivals:

Corollary 3.6 **(Market share with partially known market size, PASTA).** *Let the market size* Λ *follow a distribution* $\mathbb{P}$ *that belongs to the ambiguity set* $\mathscr{P}(m,\sigma,\underline{\lambda},\overline{\lambda})$*. Suppose the market consists of individuals that base their decision on the prior distribution of* Λ*. Then, the worst-case joining probability* $q = q^*$ *solves the equation*

$$\frac{r-p}{c} = p_1 W(s, q\lambda_1) + p_2 W(s, q\lambda_2),$$

with $\lambda_1,\ \lambda_2,\ p_1$ *and* p_2 *as defined in* Theorem 3.4.

We next present numerical results for the M/M/1 model with random market size $\Lambda \sim \mathbb{P} \in \mathscr{P}_\Lambda(2,\sigma,0,4)$, $c = 0.25$, $r = 2$ and RASTA arrivals. Figure 3.1 illustrates the equilibrium joining probabilities and revenues for various price values p and different levels of variance σ^2. As expected, both the joining probabilities and revenues decrease with an increase in the dispersion of the market size. This observation aligns with the intuition that greater variance indicates more uncertainty about the market size and, as a result, greater uncertainty about the expected wait. Another noteworthy characteristic of the ambiguous model, as compared to the model with a known arrival rate, is that the equilibrium joining probability cannot exceed a certain level since, otherwise, the M/M/1 system can become unstable.

For conciseness, let $\mathbb{P}^*$ denote the extremal two-point distribution, as defined in Theorem 3.4, in the remainder of this section. We now present expressions for the optimal price in the RASTA context. To accomplish this, we start by deriving an expression for the derivative of the expected queue length with respect to λ, using the equation for $L'(s,\rho)$ from [91]. By substituting $\rho = \lambda/(s\mu)$ into the expression for $L'(s,\rho)$ and applying the chain rule, we obtain

$$M(s,\lambda) := \frac{\partial}{\partial\lambda} L(s,\lambda) = \frac{1}{\mu}\left(\frac{C(s,\lambda)\mu(s\mu+\lambda-C(s,\lambda)\lambda)}{(\lambda-s\mu)^2} + C(s,\lambda)+1\right).$$

Let us now determine the joining probability that results in the maximal revenue for the firm.

Proposition 3.7 **(Optimal joining probability).** *Consider the M/M/s queue with RASTA arrivals. Suppose that* $c/\mu < r < c\overline{W}(s,1)$*. Then the joining probability* q^* *that maximizes the revenue for the firm, in the worst-case market* $\Lambda \sim \mathbb{P} \in \mathscr{P}_{(m,\sigma)}$*, is the unique solution in* q *of*

$$r - c\frac{\mathbb{E}_{\mathbb{P}^*}[\Lambda M(s, q\Lambda)]}{m} = 0. \tag{3.9}$$

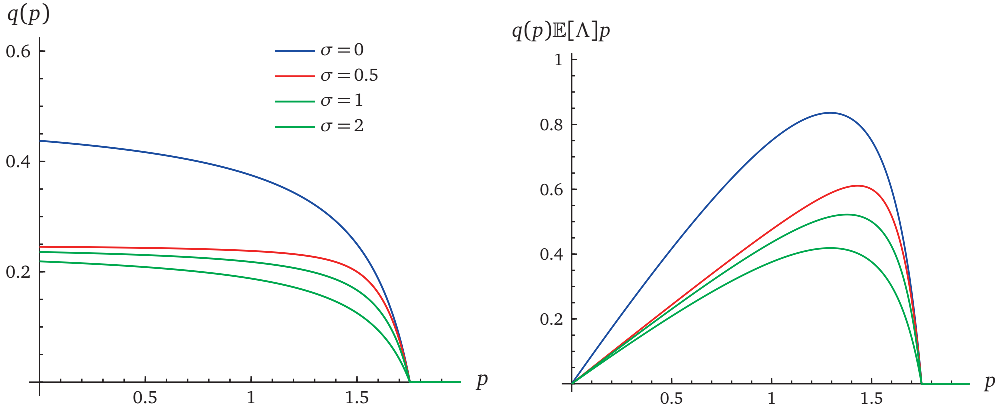


Figure 3.1: Equilibrium probability and revenue for different prices p in the M/M/1 model with random market size $\Lambda \sim \mathbb{P} \in \mathscr{P}_\Lambda(2, \sigma, 0, 4)$, $c = 0.25$, $r = 2$ and RASTA arrivals

Proof. It is a known property of RASTA arrivals that for the revenue-maximizing firm the objective function is equivalent to that of the social optimizer [109]. This leads us to consider the maximin social welfare as the objective function:

$$\max_{q\in(0,1)} \min_{\mathbb{P}\in\mathscr{P}_{(m,\sigma)}} \mathbb{E}_{\mathbb{P}}[q\Lambda r - cL(s, q\Lambda)].$$

It suffices to consider $q \in (0, 1)$ as, due to the assumption, both $q = 0$ and $q = 1$ are not equilibrium strategies. Notice that the min problem is solved by $\mathbb{P}^*$, since $L(s, q\lambda)$ has the required curvature properties. We can now solve the max part. Taking the derivative w.r.t. q (which is allowed as Λ is bounded), we obtain the first-order condition

$$r\mathbb{E}_{\mathbb{P}^*}[\Lambda] - c\mathbb{E}_{\mathbb{P}^*}[\Lambda M(s, q\Lambda)] = 0,$$

which can be written as (3.9). Since $M(s, \cdot)$ is an increasing function of λ, as $L(s, \cdot)$ is (strictly) convex, this implies that (3.9) has a unique solution that is the global maximizer, as the social welfare objective is a concave function of q. □

We next determine the minimax price p that induces the optimal joining strategy q^*.

Proposition 3.8 **(Optimal maximin price).** *For the M/M/s queue with RASTA arrivals, the optimal price for the firm that solves the maximin problem* (3.5) *is*

$$p^* = c\frac{\mathbb{E}_{\mathbb{P}^*}\left[\Lambda\left(M\left(s, q^*\Lambda\right) - W\left(s, q^*\Lambda\right)\right)\right]}{m}. \tag{3.10}$$

Proof. It suffices to show that the right-hand side of (3.10) yields a solution to $q(p) = q^*$, or equivalently, $U(q^*) - p = 0$. Indeed, plugging in the right-hand side of (3.10), we see that

$$U(q^*) - p = r - c \max_{\mathbb{P}\in\mathscr{P}_{(m,\sigma)}} \overline{W}(s,q^*) - c\frac{\mathbb{E}_{\mathbb{P}^*}\left[\Lambda\left(M\left(s,q^*\Lambda\right) - W\left(s,q^*\Lambda\right)\right)\right]}{\mathbb{E}_{\mathbb{P}^*}[\Lambda]}$$

$$= r - c\frac{\mathbb{E}_{\mathbb{P}^*}[\Lambda M(s,q\Lambda)]}{\mathbb{E}_{\mathbb{P}^*}[\Lambda]} = 0,$$

in which the final identity follows from Proposition 3.7. □

In conclusion, we have shown that our second-order bounds can aid in deriving the optimal maximin price for a firm seeking to maximize revenue. This firm serves a rational customer base that makes decisions based on limited information about the overall market size. Interestingly, this adverse market follows a computationally tractable two-point distribution, thus allowing for the application of established techniques designed for rational queueing models with stochastic arrival rates.

3.4. Other types of market information

In this section, we demonstrate the tractability of our approach when applied to various types of market information. It appears that the curvature properties, central to this work, also provide the necessary mathematical tractability for other types of ambiguity sets. Section 3.4.1 introduces two distinct measures of dispersion as alternatives to variance: mean absolute deviation and upper semivariance. Section 3.4.2 incorporates unimodality information to refine the mean-variance bounds. Section 3.4.3 examines a data-driven model in which the decision maker progressively learns the true distribution of the market size and offers a robust solution using the Wasserstein ambiguity set.

3.4.1. Alternative dispersion measures

Besides variance information, there exist other types of dispersion measures which, in conjunction with the curvature properties, lead to easy solutions to (3.1) for different ambiguity sets. For example, let $\mathscr{P}_{(m,d)} := \mathscr{P}(m,d,\underline{x},\overline{x})$ be the set of all distributions with mean m, mean absolute deviation (MAD) $d := \mathbb{E}|X-\mu|$ and support contained in the interval $[\underline{x},\overline{x}]$. The following result follows immediately from the mean-MAD bounds discussed in the previous chapters.

Theorem 3.9 **(M/M/s queue with MAD information).** *Consider an M/M/s queue with random arrival rate Λ that follows a distribution $\mathbb{P}$ belonging to the ambiguity set $\mathscr{P}(m,d,\underline{\lambda},\overline{\lambda})$. The tight upper bound for the expected wait $\mathbb{E}_{\mathbb{P}}[W(s,\Lambda)]$ is attained by a three-point distribution with*

$$\mathbb{P}(\Lambda=\underline{\lambda}) = \frac{d}{2(m-\underline{\lambda})} =: p_1,\ \mathbb{P}(\Lambda=m) = 1 - \frac{d}{2(m-\underline{\lambda})} - \frac{d}{2(\overline{\lambda}-m)} =: p_2,$$

$$\mathbb{P}(\Lambda=\overline{\lambda}) = \frac{d}{2(\overline{\lambda}-m)} =: p_3,$$

and given by

$$p_1W(s,\underline{\lambda}) + p_2W(s,m) + p_3W(s,\overline{\lambda}).$$

We next consider a particular asymmetric dispersion measure that only measures dispersion above the mean. Let $\mathscr{P}_{(m,\overline{\sigma})} := \mathscr{P}(m, \overline{\sigma}, \underline{x}, \overline{x})$ be the set of all distributions with mean m, (upper) semivariance $\overline{\sigma} := \mathbb{E}[(X-\mu)^+]^2$, where $(x)^+ = \max\{x, 0\}$, and the support contained in the interval $[\underline{x}, \overline{x}]$. Using primal-dual techniques, we prove the following result.

Theorem 3.10 **(M/M/*s* queue with semivariance information).** *Consider an M/M/s queue with random arrival rate Λ that follows a distribution $\mathbb{P}$ belonging to the ambiguity set $\mathscr{P}(m, \overline{\sigma})$. Suppose that $m \in (\underline{\lambda}, \overline{\lambda})$ and $\overline{\sigma} \in (0, \frac{(\overline{\lambda}-m)^2(m-\underline{\lambda})}{(\overline{\lambda}-\underline{\lambda})})$. Then, the tight upper bound for the expected wait corresponds to the maximum value of $\mathbb{E}_{\mathbb{P}}[W(s,\Lambda)]$ that results from the following two solutions:*

(i) *the expected wait with the expectation taken over a three-point distribution with*

$$\mathbb{P}(\Lambda = \underline{\lambda}) = p_1(x_0^*),\ \mathbb{P}(\Lambda = x_0^*) = p_2(x_0^*),\ \mathbb{P}(\Lambda = \overline{\lambda}) = p_3(x_0^*),$$

where $x_0^* \in [\mu, \overline{\lambda})$ solves

$$\max_{x_0 \in [\mu, \overline{\lambda})} p_1(x_0) W(s, \underline{\lambda}) + p_2(x_0) W(s, x_0) + p_3(x_0) W(s, \overline{\lambda}) \ \text{ s.t. } \ 0 \leqslant p_1, p_2, p_3 \leqslant 1; \ \text{or}$$

(ii) *the expected wait with the expectation taken over the two-point distribution*

$$\begin{aligned}
\mathbb{P}(\Lambda = \underline{\lambda}) &= \frac{\sqrt{4\overline{\sigma}(m-\underline{\lambda})^2 + \overline{\sigma}^2} - \overline{\sigma}}{2(m-\underline{\lambda})^2}, \\
\mathbb{P}(\Lambda = x_0^*) &= \frac{2(m-\underline{\lambda})^2}{\sqrt{4\overline{\sigma}(m-\underline{\lambda})^2 + \overline{\sigma}^2} + 2(m-\underline{\lambda})^2 + \overline{\sigma}}, \\
\text{where } x_0^* &= \frac{2m(m-\underline{\lambda}) + \sqrt{4\overline{\sigma}(m-\underline{\lambda})^2 + \overline{\sigma}^2} + \overline{\sigma}}{2(m-\underline{\lambda})}.
\end{aligned}$$

Proof. The primal problem for mean-(upper)semivariance information is given by

$$\begin{aligned}
\max_{\mathbb{P}(x) \geqslant 0} \quad & \int_{\underline{x}}^{\overline{x}} \phi(x) \mathrm{d}\mathbb{P}(x) \\
\text{s.t.} \quad & \int_{\underline{x}}^{\overline{x}} \mathrm{d}\mathbb{P}(x) = 1, \ \int_{\underline{x}}^{\overline{x}} x \, \mathrm{d}\mathbb{P}(x) = m, \ \int_{\underline{x}}^{\overline{x}} ((x-m)^+)^2 \mathrm{d}\mathbb{P}(x) = \overline{\sigma},
\end{aligned} \tag{3.11}$$

and admits the following dual:

$$\begin{aligned}
\min_{\lambda_0, \lambda_1, \lambda_2} \quad & \lambda_0 + \lambda_1 m + \lambda_2 \overline{\sigma} \\
\text{s.t.} \quad & M_{\overline{\sigma}}(x) := \lambda_0 + \lambda_1 x + \lambda_2 ((x-m)^+)^2 \geqslant \phi(x), \ \forall x \in [\underline{x}, \overline{x}].
\end{aligned} \tag{3.12}$$

The objective function $\phi(\cdot)$ is increasingly convex since it represents $W(s,\cdot)$. Under the conditions imposed on the parameters of the ambiguity set, strong duality holds and the optimal values of the primal and dual problem coincide. In addition, since the common optimal value is finite, the dual optimal solution is attained [196, Proposition 3.4]. Moreover, as both the objective function and the dual function $M_{\overline{\sigma}}(x)$ are continuous and the support $[a, b]$ is compact, the optimal primal solution is also attained [196, Corollary 3.1]. As a consequence, complementary slackness holds [196, Proposition 2.1]. We next use the complementary slackness property and structural properties of the functions ϕ and $M_{\overline{\sigma}}$ to determine which points should constitute the support of the extremal distribution. Since the primal problem has three constraints, we can restrict our search to a worst-case distribution with at most three support points [186]. Since the conditions in the theorem impose $\overline{\sigma} > 0$, we can readily exclude the case of a one-point worst-case distribution. Furthermore, we need both a point below and above the mean to satisfy the mean condition and to avoid the semivariance being equal to zero. For the support point below the mean, $\underline{x}$ is the only dual-feasible option by convexity of ϕ, as $M_{\overline{\sigma}}(\cdot)$ is linear for all $x \in [\underline{x}, m]$. We next consider two dual solutions that correspond to cases (i) and (ii) stated in the theorem, which assert a worst-case two- and three-point distribution, respectively.

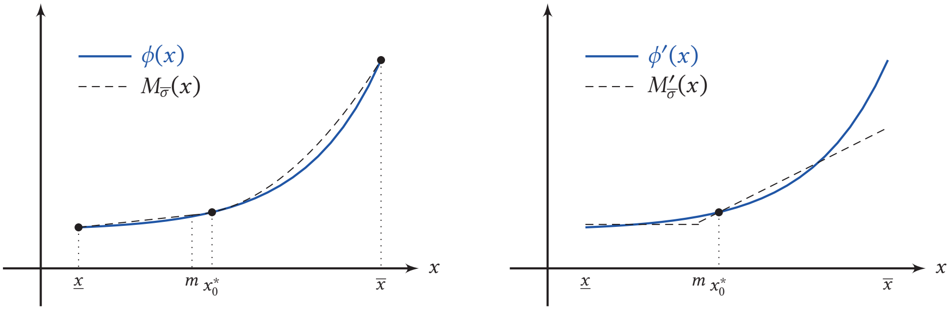


Figure 3.2: Objective function $\phi(x)$, dual function $M_{\overline{\sigma}}(x)$ and their derivatives, where $M'_{\overline{\sigma}}(x)$ is interpreted as the right-derivative

(i) Fixing the first support point to $\underline{x}$, we next seek the other two points using another complementary slackness argument. The dual function $M_{\overline{\sigma}}$ can only be tangent to ϕ at one unique point x_0 on the interval $[m, \overline{x}]$. This follows from linearity of $M'_{\overline{\sigma}}(x)$ and convexity of $\phi'(x)$, as illustrated in Figure 3.2, which rules out the possibility of a second tangent point. To see this, assume that there exists a second tangent point $\tilde{x}_0$. To remain dual-feasible, the dual function has to satisfy $M'_{\overline{\sigma}}(x) \leqslant \phi'(x)$ for $x \uparrow \tilde{x}_0$, and $M'_{\overline{\sigma}}(x) \geqslant \phi'(x)$ for $x \downarrow \tilde{x}_0$. Therefore, $M'_{\overline{\sigma}}(x)$ has to intersect $\phi'(x)$ from below at $\tilde{x}_0$, but since this already occurred at x_0^*, this cannot happen a second time as otherwise $M'_{\overline{\sigma}}(x)$ has to be nonlinear or $\phi'(x)$ nonconvex. Hence, we arrive at a contradiction. The dual function can only coincide with $\phi(x)$ one more time, at the upper bound of the support, $x = \overline{x}$. Therefore, if the worst case is given by a three-point distribution, then it has to admit this specific form as a result of complementary slackness. The probabilities, as functions of x_0, follow from solving the moment constraints in (3.11). As x_0 is yet to be determined, the maximum value follows from solving the univariate optimization problem stated in the claim. There always exists a feasible solution to this optimization problem. Specifically, letting $x_0 = m$, we obtain the three-point distribution

$$p_1 = \frac{\overline{\sigma}}{(\overline{x} - m)(m - \underline{x})}, \ p_2 = 1 - \frac{\overline{\sigma}(\overline{x} - \underline{x})}{(m - \underline{x})(\overline{x} - m)^2}, \ p_3 = \frac{\overline{\sigma}}{(\overline{x} - m)^2},$$

of which is easily verified, using the conditions in the claim, that it is primal feasible for all possible parameter combinations.

(ii) We next consider the two-point solution. If the dual function only coincides with ϕ at $\underline{x}$ and some x_0, then the resulting two-point distribution can be directly derived from the moment conditions

$$p_1 + p_2 = 1,\ p_1\underline{\lambda} + p_2 x_0^* = m,\ p_2(x_0^* - m)^2 = \overline{\sigma},$$

which is a system of three equations with three unknowns with a unique solution, as stated in the claim. Further, from the assumptions in the claim, it follows that this distribution is always feasible in the primal. Taking the maximum expected value of the assertions (i) and (ii) completes the proof of the claim. □

Although upper semivariance may be regarded as a useful dispersion measure, as it quantifies the upside risk of a substantially large market on the expected wait, its use as a dispersion measure has a drawback. Despite being able to restrict some of the support points using the curvature properties of ϕ, the worst-case distribution still partly depends on the objective function in some cases, necessitating an additional numerical step to determine x_0^*. In contrast, MAD dispersion leads to an extremal distribution that no longer depends on the exact expression for $W(s, \lambda)$, thus yielding similar computational advantages as the variance-based ambiguity set.

Table 3.2 displays the numerical results for the tight mean-MAD bounds, utilizing the same numerical setup as Table 3.1. We observe that the MAD bounds perform comparably to the variance-based bounds. For low-utilization regimes, the variance bounds appear much sharper, whereas for high-utilization levels, the MAD bounds appear to result in values closer to the ground truth. However, we want to stress that a direct comparison is meaningless, as the underlying ambiguity sets might contain vastly different distributions.

Table 3.2: Numerical bounds for the expected wait in the M/M/2 queue with beta arrival rate distribution

	$\overline{\lambda} = 1.95$			$\overline{\lambda} = 1.99$		
ρ	$\mathbb{E}[W(2,\Lambda)]$	Var-UB	MAD-UB	$\mathbb{E}[W(2,\Lambda)]$	Var-UB	MAD-UB
0.2	1.0449	1.1199	1.5325	1.0449	1.4312	3.5291
0.5	1.3440	1.5315	2.1166	1.3440	2.3236	5.3162
0.7	2.0096	2.5002	3.2631	2.0098	4.6625	8.5942
0.8	2.9465	3.9446	4.7258	2.9497	8.6515	12.7120
0.9	6.6019	9.0133	9.0679	6.9387	25.0550	24.8840

3.4.2. Sharper bounds for unimodal market size

Often more is known about the market-size distribution than just the mean and variance. For example, we may have some structural information regarding the distribution of the market size. In this section, we show how to use this structural information to improve the tight bounds, corresponding to the situation in which the service provider has the additional information that the market-size distribution is unimodal, which can be understood as there being one primary market segment that constitutes the bulk of the arrivals requesting service.

A random variable is said to be unimodal with mode $\overline{m}$ if its distribution can be characterized as the mixture of a Dirac distribution $\delta_{\overline{m}}$ and a distribution function that is a concave function on $[\underline{x}, m]$ and a convex function on $(m, \overline{x}]$. Some examples of unimodal distributions are the normal, exponential and beta probability distributions. In order to incorporate such unimodality, we apply Khinchine's characterization theorem, which states that a random variable Y has a unimodal distribution with mode zero if, and only if, there exists a random variable Y such that $Y = UV$, where U is a uniformly distributed random variable on $[0, 1]$ independent of V (see, e.g., [79, p. 158]). We adopt the following approach from [38], which is based on the, slightly more general, results derived by Kemperman [131]. First, we transform the problem from the unimodal random variable X to the auxiliary variable V. Then, we establish tight bounds for V and use them to obtain bounds for X. Notice that if X is unimodal with mode $\overline{m}$, then $Y = X - \overline{m}$ is also unimodal but with mode 0. Hence, according to Khinchine's theorem, we have $Y = UV$, where U is uniformly distributed over $[0, 1]$. According to Kemperman's approach, we can transform the moment problem in terms of the random variable X into one in terms of the auxiliary random variable V using that for any function ϕ, $\mathbb{E}[\phi(Y)] = \mathbb{E}[\phi^*(V)]$, where

$$\phi^*(x) = \frac{1}{x}\int_0^x \phi(t)dt = \mathbb{E}[\phi(UV) \,|\, V = x]. \tag{3.13}$$

By taking $\phi(x) = x^k$, we can relate the moments of Y to those of V through

$$\phi^*(x) = \frac{1}{(k+1)}x^k,$$

so that $\mathbb{E}[V^k] = (k+1)\mathbb{E}[Y^k]$. The mean and variance of V, in terms of the mean and variance of X, are given by

$$m_V = \mathbb{E}[V] = 2\mathbb{E}[Y] = 2\left(m_X - \overline{m}\right).$$
$$\sigma_V^2 = \mathbb{E}[V^2] - (\mathbb{E}[V])^2 = 3\mathbb{E}[Y^2] - 4(\mathbb{E}[Y])^2 = 3\sigma_Y^2 - (\mathbb{E}[Y])^2 = 3\sigma_X^2 - (m_X - \overline{m})^2.$$

We can use the same integral relationship to find tight bounds for $\mathbb{E}[\phi(X)]$ when X is unimodal with mode $\overline{m}$. Specifically, this problem is equivalent to finding bounds for $\mathbb{E}[\phi^*(V)]$ subject to the moment constraints on V, where

$$\phi^*(x) = \frac{1}{x}\int_0^x \phi(u + \overline{m})\mathrm{d}u.$$

We already have the solution for this transformed problem, since the (conditional) expectation operator (3.13) preserves the curvature properties. By substituting the transformed moments into the result stated in Lemma 3.2, we obtain the following result.

Theorem 3.11 **(M/M/s queue with unimodality).** *Consider an M/M/s queue with random arrival rate Λ. Suppose that Λ is unimodal with mode $\overline{m}$, mean m, variance σ^2 and has support contained in $[\underline{\lambda}, \overline{\lambda}]$. Let*

$$W^*(s,x) = \frac{1}{x}\int_m^{x+m} W(s,t)\mathrm{d}t.$$

Then, the tight upper bound for the expected wait $\mathbb{E}_{\mathbb{P}}[W(s,\Lambda)]$ is given by

$$(1-p)W^*(s,\lambda_1) + pW^*(s,\lambda_2),$$

where

$$\lambda_1 = 2m - 2\overline{m} - \frac{3\sigma^2 - (m-\overline{m})^2}{\overline{\lambda} + \overline{m} - 2m}, \quad \lambda_2 = \overline{\lambda} - \overline{m}, \quad p = \frac{3\sigma^2 - (m-\overline{m})^2}{3\sigma^2 - (m-\overline{m})^2 + \left(\overline{\lambda} + \overline{m} - 2m\right)^2}.$$

Table 3.3 demonstrates the performance of the tight mean-variance bounds that incorporate unimodality information. Evidently, incorporating this structural information significantly sharpens the bounds compared to the standard mean-variance bounds without this information. Nevertheless, the bounds still substantially diverge from the true value for high utilization levels.

Table 3.3: Numerical bounds for the expected wait in the M/M/2 queue with unimodal beta arrival rate distribution

	$\overline{\lambda} = 1.95$			$\overline{\lambda} = 1.99$		
ρ	$\mathbb{E}[W(2,\Lambda)]$	Var-UB	Unimod-UB	$\mathbb{E}[W(2,\Lambda)]$	Var-UB	Unimod-UB
0.2	1.0449	1.1199	1.0583	1.0449	1.4312	1.0686
0.5	1.3440	1.5315	1.3879	1.3440	2.3236	1.4277
0.7	2.0096	2.5002	2.1502	2.0098	4.6625	2.3124
0.8	2.9465	3.9446	3.2477	2.9497	8.6515	3.7098
0.9	6.6019	9.0133	7.1468	6.9387	25.0550	9.3105

3.4.3. Data-driven setting: Learning the market size

We next derive some first reformulations for the model with a Wasserstein ambiguity set [85, 157]. In this model, the worst-case is taken over a set of distributions that are "sufficiently" close to the empirical distribution obtained from data. Here, the distance between distributions is measured with the Wasserstein metric. We first define the notion of a Wasserstein ambiguity set. Let $\mathscr{P}_0([\underline{\lambda}, \overline{\lambda}])$ be the set of all probability distributions $\mathbb{P}$ of Λ supported on $[\underline{\lambda}, \overline{\lambda}]$. For $r \geqslant 1$, the r-Wasserstein distance between two distributions of Λ is then defined as

$$d_r(\mathbb{P}_1, \mathbb{P}_2) = \inf\left\{\left(\int_{[\underline{\lambda},\overline{\lambda}]^2} |\lambda_1 - \lambda_2|^r \mathrm{d}\mathbb{Q}(\lambda_1,\lambda_2)\right)^{\frac{1}{r}}\right\} \tag{3.14}$$

in which $\mathbb{Q}$ is a joint distribution of Λ_1 and Λ_2 with marginals $\mathbb{P}_1$ and $\mathbb{P}_2$, respectively. Assume we observe n independent realizations given by $\{\hat{\lambda}_i\}_{i=1,\dots,n}$, and define the empirical distribution as $\hat{\mathbb{P}}_n := \frac{1}{n}\sum_{i=1}^n \delta_{\hat{\lambda}_i}$ with δ_x being the Dirac measure with mass concentrated on x. Now, define the r-Wasserstein ambiguity set as

$$\mathscr{B}_\epsilon^r(\hat{\mathbb{P}}_n) := \left\{ \mathbb{P} \in \mathscr{P}_0([\underline{\lambda}, \overline{\lambda}]) \, : \, d_r(\mathbb{P}, \hat{\mathbb{P}}_n) \leqslant \epsilon \right\} \tag{3.15}$$

with ϵ as the radius of the Wasserstein ball and $\hat{\mathbb{P}}_n$ as the reference distribution. We next present a result for the setting with $r = 1$.

Theorem 3.12 **(M/M/s queue with Wasserstein ambiguity).** *Consider an M/M/s queue with a random arrival rate Λ that follows a distribution $\mathbb{P}$ belonging to the ambiguity set $\mathscr{B}_\epsilon^1(\hat{\mathbb{P}}_n)$. The tight upper bound for $\mathbb{E}_\mathbb{P}[W(s,\Lambda)]$ coincides with the optimal objective value of the linear program*

$$\begin{aligned} \inf_{\alpha\in\mathbb{R}_+,\boldsymbol{\beta}\in\mathbb{R}^n} \; & \alpha\epsilon + \frac{1}{n}\sum_{i=1}^n \beta_i \\ \text{s.t.} \; & \beta_i - \alpha(\underline{\lambda} - \hat{\lambda}_i) \geqslant W(s,\underline{\lambda}), \quad \forall i = 1,\dots,n, \\ & \beta_i \geqslant W(s,\hat{\lambda}_i), \quad \forall i = 1,\dots,n, \\ & \beta_i + \alpha(\overline{\lambda} - \hat{\lambda}_i) \geqslant W(s,\overline{\lambda}), \quad \forall i = 1,\dots,n. \end{aligned} \tag{3.16}$$

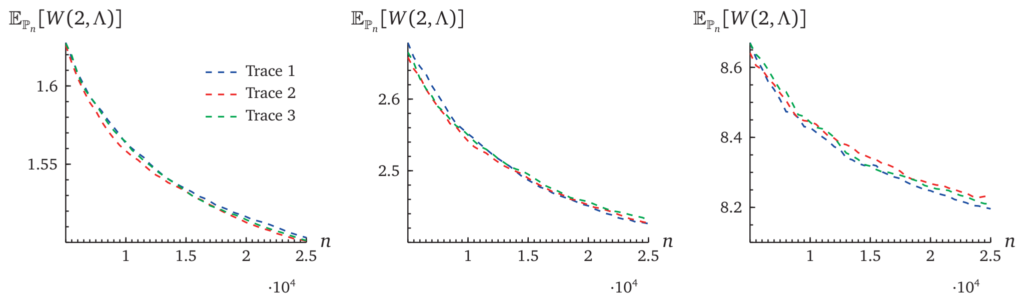
Proof. The strong dual of $\sup_{\mathbb{P}\in\mathscr{B}_\epsilon^1(\hat{\mathbb{P}}_n)} \mathbb{E}_\mathbb{P}[W(s,\Lambda)]$ is given by [157, Theorem 4.2]

$$\begin{aligned} \inf_{\alpha\in\mathbb{R}_+,\boldsymbol{\beta}\in\mathbb{R}^n} \; & \alpha\epsilon + \frac{1}{n}\sum_{i=1}^n \beta_i \\ \text{s.t.} \; & \beta_i + \alpha|\lambda - \hat{\lambda}_i| \geqslant W(s,\lambda) \quad \forall i = 1,\dots,n, \; \forall \lambda \in [\underline{\lambda}, \overline{\lambda}]. \end{aligned} \tag{3.17}$$

Given that the expected wait is convex in λ, it suffices to verify that the semi-infinite constraints are satisfied at the three points $\lambda = \underline{\lambda}, \hat{\lambda}_i, \overline{\lambda}$, from which the claim follows. □

Computing bounds for the expected wait using Theorem 3.12 brings substantial benefits, especially in the data-driven setting. By adjusting the radius ϵ of the Wasserstein ball, one can control the degree of conservatism of the bound, where $\epsilon = 0$ represents the singleton set containing the empirical distribution based on the n sampled data points. Furthermore, it has been demonstrated that the data-driven bound in (3.16) converges to the true value of the stochastic model as $n \to \infty$ [157].

To demonstrate the effectiveness of the data-driven bounds, we conducted a numerical experiment using the same setup as presented in Table 3.1 as our ground truth. The arrival rate Λ is modeled by a beta distribution, and we generated three traces of the data-driven bounds for a varying number of data points, n, ranging from 5,000 to 20,000. These bounds were computed for low-, medium-, and high-utilization scenarios. For the actual values of $\mathbb{E}_\mathbb{P}[W(s,\Lambda)]$, please refer to Table 3.1. Notice that the data-driven bounds converge to the true values as the number of data points increases while the radius of the Wasserstein ball shrinks with $\epsilon = 1/\sqrt{n}$. However, convergence occurs at a slower rate for higher utilization levels. Furthermore, as the utilization level increases, the behavior of the computed traces becomes more erratic.



(a) $m = 1$ **(b)** $m = 1.4$ **(c)** $m = 1.8$

Figure 3.3: Data-driven bounds for the expected wait of an M/M/2 queue with beta distributed arrival rate and $\epsilon = 1/\sqrt{n}$

3.5. Further applications

This section attempts to showcase the extensive range of applications for our approach and position it within a more general framework. In Section 3.5.1, we apply the second-order bounds to a staffing optimization problem. In Section 3.5.2, we demonstrate that our results are also valid for other types of queues with uncertain input parameters. Last, in Section 3.5.3, we explore the extension wherein the service rate parameter is likewise a random parameter.

3.5.1. Staffing problems

An alternative optimization problem that one can consider is staffing. The goal of staffing is to strike the right balance between the capacity allocation costs and delay costs incurred. To be specific, let a be the server allocation costs per unit time, and let h be the penalty cost per delayed job per unit time. Here, delay is time in the queue only. Assume that the servers work at unit rate. This yields, for a known arrival rate, the total cost function

$$K(s,\lambda) := a\, s + h\lambda \frac{C(s,\lambda)}{s-\lambda}.$$

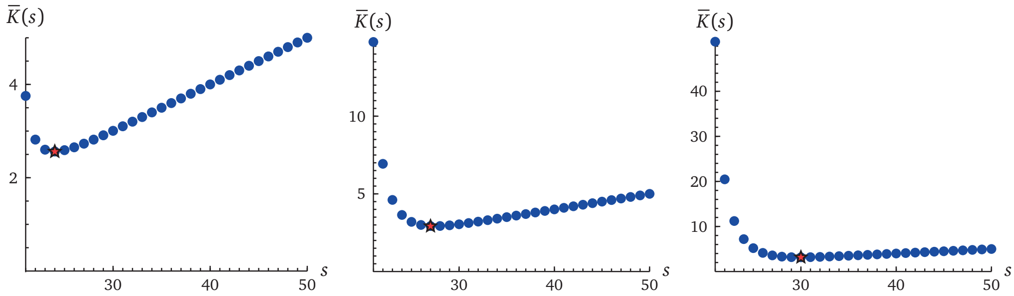
For uncertain arrival rate $\Lambda \sim \mathbb{P} \in \mathscr{P}_{(m,\sigma)}$, the two-point extremal distribution also yields the worst case for the expected delay costs rate since

$$\lambda \frac{C(s,\lambda)}{s-\lambda} = \lambda W(s,\lambda) - \frac{\lambda}{\mu} = L(s,\lambda) - \frac{\lambda}{\mu}$$

is still increasingly convex in λ. Assume that the decision maker aims to minimize total costs given a mixed Poisson arrival process with a random arrival rate Λ. The decision maker considers the time-average metrics of the system. Therefore, we seek to minimize the worst-case costs

$$\overline{K}(s,\lambda) := as + h\left(p_1\lambda_1 \frac{C(s,\lambda_1)}{s-\lambda_1} + p_2\lambda_2 \frac{C(s,\lambda_2)}{s-\lambda_2}\right),$$

where λ_1, λ_2, p_1 and p_2 are defined in Theorem 3.4. We now ask for the capacity level s that minimizes $\overline{K}(s,\lambda)$, denoted by $s^* \in \arg\min_{s>\overline{\lambda}} \overline{K}(s,\lambda)$. We impose $s > \overline{\lambda}$ as this ensures that the resulting system is stable. Notice that the search for this global optimum can be performed efficiently using standard numerical techniques, as the objective is convex in the decision variable, the number of servers s. Here, (discrete) convexity as a function of the number of servers follows from Harel [105, Proposition 5]. We emphasize that the quality-of-service model can be applied in this context as well. Given that $C(s,\cdot)$ is convex with respect to λ, the MAD-based ambiguity yields a mathematically tractable solution.



(a) $m = 5$ **(b)** $m = 10$ **(c)** $m = 15$

Figure 3.4: Optimal staffing example with $a = 0.1$, $h = 5$, $\overline{\lambda} = 20$ while fixing $\frac{\sigma^2}{m^2} = 0.2$

We proceed by discussing a numerical example. Consider $\mu = 1$, $a = 0.1$, $h = 5$, and $\overline{\lambda} = 20$. Figure 3.4 displays the costs for different staffing levels as we vary the mean arrival rate, while keeping the squared coefficient of variation fixed at $\sigma^2/m^2 = 0.2$. The bounds diverge to infinity for $s \leqslant \overline{\lambda}$, resulting in an unstable system. As the mean arrival rate increases, the optimal value of s^* also increases, and the marginal cost of adding servers above this optimal value appears to diminish.

3.5.2. Other queueing systems

We next exploit convexity in conjunction with mean-MAD ambiguity so that we can handle an even more diverse set of models. Establishing convexity of mean performance measures, such as $W(s,\lambda)$, is of great importance for the efficient optimization of stochastic systems, particularly for the optimal design of queueing systems. Our problem is not much different from these parametric optimization models. We also seek to establish convexity of the total system time as a function of these parameters, for example, $W(\lambda) := \mathbb{E}[\widetilde{W}(\lambda)]$, where the distribution of $\widetilde{W}$ depends on the underlying stochastic model. However, for our application, convexity results in tractable solutions to the maximization step in a minimax problem, rather than the optimal value of these parameters in order to minimize waiting costs. Hence, even though the ultimate goal is different, our models can still benefit from the same convexity results.

Often direct algebraic manipulations on closed-form expressions for the mean performance measure are considered. However, this algebra turns out to be rather tedious, even for relatively simple models. For convexity in the service rate μ, Weber [220] presented a powerful approach for the G/G/1 queue by analyzing the sample-path dynamics of the stochastic process. Consider the G/G/1 queueing system with (parameterized) service times $S_n(\mu) = S_n/\mu$ and interarrival times T_n, for all $n \geqslant 1$. From Lindley's recursion,

$$\widetilde{W}_n(\mu) = \left(\widetilde{W}_{n-1}(\mu) + S_n/\mu - T_n\right)^+, \quad n \geqslant 1, \quad \widetilde{W}_0(\mu) = 0, \tag{3.18}$$

it is easily observed that the waiting times $\widetilde{W}_n(\cdot)$ are decreasing and convex in μ for all sample paths so that this result must hold true for the expected wait $W(\mu)$ as well.

Let us discuss two convexity results that arise from these sample-path techniques. First, consider the single-server queue with Poisson arrivals, for which the arrival rate is given by the random variable Λ, and generally distributed service times with mean $1/\mu$, i.e., a variant of the M/G/1 system, but with mixed Poisson input. The goal is to determine bounds for the expected waiting time of a customer, where the expected wait is defined as a function of the arrival rate λ. It is well known that the expected wait in the M/G/1 queue, denoted here as $W_*(\cdot)$, is a convex function of the arrival rate λ (see, e.g., [82]). For a proof of convexity in λ using sample-path arguments, see [195]. As a result, the three-point distribution constitutes the worst-case for the expected wait of the M/G/1 queue.

Proposition 3.13 **(M/G/1 queue with MAD information).** *Let the market size Λ have a distribution $\mathbb{P}$ that resides in the ambiguity set $\mathcal{P}(m, d, \underline{\lambda}, \overline{\lambda})$, and let $W_*(\lambda)$ denote the expected wait in the M/G/1 queue. Suppose that $\overline{\lambda} < \mu$. Then, the tight upper bound for $\mathbb{E}_{\mathbb{P}}[W_*(\Lambda)]$ is attained by a three-point distribution as in* Theorem 3.9.

Next, consider a multi-server queue with s servers, Poisson arrivals and deterministic service times (i.e., an M/D/s queueing system). For this system, assigning customers to servers cyclically preserves the first-come-first-served queueing policy, due to the deterministic service times. Whenever there is an idle server, there is no waiting customer. Each server thus serves the same set of customers, and as a result, each queue can view its channel as a separate M/D/1 queue. Convexity of $W_*(s, \lambda)$ in λ follows from Theorem 3(b) in [104]. Hence, our result for the M/G/1 queue directly implies the following proposition:

Proposition 3.14 **(M/D/s queue with MAD information).** *Let the market size Λ have a distribution $\mathbb{P}$ that resides in the ambiguity set $\mathcal{P}(m, d, \underline{\lambda}, \overline{\lambda})$, and let $W_*(s, \lambda)$ denote the expected wait in the M/D/s queue. Then, the tight upper bound for $\mathbb{E}_{\mathbb{P}}[W_*(s, \Lambda)]$ is attained by a three-point distribution as in* Theorem 3.9.

A common misconception is that the expected waiting time in the G/G/1 queue is a convex function of the arrival rate when the interarrival times in the Lindley recursion (3.18) are scaled by $1/\lambda$. In their brief note, Fridgeirsdottir and Chiu [82] rightfully claim that much of the congestion pricing literature is flawed due to this unfounded assumption. However, as the authors point out, a different performance metric is utilized in the majority of these analyses. So, although the expected wait is not necessarily a convex function of λ, Fridgeirsdottir and Chiu [82] demonstrate that the expected delay cost rate, $\lambda W(\lambda)$, is convex in the arrival rate, also for the G/G/1 queue. We next turn back to the rational queueing model. Now say the firm searches for a socially optimal fee that maximizes social welfare, rather than its own financial benefits. This leads us to consider the following maximin problem (see, e.g., [48] for the original problem with full distributional information):

$$\max_{p} \min_{\mathbb{P}\in\mathcal{P}_{(m,d)}} \mathbb{E}_{\mathbb{P}}[q(p)\Lambda(r - cW(q(p)\Lambda))]$$

Again, $q(p)$ is an implicit function of p that solves the utility equation $U(q) = 0$. Now notice that $\Lambda W(\Lambda)$ is precisely the expected delay cost rate, which is convex in the arrival rate parameter. Hence, for mean-MAD information, we can obtain tight bounds for the expected social welfare, even if the underlying arrival process conditioned on the event $\Lambda = \lambda$ is not a Poisson process, but instead, an arrival process which is characterized by the uncertain parameter Λ.

The applications above are just three examples of a vast range of uses. In principle, all queueing systems for which the steady-state performance measures are convex functions of the parameters are amenable to minimax analysis with MAD dispersion. However, when it comes to variance, convexity on its own is insufficient. The second-order bounds necessitate the third-order property, a convex derivative. Here, a sample-path method appears to be impractical for many models. To see this, consider the evolution of the waiting time as in (3.18). Clearly, the underlying sample path is not continuously differentiable in λ (due to the $(\cdot)^+$ operator). Despite the fact that the derivative does not exist in the sample-path sense, *ad hoc* approaches, in which one works directly with algebraic formulae for the steady-state expected wait, will still work. As a concrete example of this direct approach, consider the Pollaczek-Khinchine formula for the expected wait in the M/G/1 queue, given by

$$W_*(\lambda) = \frac{\rho + \lambda\mu\,\mathrm{Var}(S)}{2(\mu - \lambda)} + \mu^{-1}.$$

Now observe that

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}W_*(\lambda) = \frac{1 + \mu^2\,\mathrm{Var}(S)}{2(\lambda - \mu)^2} \geqslant 0, \qquad \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}W_*(\lambda) = \frac{1 + \mu^2\,\mathrm{Var}(S)}{(\mu - \lambda)^3} \geqslant 0,$$

and moreover,

$$\frac{\mathrm{d}^3}{\mathrm{d}\lambda^3}W_*(\lambda) = \frac{3(1 + \mu^2\,\mathrm{Var}(S))}{(\lambda - \mu)^4} \geqslant 0, \quad \text{for } \lambda < \mu.$$

Hence, $W_*(\lambda)$ is increasingly convex, and $\mathbb{E}[W_*(\Lambda)]$ can be bounded using Lemma 3.2, yielding the following result.

Proposition 3.15 **(M/G/1 queue with variance information).** *Let the market size Λ have a distribution $\mathbb{P}$ that resides in the ambiguity set $\mathcal{P}(m, \sigma, \underline{\lambda}, \overline{\lambda})$, and let $W_*(\lambda)$ denote the expected wait in the M/G/1 queue. Then, the tight upper bound for $\mathbb{E}_{\mathbb{P}}[W_*(\Lambda)]$ is attained by the two-point distribution stated in* Theorem 3.4.

Thus, despite the need for some tedious algebra, the second-order bounds that incorporate variance information have the potential to be widely applicable to other stochastic systems.

3.5.3. Random service rate

In practice, it might turn out that other parameters besides the arrival rate are uncertain, with only partial information available. It could be advantageous to establish second-order bounds for the service rate at which the different servers operate. It seems reasonable to assume that decision makers may not always have precise knowledge of server rates, as these rates could depend on specific server characteristics. In this regard, we consider the scenario in which the service rate μ is replaced by a random service rate M. Convexity of the mean queue size of the M/M/s queue with respect to the system utilization ρ is a classic result; see e.g. [91] and [142]. The following result pertains to the multi-server setting, where we again will rely on the results in [182].

Lemma 3.16 **(Derivative w.r.t. μ is concave).** *The expected wait $W(s,\lambda,\mu)$ as function of μ has a concave derivative.*

Proof. Denote $\phi_\mu(x) := W(s,\lambda,x)$. To prove that $\phi'_\mu(x)$ is concave, we will use the result in Randhawa [182] which proves convexity of L' as function of ρ. Interpreting L indeed as function of ρ, define $h(x) := L$ and $g(\mu) := \lambda/(s\mu)$, where μ is the variable and the parameters λ and s are fixed to arbitrary constants. To see that the composite function $(h \circ g)(\mu) = L(\lambda/(s\mu))$ is decreasingly convex, notice that

$$\frac{\partial^3}{\partial \mu^3}(h\circ g)(\mu) = \underbrace{g'''(\mu)h'(g(\mu))}_{g'''<0,h'\geqslant 0} + \underbrace{g'(\mu)^3 h'''(g(\mu))}_{g'<0,h'''\geqslant 0} + \underbrace{3g'(\mu)g''(\mu)h''(g(\mu))}_{g'<0,g''>0,h''\geqslant 0} \leqslant 0.$$

Little's law indicates that $W(s,\lambda,\mu) = L(s,\lambda,\mu)/\lambda$, with $\lambda > 0$ a fixed constant, which completes the proof. □

Hence, $W(s,\lambda,\mu)$ is a decreasing, convex function of μ with a concave derivative. By Lemma 3.2(i), the extremal distribution is given by a two-point distribution with support $\{\mu_1,\mu_2\} = \{\underline{\mu}, m_\mu + \frac{\sigma^2}{m_\mu - \underline{\mu}}\}$ and respective probabilities

$$p_1 = \frac{\sigma^2}{(m_\mu - \underline{\mu})^2 + \sigma^2}, \quad p_2 = \frac{(m_\mu - \underline{\mu})^2}{(m_\mu - \underline{\mu})^2 + \sigma^2}.$$

That is, the worst case is given by servers which either operate at their lowest service capacity or just above the mean service rate.

Moreover, it is possible to work with multiple uncertain parameters simultaneously. We next consider

$$\max_{\mathbb{P}_\Lambda \in \mathcal{P}_{(m_\Lambda,\sigma_\Lambda)}} \max_{\mathbb{P}_M \in \mathcal{P}_{(m_M,\sigma_M)}} \mathbb{E}_{\mathbb{P}_M \otimes \mathbb{P}_\Lambda}[W(s,\Lambda,M)],$$

which turns out to be solvable in closed form:

Theorem 3.17 **(M/M/s queue with partially known service rate and market size).** *Let the service rate M follow a marginal distribution $\mathbb{P}_M \in \mathcal{P}_M := \mathcal{P}_{(m_M,\sigma_M)}$ and the market size Λ follow a marginal distribution $\mathbb{P}_\Lambda \in \mathcal{P}_\Lambda := \mathcal{P}_{(m_\Lambda,\sigma_\Lambda)}$. Suppose that $\overline{\lambda} < s\underline{\mu}$. If M and Λ are independent,* *the sharpest possible upper bound for* $\mathbb{E}_{\mathbb{P}}[W(s,\Lambda,M)]$ *is attained by the product measure* $\mathbb{P}_M^* \otimes \mathbb{P}_\Lambda^*$, *with* $\mathbb{P}_\Lambda^*$ *and* $\mathbb{P}_M^*$ *as defined in* Lemma 3.2(i) *and* (ii), *respectively.*

Proof. The tower rule yields

$$\max_{\mathbb{P}_M \in \mathscr{P}_M} \mathbb{E}_{\mathbb{P}}[W(s,\Lambda,M)] = \max_{\mathbb{P}_M \in \mathscr{P}_M} \mathbb{E}_M[\mathbb{E}_\Lambda[W(s,\Lambda,M)|M]],$$

where the expectation is well defined due to $s\underline{\mu} > \overline{\lambda}$. Hence, fixing the distribution of Λ gives the moment problem

$$\max_{\mathbb{P}_M \in \mathscr{P}_M} \int_{[\underline{\mu},\overline{\mu}]} \mathbb{E}_\Lambda[W(s,\Lambda,\mu)]\mathrm{d}\mathbb{P}(\mu).$$

For λ a fixed constant, $W(s,\Lambda,\mu)$ has the required "curvature" properties, so $\mathbb{P}_M^*$ is the solution to the moment problem. For $\mathbb{P}_M^*$ to work also for $\phi_\mu(x) := \mathbb{E}_\Lambda[W(s,\Lambda,x)]$, we need to check the signs of the first three derivatives. We interchange differentiation and expectation operations (which is allowed as Λ is bounded), so that

$$\phi_\mu^{(k)}(x) = \frac{\mathrm{d}^{(k)}}{\mathrm{d}x^{(k)}}\mathbb{E}_\Lambda[W(s,\Lambda,x)] = \mathbb{E}_\Lambda\left[\frac{\mathrm{d}^{(k)}}{\mathrm{d}x^{(k)}}W(s,\Lambda,x)\right].$$

As the expectation operator is nonnegative,

$$\frac{\mathrm{d}^{(k)}}{\mathrm{d}x^{(k)}}W(s,\lambda,x) \geqslant 0,\ \forall\lambda \implies \mathbb{E}_\Lambda\left[\frac{\mathrm{d}^{(k)}}{\mathrm{d}x^{(k)}}W(s,\Lambda,x)\right] \geqslant 0,$$

and likewise,

$$\frac{\mathrm{d}^{(k)}}{\mathrm{d}x^{(k)}}W(s,\Lambda,x) \leqslant 0,\ \forall\lambda \implies \mathbb{E}_\Lambda\left[\frac{\mathrm{d}^{(k)}}{\mathrm{d}x^{(k)}}W(s,\Lambda,x)\right] \leqslant 0,$$

which proves that $\mathbb{P}_M^*$ maximizes $\phi_\mu(x)$ for any distribution of Λ. Substituting the two-point distribution gives the moment problem

$$\max_{\mathbb{P}_\Lambda \in \mathscr{P}_\Lambda} \mathbb{E}_{\mathbb{P}_\Lambda}[p_1 W(s,\Lambda,\mu_1) + p_2 W(s,\Lambda,\mu_2)].$$

Based on the same line of reasoning, this expression is maximized by $\mathbb{P}_\Lambda^*$. □

Theorem 3.17 demonstrates another favorable property of our second-order bounds. Namely, the extremal distributions are independent of the precise form of $W(s,\lambda,\mu)$, allowing us to apply the univariate results recursively when considering multiple uncertain parameters.

3.6. Conclusions and outlook

Let us conclude. We have derived novel second-order bounds for the expected wait in an M/M/s queueing system driven by a mixed Poisson process, in which the arrival rate parameter is inherently a random variable. We have been able to derive these tight bounds for the M/M/s queue by exploiting the unique curvature properties of its steady-state queueing metrics, in conjunction with semi-infinite programming techniques from distributionally robust analysis. Furthermore, we have demonstrated that within a rational queueing context, these bounds can be used to determine the optimal maximin price for a firm catering to a market of delay-sensitive, rational customers with limited information about the total market size.

Looking forward, it seems worthwhile to search for other information sets that exhibit useful properties similar to variance and MAD, which proved crucial in our distributionally robust analyses. It can also be interesting to consider other data-driven ambiguity sets, such as the Wasserstein ambiguity set discussed in Section 3.4.3, which are equipped to handle limited data, as estimates of means and dispersion measures might not always be statistically accurate enough. Additionally, it is insightful to investigate to which other stochastic systems analogous types of distributionally robust techniques to control for parameter uncertainty can be applied, a topic we briefly touched upon in Section 3.5.2.

4

Tight tail probability bounds for distribution-free decision making

4.1. Introduction

Chebyshev's inequality provides an upper bound on the tail probability of a random variable using only its first two moments [31, 42]. This inequality is tight, meaning it cannot be improved in general. However, Chebyshev's inequality can be criticized for only being attained by pathological distributions that abuse the unboundedness of the underlying support. Indeed, the worst-case distribution takes values on merely two support points, which can be regarded unrealistic [214]. A variant of the Chebyshev inequality that was already considered in [86] restricts the distributions it considers to be unimodal. This yields an improvement by a factor $\frac{4}{9}$ over the classical Chebyshev inequality. This idea of including unimodality has been extended to the multivariate case recently as well [214].

Multivariate generalizations of Chebyshev's inequality have also been studied. In [29] and [215] generalizations are studied through formulating a convex optimization problem, given that the prescribed confidence region can be described by polynomial or linear and quadratic inequalities, respectively. In [92] on the other hand, closed-form variants of Chebyshev's inequality are provided for different dispersion measures than the variance. Generalized versions

This chapter is based on the research paper [188].

of Chebyshev's inequality for products of random variables that focus on a one-sided inequality have also received some attention recently [190].

All the above-mentioned inequalities, however, still assume an unbounded support. In many practical applications some information on the minimum and maximum of uncertain parameters is known. This is particularly true for OR applications that consider uncertain parameters that are known to be nonnegative, such as inventory management, service operations, appointment scheduling and pricing mechanisms. De Schepper and Heijnen [64] derived tail probability bounds that incorporate the upper bound of the random variable's range. These bounds are attained by discrete distributions, supported on two or three atoms.

Next to restricting the support to be contained in a bounded interval, a second potential improvement with this novel setup compared to Chebychev's inequality concerns robustness for outliers. Whereas outliers greatly influence the (sample) variance, the mean absolute deviation from the mean (MAD) is less sensitive for large deviations from the mean, and hence a potentially more robust measure of statistical dispersion in data. We therefore propose to replace variance with MAD. Using the MAD comes with additional advantages. We show that the set of extremal distributions for which the derived tail bounds are tight is more varied than a single pathological distribution: it consists of an infinite number of mixed distributions instead. Second, because the MAD is a linear function, it allows for elegant closed-form bounds, a feature we shall leverage when applying the bounds to domain-specific OR questions.

The solution to a generalized moment problem will give a tight upper bound on the tail probability of all random variables with a given bounded support, mean and MAD. This new robust bound is of a similar simplicity and generality as the original Chebyshev inequality, and can therefore be used in various applications. The worst-case distribution that achieves the tight bounds is, however, more complicated than the two-point distribution of the Chebyshev inequality, and is a mixed distribution with up to three discrete parts and one continuous part. We also derive a number of additional tail probability bounds: the tight lower bound for mean, MAD and bounded support information, and the tight upper bound where we condition on the median and the mean absolute deviation from the median.

Recent advances in distributionally robust optimization (DRO) also exploit MAD-based ambiguity sets to obtain closed-form expressions for stochastic quantities such as the minimum and maximum expectation of a convex function [87, 179]. These closed-form expressions are then used to solve minimax and maximin optimization problems that arise naturally in decision making under uncertainty. Postek et al. [179] specifically used results from Ben-Tal and Hochman [19] on tight upper and lower bounds on the expectation of a convex function of a random variable.

4.1.1. Contributions and outline

This chapter presents the first closed-form solution for the problem with a combination of mean, MAD and support constraints, and a *nonconvex* objective function. This proof method is not restricted to the indicator function that models the tail probability and works for a much larger class of (measurable) functions, as explained in Section 1.2.1.

We apply the robust bounds for distribution-free analysis of three applications that can be subjected to minmax or maxmin optimization. We start with the newsvendor model, the basic single-period inventory model that searches for the optimal order quantity in view of overage and underage costs. The second application is stop-loss reinsurance, in which an insurance company faces a claim which it pays up to a predefined level, while the reinsurance company covers the remainder. We study this problem from both the insurer's and reinsurer's perspective, the latter of which requires an extension of our tail probability bound. Last, we study a continuous optimization problem from radiotherapy optimization with an ambiguous chance constraint. Application of the derived tail probability bound yields a computationally tractable convex reformulation that can be solved with traditional methods. The three applications illustrate different aspects of the derived tail probability and the primal-dual proof used to obtain it. First, the newsvendor example is a direct application of the bound to a classical OR problem. Second, the stop-loss reinsurance application illustrates how the primal-dual proof technique can be extended to more complex functionals than the tail probability. Third, the radiotherapy optimization example highlights the connection of our result to the field of distributionally robust optimization. We should say that the models have been chosen somewhat arbitrarily, and there are many other OR questions where tail probability bounds under mean-MAD constraints can prove useful.

The chapter is organized as follows. We present the tail probability bounds in Section 4.2 and the three applications in Section 4.3. Section 4.4 presents some conclusions and several directions for future research.

4.2. Tail probability bounds

In this section, we derive novel bounds for the tail probability $\mathbb{P}(X \geqslant t)$ of a random variable X. We solve the semi-infinite linear program (LP) in Section 4.2.1. In Section 4.2.2, we derive more bounds, based on different ambiguity sets. We compare the novel bounds with some existing bounds in Section 4.2.3.

4.2.1. Tight lower and upper bounds

Let the ambiguity set $\mathscr{P}_{(\mu,\sigma)}$ contain all distributions with a given mean μ and variance σ^2, and let the random variable X follows some distribution $\mathbb{P} \in \mathscr{P}_{(\mu,\sigma)}$. Chebyshev's inequality (the one-sided version also known as Cantelli's inequality) then follows from the worst-case distribution that solves the optimization problem $\sup_{X\sim\mathbb{P}\in\mathscr{P}_{(\mu,\sigma)}} \mathbb{E}_{\mathbb{P}}\left[\mathbb{1}\{X \geqslant t\}\right]$, yielding the upper bound

$$\mathbb{P}\left(X \geqslant t\right) \leqslant \frac{\sigma^2}{\sigma^2 + (t-\mu)^2}. \tag{4.1}$$

This worst-case distribution takes only values on the points $\mu - \sigma^2/(t-\mu)$ and t (with probabilities $(t-\mu)^2/(\sigma^2+(t-\mu)^2)$ and $\sigma^2/(\sigma^2+(t-\mu)^2)$, resp.), which can be regarded as conservative. In obtaining our novel robust tail bounds, we instead need to solve

$$\sup_{X\sim\mathbb{P}\in\mathscr{P}_{(\mu,b,d)}} \mathbb{E}_{\mathbb{P}}\left[\mathbb{1}\{X \geqslant t\}\right], \tag{4.2}$$

with $\mathscr{P}_{(\mu,b,d)}$ the ambiguity set that contains all distributions with a given mean μ, support $[0,b]$ and mean absolute deviation d, i.e.,

$$\mathscr{P}_{(\mu,b,d)} = \{\mathbb{P} : \mathscr{B} \to [0,1] \,|\, \mathbb{P}(X \in [0,b]) = 1,\ \mathbb{E}_{\mathbb{P}}[X] = \mu,\ \mathbb{E}_{\mathbb{P}}[|X-\mu|] = d\} \tag{4.3}$$

with $\mathscr{B}$ the Borel σ-algebra of the closed set $[0,b]$. Optimization problem (4.2) is a semi-infinite LP that is reminiscent of those arising in moment problems, and typically does not allow for an analytic (closed-form) solution. However, using the MAD-based ambiguity set, the dual program to (4.2) can be solved explicitly. Since $\mathbb{P}$ is a probability measure it should satisfy the constraint $\int_{x\in[0,b]} \mathrm{d}\mathbb{P}(x) = 1$. Moreover, this probability measure should satisfy the mean and MAD constraints $\int_{x\in[0,b]} x\, \mathrm{d}\mathbb{P}(x) = \mu$ and $\int_{x\in[0,b]} |x-\mu|\, \mathrm{d}\mathbb{P}(x) = d$. Under these constraints, we solve the semi-infinite linear program

$$\sup_{\mathbb{P}\in\mathscr{P}_{(\mu,b,d)}} \int_x \mathbb{1}_{\{x\geqslant t\}}\, \mathrm{d}\mathbb{P}(x), \tag{4.4}$$

which gives our first main result.

THEOREM 4.1. *Consider a random variable X with a distribution $\mathbb{P}$ in $\mathscr{P}_{(\mu,b,d)}$. Then,*

$$\sup_{\mathbb{P}\in\mathscr{P}_{(\mu,b,d)}} \mathbb{P}(X \geqslant t) = \sup_{\mathbb{P}\in\mathscr{P}_{(\mu,b,d)}} \mathbb{P}(X > t) = \begin{cases} 1, & t \in [0,\tau_1], \\ \frac{\mu}{t} - \frac{d(b-t)}{2t(b-\mu)}, & t \in [\tau_1,\mu], \\ 1 - \frac{d}{2\mu}, & t \in [\mu,\tau_2], \\ \frac{d}{2(t-\mu)}, & t \in [\tau_2,b], \end{cases} \tag{4.5}$$

with τ_1 and τ_2 given by

$$\tau_1 = \mu - \frac{d(b-\mu)}{2(b-\mu)-d}, \quad \tau_2 = \mu + \frac{d\mu}{2\mu-d}.$$

Proof. Let $\mathscr{M}^+$ be the set of non-negative measures defined on the measurable space $([0,b],\mathscr{B})$. We need to solve

$$\begin{aligned} \sup_{\mathbb{P}\in\mathscr{M}^+} \quad & \int_x \mathbb{1}_{\{x\geqslant t\}}\, \mathrm{d}\mathbb{P}(x) \\ \text{s.t.} \quad & \int_x \mathrm{d}\mathbb{P}(x) = 1,\ \int_x x\, \mathrm{d}\mathbb{P}(x) = \mu,\ \int_x |x-\mu|\, \mathrm{d}\mathbb{P}(x) = d. \end{aligned} \tag{4.6}$$

Consider the dual of (4.6),

$$\begin{aligned} \inf_{\lambda_0,\lambda_1,\lambda_2} \quad & \lambda_0 + \lambda_1\mu + \lambda_2 d \\ \text{s.t.} \quad & \mathbb{1}_{\{x\geqslant t\}} \leqslant \lambda_0 + \lambda_1 x + \lambda_2|x-\mu| =: F(x),\ \forall x \in [0,b]. \end{aligned} \tag{4.7}$$

The constraint of the dual problem requires $F(\cdot)$ to majorize $\mathbb{1}\{x \geqslant t\}$. Note that $F(\cdot)$ has a "kink" at $x = \mu$, that is, $F(\cdot)$ is piecewise linear and can only change direction in $x = \mu$. Solving (4.7)

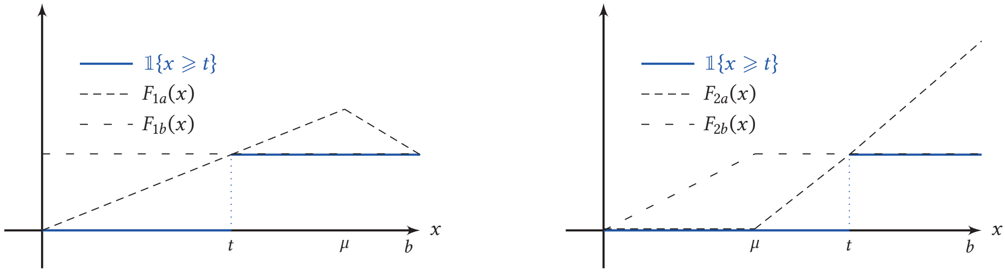


Figure 4.1: The majorizing functions under the different scenarios

boils down to finding the tightest majorant. We have four candidates for the solution, which are depicted in Figure 4.1. When $t \in [0, \mu]$, $F(x)$ touches $\mathbb{1}\{x \geqslant t\}$ in $\{0, t, b\}$ (scenario 1a), or $F(x) = 1$ and touches $\mathbb{1}\{x \geqslant t\}$ in $[t, b]$ (scenario 1b). When $t \in [\mu, b]$, $F(x)$ touches in $[0, \mu] \cup \{t\}$ (scenario 2a) or in $\{0\} \cup [t, b]$ (scenario 2b).

Scenario 1a implies $F(0) = 0$, $F(t) = F(b) = 1$, which gives as dual solution

$$\lambda_2 = -\frac{b - t}{2t(b - \mu)}, \quad \lambda_1 = \frac{1}{t} + \lambda_2, \quad \lambda_0 = -\lambda_2 \mu,$$

and objective value

$$\lambda_0 + \lambda_1 \mu + \lambda_2 d = \frac{\mu}{t} - \frac{d(b - t)}{2t(b - \mu)}.$$

The next step is to find a feasible solution for the primal problem that yields the same objective value as the solution to the dual problem. By weak duality of semi-infinite linear programming, a feasible solution to the dual problem gives a valid upper bound for the optimal primal solution value. A feasible primal solution with an objective value equal to this upper bound results in strong duality. Next, we provide a constructive approach to find such a primal solution. Assume that strong duality holds. The primal maximizer $\mathbb{P}^*$ and the dual minimizer $(\lambda_0^*, \lambda_1^*, \lambda_2^*)$ are then related as

$$\int_x \mathbb{1}\{x \geqslant t\} \, \mathrm{d}\mathbb{P}^*(x) = \int_x (\lambda_0^* + \lambda_1^* x + \lambda_2^* |x - \mu|) \, \mathrm{d}\mathbb{P}^*(x). \tag{4.8}$$

Due to dual feasibility, we must have that $\lambda_0^* + \lambda_1^* \mu + \lambda_2^* d - \mathbb{1}\{x \geqslant t\} \geqslant 0$ pointwise for each $x \in [0, b]$. This inequality combined with equation (4.8) is also known as the complementary slackness relation in (semi-infinite) linear programming. Complementary slackness implies that the worst-case probability distribution is supported on the points where the dual function $F^*(x) = \lambda_0^* + \lambda_1^* x + \lambda_2^* |x - \mu|$ coincides with the indicator function $\mathbb{1}\{x \geqslant t\}$. For scenario 1a we have one unique option, a three-point distribution on $\{0, t, b\}$. The corresponding optimal probabilities of (4.6) follow from solving

$$p_0 + p_t + p_b = 1, \quad p_t t + p_b b = \mu, \quad p_0 \mu + p_t(\mu - t) + p_b(b - \mu) = d.$$

This gives

$$p_0 = 1 - p_t - p_b, \quad p_t = \frac{\mu}{t} - \frac{bd}{2t(b - \mu)}, \quad p_b = \frac{d}{2(b - \mu)}, \tag{4.9}$$

and hence

$$\int_x \mathbb{1}\{x \geqslant t\}\,\mathrm{d}\mathbb{P}(x) = p_t + p_b = \frac{\mu}{t} - \frac{d(b-t)}{2t(b-\mu)}.$$

Since strong duality holds as the primal and dual objective values agree, (4.9) is the optimal solution.

Scenario 1b implies $F(0) = F(t) = F(b) = 1$ and hence $\lambda_0 = 1$, $\lambda_1 = \lambda_2 = 0$ with objective value 1. One feasible primal solution is e.g. $p_t = \frac{d}{2(\mu-t)}$, $p_b = \frac{d}{2(b-\mu)}$, $p_\mu = 1 - p_t - p_b$, with objective 1. Note that this solution is not a unique optimum, as the dual function $F^*_{1b}(x)$ coincides with $\mathbb{1}\{x \geqslant t\}$ on the entire interval $[t,b]$. Therefore, one can construct an arbitrary (discrete, continuous or mixed) probability distribution with support on the interval $[t,b]$, which then serves as the worst-case distribution, as long as the mean and MAD conditions are satisfied.

Scenario 2a implies $F(0) = F(\mu) = 0$, $F(t) = 1$, which gives

$$\lambda_1 = \lambda_2 = \frac{1}{2(t-\mu)}, \quad \lambda_0 = -\frac{\mu}{2(t-\mu)},$$

and objective value

$$\lambda_0 + \lambda_1\mu + \lambda_2 d = \frac{d}{2(t-\mu)}.$$

Solving the optimal probabilities of (4.6), where we take $\{0,\mu,t\}$ for the support of the worst-case distribution, indeed confirms that $p_t = \frac{d}{2(t-\mu)}$.

Scenario 2b gives $F(0) = 0$, $F(\mu) = F(b) = 1$, which results in

$$\lambda_0 = \frac{1}{2}, \quad \lambda_1 = \frac{1}{2\mu}, \quad \lambda_2 = -\frac{1}{2\mu},$$

and dual objective value

$$\lambda_0 + \lambda_1\mu + \lambda_2 d = 1 - \frac{d}{2\mu}.$$

Solving (4.6) with support $\{0,t,b\}$ confirms that $p_0 = \frac{d}{2\mu}$.

The proof is then completed by finding the minimum for each scenario and determining the values of τ_1 and τ_2 for scenarios 1 and 2, respectively. We remark that the proof is identical for the strict inequality. Because the majorant is a continuous function, it is irrelevant whether the indicator function that is majorized is lower or upper semi-continuous. □

We mention some noteworthy characteristics of the bound in Theorem 4.1. The bound is continuous in $t = \mu$. If the support is symmetric around μ, then the worst-case probability is at least $1/2$ for $t \in [0,\mu]$. The upper bound for $t \in [\mu,b]$ is increasing for $d \leqslant 2\mu(t-\mu)/t$ and decreasing for larger values of d. This last observation in particular is interesting as one might anticipate the bound to increase with MAD. This also implies that when MAD is unknown, the worst-case probability based on only the support and mean is given by the result of Theorem 4.1 for $d = 2\mu(t-\mu)/t$. This indeed returns Markov's inequality. We also mention that the support information $[0,b]$ can easily be extended to $[a,b]$ with $a \in \mathbb{R}$ by shifting the distribution accordingly. The tail bounds for the second and third interval then change into

$$\frac{\mu - a}{t - a} - \frac{d(b-t)}{2(t-a)(b-\mu)} \quad \text{and} \quad 1 - \frac{d}{2(\mu - a)}, \tag{4.10}$$

respectively. Similarly, the result can be adapted to a support that is only bounded from below or above. For such supports, one of the cases in (4.5) disappears. Specifically, when the support of X is given by $[0, \infty)$, it follows that $\tau_1 = \mu$, while for the support $(-\infty, b]$, it follows that $\tau_2 = \mu$.

For a tight lower bound on $\mathbb{P}(X > t)$, we can use the results and the remark above on a slightly altered version of the input. The idea is formalized in the following result:

Corollary 4.2. *For a random variable X with a distribution $\mathbb{P}$ in $\mathcal{P}_{(\mu,b,d)}$,*

$$\inf_{\mathbb{P}\in\mathcal{P}_{(\mu,b,d)}} \mathbb{P}(X \geqslant t) = \inf_{\mathbb{P}\in\mathcal{P}_{(\mu,b,d)}} \mathbb{P}(X > t) = \begin{cases} 1 - \frac{d}{2(\mu-t)}, & t \in [0, \tau_1], \\ \frac{d}{2(b-\mu)}, & t \in [\tau_1, \mu], \\ \frac{\mu-t}{b-t} + \frac{dt}{2\mu(b-t)}, & t \in [\mu, \tau_2], \\ 0, & t \in [\tau_2, b], \end{cases} \tag{4.11}$$

with

$$\tau_1 = \mu - \frac{d(b-\mu)}{2(b-\mu)-d}, \quad \tau_2 = \mu + \frac{d\mu}{2\mu - d}.$$

Proof. We reformulate the infimum as follows:

$$\begin{aligned} \inf_{\mathbb{P}\in\mathcal{P}_{(\mu,b,d)}} \mathbb{P}(X > t) &= 1 - \sup_{\mathbb{P}\in\mathcal{P}_{(\mu,b,d)}} \mathbb{P}(X \leqslant t) \\ &= 1 - \sup_{\mathbb{P}\in\tilde{\mathcal{P}}_{(\mu,\tilde{b},d)}} \mathbb{P}(X \geqslant \tilde{t}), \end{aligned}$$

where

$$\tilde{\mathcal{P}}_{(\mu,\tilde{b},d)} = \{\mathbb{P} : \mathcal{B} \to [0,1] \,|\, \mathbb{P}(X \in [\tilde{a}, \tilde{b}]) = 1,\ \mathbb{E}_{\mathbb{P}}[X] = \mu,\ \mathbb{E}_{\mathbb{P}}[|X - \mu|] = d\},$$

and $\tilde{a} = 2\mu - b$, $\tilde{b} = 2\mu$ and $\tilde{t} = 2\mu - t$. This transformation essentially flips the support around μ such that $\mu - 0 = \tilde{b} - \mu$ and $b - \mu = \mu - \tilde{a}$. Therefore, the maximum probability below t is equal to the maximum probability above the flipped threshold $\tilde{t}$. Plugging in the results from Theorem 4.1 for $t \in (a, b]$ then yields (4.11). Similarly, the result for $\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}(X \geqslant t)$ can be obtained. □

We now describe in more detail the worst-case distributions that are revealed in the proof of Theorem 4.1.

Corollary 4.3. *Consider the set of worst-case distributions*

$$\mathcal{P}^* = \operatorname*{argsup}_{\mathbb{P}\in\mathcal{P}_{(\mu,b,d)}} \mathbb{E}_{\mathbb{P}}[\mathbb{1}\{X \geqslant t\}].$$

Then,

(i) *if* $t \in [0, \tau_1]$, $\mathscr{P}^* = \{\mathbb{P} \in \mathscr{P}_{(\mu,b,d)} \,|\, \mathbb{P}(X \in [t,b]) = 1\}$, *all distributions in* $\mathscr{P}_{(\mu,b,d)}$ *that are supported on the interval* $[t,b]$*;*

(ii) *if* $t \in [\tau_1, \mu]$, $\mathscr{P}^* = \{\mathbb{P} : \mathbb{P}(X = 0) = 1 - \frac{\mu}{t} + \frac{d(b-t)}{2t(b-\mu)}, \mathbb{P}(X = t) = \frac{\mu}{t} - \frac{bd}{2t(b-\mu)}, \mathbb{P}(X = b) = \frac{d}{2(b-\mu)}\}$, *the three-point distribution as derived in scenario 1a in the proof of Theorem 4.1;*

(iii) *if* $t \in [\mu, \tau_2]$, $\mathscr{P}^* = \{\mathbb{P} \in \mathscr{P}_{(\mu,b,d)} \,\big|\, \mathbb{P}(X = 0) = \frac{d}{2\mu}, \mathbb{P}(X \in [t,b]) = 1 - \frac{d}{2\mu}\}$, *all discrete/mixed distributions with probability mass* $\frac{d}{2\mu}$ *on 0 and the remainder of its probability mass supported on* $[t,b]$*;*

(iv) *if* $t \in [\tau_2, b]$, $\mathscr{P}^* = \{\mathbb{P} \in \mathscr{P}_{(\mu,b,d)} \,\big|\, \mathbb{P}(X = t) = \frac{d}{2(t-\mu)}, \mathbb{P}(X \in [0,\mu]) = 1 - \frac{d}{2(t-\mu)}\}$, *all discrete/mixed distributions with probability mass* $\frac{d}{2(t-\mu)}$ *on t and the remainder of its probability mass supported on* $[0,\mu]$.

Proof. The proof follows almost directly from the complementary slackness relation explained in the proof of Theorem 4.1. For $t \in [0, \tau_1]$ the dual solution function coincides with $\mathbb{1}\{x \geqslant t\}$ on the interval $[t,b]$. Hence, all distributions that are supported on this interval and comply with the mean and MAD requirements are possible candidates for the worst-case distribution. Next, one can apply a similar reasoning for $t \in [\mu, \tau_2]$ and $t \in [\tau_2, b]$. The worst-case distribution can exist on the range where the dual solution function $F^*(x)$ and the indicator function coincide. To attain the same optimal value, the probability mass on the singletons is chosen accordingly. Finally, note that the second case is already shown in the proof of Theorem 4.1. □

Observe that when t equals τ_1, μ, or τ_2, there is only a single discrete extremal distribution. Figure 4.2 provides examples of the worst-case distributions for several different parameter settings and values of t. For the sake of exposition, all depicted examples have a continuous part that is uniform over its supported interval. Corollary 4.3 shows that the ambiguity set $\mathscr{P}_{(\mu,b,d)}$ results in a non-trivial collection of worst-case distributions; that is, the mean-MAD approach results in a set that does not solely include discrete distributions with a small number of atoms for $t \notin [\tau_1, \mu] \cup \{\tau_2\}$.

4.2.2. Sharp bounds for other types of ambiguity

The primal-dual method is a general approach, often used in DRO, with a much wider range of possible applications. This subsection demonstrates that the semi-infinite programming problems can be adapted to different ambiguity sets, thus incorporating other types of information. We first derive alternative (tight) bounds for the tail probability, where we use a different measure of central tendency: the median. We then turn back to mean-MAD information and derive sharper bounds with a commonly used skewness measure that complements the mean-MAD ambiguity set, i.e., $\mathbb{P}(X \geqslant \mu)$.

For the first adjustment, assume we know the following parameters: the support $[0,b]$, the median m and the mean absolute deviation (from the median) d_m. Let us suppose further, for simplicity, that the median is uniquely defined. We now obtain a different set of distributions,

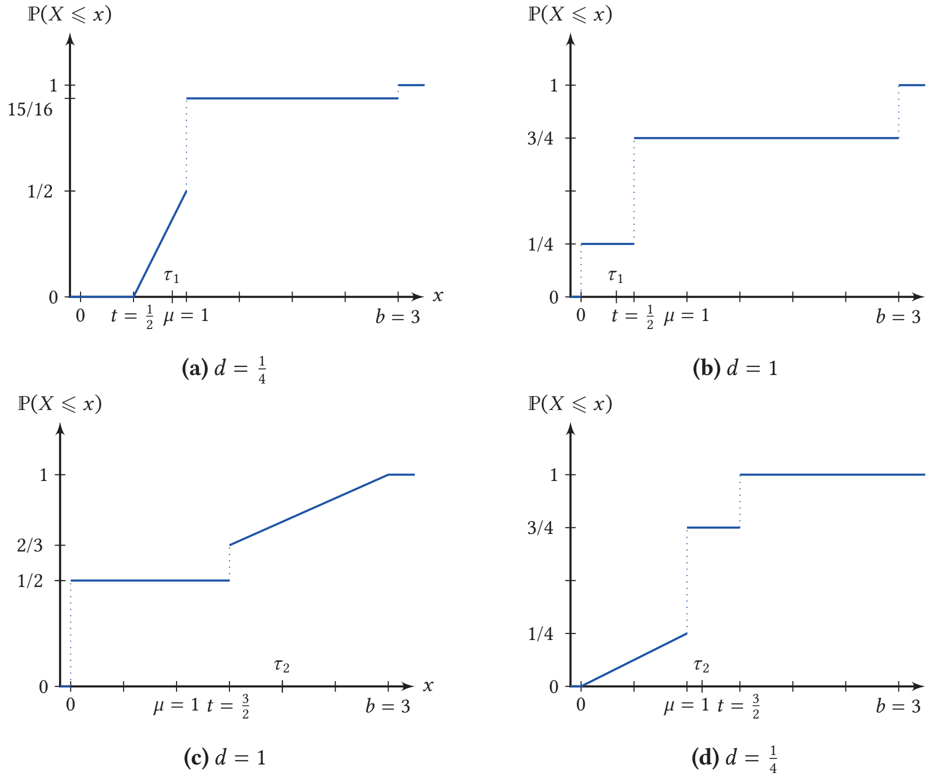


Figure 4.2: Examples of the extremal distributions that attain the tail probability bound as described in Corollary 4.3

namely,

$$\mathscr{P}_{(m,b,d_m)} = \Big\{ \mathbb{P} : \mathscr{B} \to [0,1] \,|\, \mathbb{P}(X \in [0,b]) = 1,\ \mathbb{P}(X \geqslant m) \geqslant \frac{1}{2}, \\ \mathbb{P}(X \leqslant m) \geqslant \frac{1}{2},\ \mathbb{E}_{\mathbb{P}}[|X - m|] = d_m \Big\}.$$

If the distribution of X resides in this ambiguity set, the tight bounds are given by the optimal value of

$$\begin{aligned} \sup_{\mathbb{P} \in \mathscr{M}^+} \quad & \int_x \mathbb{1}_{\{x \geqslant t\}} \mathrm{d}\mathbb{P}(x) \\ \text{s.t.} \quad & \int_x \mathrm{d}\mathbb{P}(x) = 1,\ \int_x \mathbb{1}_{\{x \leqslant m\}}\, \mathrm{d}\mathbb{P}(x) \geqslant \frac{1}{2}, \\ & \int_x \mathbb{1}_{\{x \geqslant m\}}\, \mathrm{d}\mathbb{P}(x) \geqslant \frac{1}{2},\ \int_x |x - m|\, \mathrm{d}\mathbb{P}(x) = d_m. \end{aligned} \tag{4.12}$$

This then gives the dual problem

$$\begin{aligned} \inf_{\substack{\lambda_0,\lambda_2;\\ \lambda_1^-,\lambda_1^+\leqslant 0}} \quad & \lambda_0 + (\lambda_1^- + \lambda_1^+)\frac{1}{2} + \lambda_2 d_m \\ \text{s.t.} \quad & \mathbb{1}_{\{x\geqslant t\}} \leqslant \lambda_0 + \lambda_1^- \mathbb{1}_{\{x\leqslant m\}} + \lambda_1^+ \mathbb{1}_{\{x\geqslant m\}} + \lambda_2 |x-m| =: F_m(x), \ \forall x \in [0,b]. \end{aligned} \tag{4.13}$$

We can solve the dual problem by exploiting the structure of $F_m(x)$, like with the proof of Theorem 4.1. The different scenarios are depicted in Figure 4.3. The following theorem presents the median-MAD bounds. The details of the proof are relegated to the appendix.

Theorem 4.4. *For a random variable X with a distribution $\mathbb{P} \in \mathscr{P}_{(m,b,d_m)}$,*

$$\sup_{\mathbb{P}\in\mathscr{P}_{(m,b,d_m)}} \mathbb{P}(X \geqslant t) = \begin{cases} \inf\left\{1, \frac{b-2d_m}{2t} + \frac{1}{2}\right\}, & t \in [0,m), \\ \inf\left\{\frac{1}{2}, \frac{d_m}{(t-m)}\right\}, & t \in [m,b]. \end{cases} \tag{4.14}$$

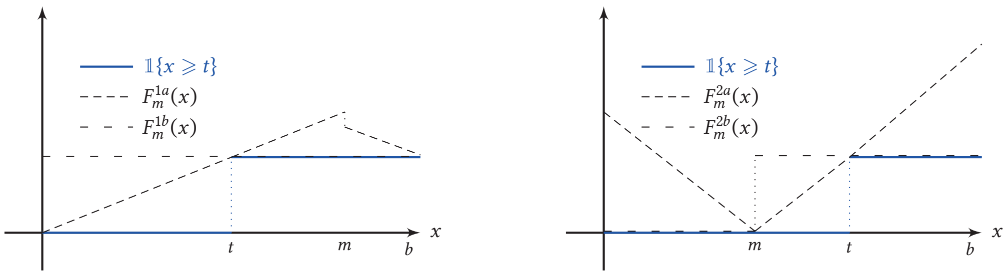


Figure 4.3: The function $F_m(x)$ under scenarios 1a, 1b, 2a and 2b

In robust statistics the median is widely considered as a more suitable location parameter than the mean when the distribution is estimated from historical data and contaminated with outliers through another (possibly fat-tailed) distribution [39]. As the former puts less importance on the tail of the distribution, it is barely affected by those outliers. The median and MAD around the median are the robust variants of, respectively, the mean and standard deviation. Next to describing the underlying distribution more accurately, the median might also provide a better measure of central tendency for the quantity that we are estimating from historical data. For example, the median wealth of a population is a better measure of "typical" wealth than the expected value since the distribution of wealth is often skewed. To illustrate the median-MAD tail bound (4.14), we plot a small example in which the ground truth follows a Pareto distribution, see Figure 4.4. Here the "true" distribution resides in the ambiguity set $\mathscr{P}_{(m,b,d_m)}$ with $m = \sqrt{2}$, $d_m = 2\sqrt{2} - 2$ and unbounded support. When the shape parameter of the Pareto distribution equals 2, the variance is infinite, rendering Chebyshev's inequality useless, but the median-MAD bound can still be computed. In the remainder of this chapter, we will focus on the mean-MAD ambiguity set now that we demonstrated our approach is also applicable to other types of information.

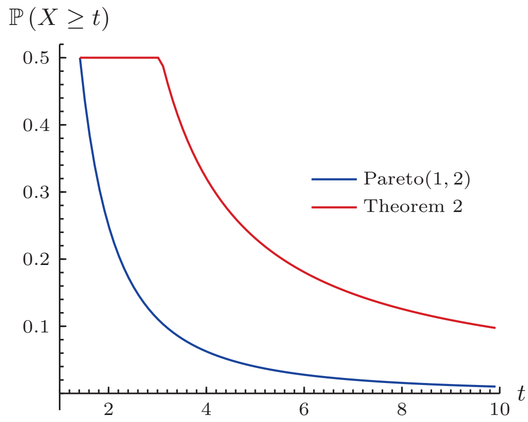


Figure 4.4: The median-MAD bounds for the tail probability

We next consider the tail bounds when also a specific measure of skewness is known: $\beta = \mathbb{P}(X \geqslant \mu)$. Since this statistic contains information about the distribution of X relative to its mean μ, it is often combined with mean-MAD information [19, 179]. Define the restricted ambiguity set

$$\mathscr{P}_{(\mu,b,d,\beta)} = \{\mathbb{P} : \mathbb{P} \in \mathscr{P}_{(\mu,b,d)}, \ \mathbb{P}(X \geqslant \mu) = \beta\}. \tag{4.15}$$

Using this ambiguity set results in new tight bounds. These results are stated in the following two results for which the primal-dual proofs are also provided in Appendix B.2.

Theorem 4.5. *For a random variable X with a distribution $\mathbb{P} \in \mathscr{P}_{(\mu,b,d,\beta)}$,*

$$\sup_{\mathbb{P} \in \mathscr{P}_{(\mu,b,d,\beta)}} \mathbb{P}(X \geqslant t) = \begin{cases} 1, & t \in [0, \tau_1], \\ \frac{(1-\beta)\mu + \beta t}{t} - \frac{d}{2t}, & t \in [\tau_1, \mu), \\ \beta, & t \in [\mu, \tau_2], \\ \frac{d}{2(t-\mu)}, & t \in [\tau_2, b], \end{cases} \tag{4.16}$$

with

$$\tau_1 = \mu - \frac{d}{2(1-\beta)}, \quad \tau_2 = \mu + \frac{d}{2\beta}.$$

Theorem 4.6. *For a random variable X with a distribution $\mathbb{P} \in \mathscr{P}_{(\mu,b,d,\beta)}$,*

$$\inf_{\mathbb{P} \in \mathscr{P}_{(\mu,b,d,\beta)}} \mathbb{P}(X > t) = \begin{cases} 1 - \frac{d}{2(\mu-t)}, & t \in [0, \tau_1], \\ \beta, & t \in [\tau_1, \mu), \\ \frac{\beta(\mu-t)}{(b-t)} + \frac{d}{2(b-t)}, & t \in [\mu, \tau_2], \\ 0, & t \in [\tau_2, b] \end{cases} \tag{4.17}$$

with

$$\tau_1 = \mu - \frac{d}{2(1-\beta)}, \quad \tau_2 = \mu + \frac{d}{2\beta}.$$

Note that for these bounds equality between $\mathbb{P}(X \geqslant t)$ and $\mathbb{P}(X > t)$ does not hold. The reason for this is demonstrated in the proofs of both the upper and lower bound. Since the function $F(x)$ in the dual problem now has a jump discontinuity for $x = \mu$, the optimal solution depends on whether the indicator function is upper or lower semi-continuous.

In Figure 4.5, the upper and lower bounds are depicted for the ambiguity set that considers all distributions with $\mu = 0.5$, $d = 0.1875$, $\beta = 0.5$, $a = 0$ and $b = 1$. As a point of reference, the Beta(2, 2) tail distribution, which is a member of the ambiguity set, is also plotted.

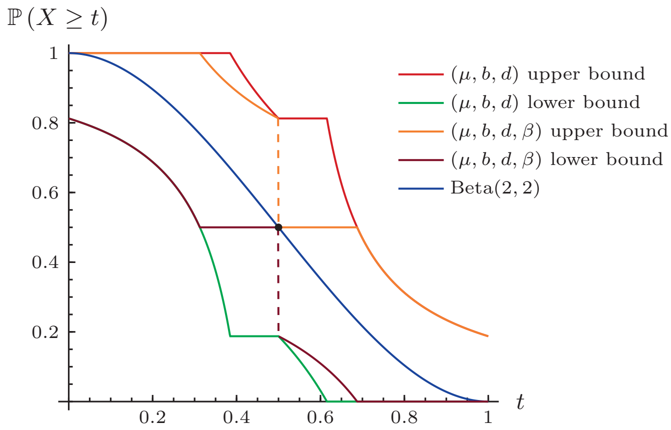


Figure 4.5: An illustration of the mean-MAD-β bounds for the tail probability

Next to summary statistics such as the mean, MAD and median, it is also possible to impose structural properties of the underlying distributions, like unimodality and symmetry, by altering the constraints of the dual problem. By cleverly exploiting conic duality, Popescu [177] shows how to adapt the dual problem to take into account these additional conditions. Moreover, we can apply the techniques discussed in this section to other objectives than the indicator function; see, for example, Section 4.3.2.

4.2.3. Comparison with other bounds

Closely related to our results is the discussion in section 4.1 of [87]. They consider, among others, the ambiguity set

$$\tilde{\mathscr{P}}_{(\mu,b,d)} = \{\mathbb{P} : \mathscr{B} \to [0,1] \, : \, \mathbb{P}\left[X \in [0,b]\right] = 1, \;\; \mathbb{E}_{\mathbb{P}}\left[X\right] = \mu, \;\; \mathbb{E}_{\mathbb{P}}\left[|X-\mu|\right] \leqslant d\}.$$

The only difference with the ambiguity set we use is the inclusion of all distributions with a lower mean absolute deviation. This has major implications for the maximum and minimum probability that X exceeds t, however. First of all, it should be noted that the distribution with all its probability mass on μ is an element of $\tilde{\mathscr{P}}_{(\mu,b,d)}$ for any value of d. This means that for any $t \leqslant \mu$ it holds that

$$\sup_{\mathbb{P}\in\tilde{\mathscr{P}}_{(\mu,b,d)}} \mathbb{P}\left(X \geqslant t\right) = 1.$$

Moreover, for any $t > \mu$ and $d > 2\mu\left(t-\mu\right)/t$, the maximum probability of X exceeding t is attained by a distribution with a mean absolute deviation equal to $2\mu\left(t-\mu\right)/t$, which is explained by the observation that the bound we obtain is decreasing in d for $d > 2\mu\left(t-\mu\right)/t$.

Clearly, because of the above observations, the theoretical maximum of $\mathbb{P}\left(X > t\right)$ has a much simpler closed-form solution than (4.5) for the ambiguity set $\tilde{\mathscr{P}}_{(\mu,b,d)}$. A big downside is that many of the distributions contained in $\tilde{\mathscr{P}}_{(\mu,b,d)}$ but not in $\mathscr{P}_{(\mu,b,d)}$ are unrealistic. Especially when the mean absolute deviation is known or can be accurately estimated, there is little reason to consider distributions with a different (in this case lower) mean absolute deviation. For large values of d relative to t in particular, using $\tilde{\mathscr{P}}_{(\mu,b,d)}$ can lead to an overestimation of the maximum value of $\mathbb{P}\left(X > t\right)$. The observation that the maximum value of $\mathbb{P}\left(X > t\right)$ is decreasing in d for large values of d also means that considering distributions with a lower mean absolute deviation can lead to a higher bound on $\mathbb{P}\left(X > t\right)$.

Comparing the result of Theorem 4.1 to Cantelli's inequality (4.1) is harder, since we assume the mean absolute deviation to be known, but not the variance. Hence, some relation between these two quantities is needed to be able to make a comparison. In particular, we will use that

$$d^2 \leqslant \sigma^2 \leqslant \frac{d(b-a)}{2}. \tag{4.18}$$

Throughout the comparison below we assume that d is given and compare the bound obtained in Theorem 4.1 with Cantelli's bound for different values of σ satisfying (4.18). Figure 4.6 illustrates this comparison for a simple numerical example with the following parameters: $a = -1$, $\mu = 0$, $b = 1$, $d = \frac{1}{4}$. We consider three values for σ: $\sigma = d = \frac{1}{4}$, $\sigma = \frac{1}{3}$ and $\sigma = \sqrt{d(b-a)/2} = \frac{1}{2}$.

Figure 4.6 gives rise to a number of interesting observations. First of all, we note that since Cantelli's bound is 1 for any $t \leqslant \mu$, the bound from Theorem 4.1 is at most Cantelli's bound as it includes an interval for which it is not 1. Furthermore, the flat area in the blue line corresponds to the values of t such that

$$\min\left\{\frac{d}{2\left(t-\mu\right)}, 1-\frac{d}{2(\mu-a)}\right\} = 1-\frac{d}{2(\mu-a)},$$

which corresponds to all $\mu \leqslant t \leqslant \tau_2 := \mu + \frac{d(\mu-a)}{2(\mu-a)-d}$. Moreover, we note that for $\sigma = d$, Cantelli's bound is lower than (4.5) for all $\tau_2 \leqslant t \leqslant b$. This is true for all parameters as:

$$\begin{aligned}\frac{d^2}{d^2+\left(t-\mu\right)^2} &= \frac{d^2}{\left(d-(t-\mu)\right)^2+2d(t-\mu)}\\ &\leqslant \frac{d^2}{2d(t-\mu)} = \frac{d}{2\left(t-\mu\right)}.\end{aligned}$$

In particular, for $\sigma = d$ Cantelli's bound and (4.5) always coincide at $t = \mu + d$, since

$$\frac{d^2}{d^2+d^2} = \frac{1}{2} = \frac{d}{2d}.$$

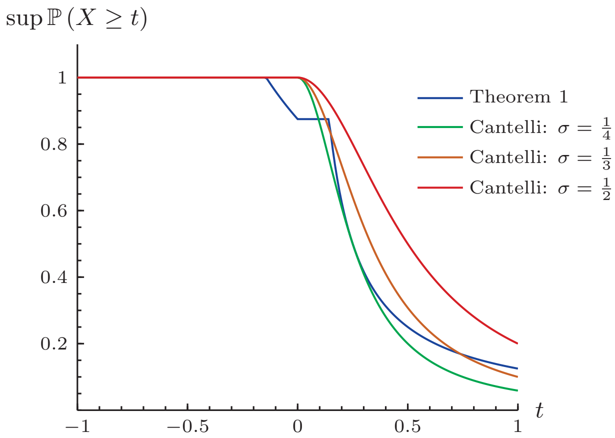


Figure 4.6: A comparison of the bound in Theorem 4.1 with Cantelli's bound for three different values of σ with the parameter values $a = -1$, $\mu = 0$, $b = 1$ and $d = \frac{1}{4}$

If, on the other hand, we choose $\sigma = \sqrt{d(b-a)/2}$, its highest possible value, Cantelli's bound is higher than (4.5). This is true for all parameter values as well, as Cantelli's bound is increasing in σ and must thus be at least (4.5) for its highest possible value.

For intermediate values of σ, we observe behavior similar to the line corresponding to $\sigma = \frac{1}{3}$ in Figure 4.6. More specifically, we find that (4.5) is lower than Cantelli's bound for all t in the two intervals $[0, \hat{\tau}]$ and$[\underline{\tau}, \overline{\tau}]$, with the three boundaries given by

$$\hat{\tau} = \mu + \sqrt{\frac{d\sigma^2}{2(\mu - a) - d}},$$

$$\underline{\tau} = \frac{\sigma^2}{d} - \sigma\sqrt{\frac{\sigma^2}{d^2} - 1},$$

$$\overline{\tau} = \min\left\{b, \frac{\sigma^2}{d} + \sigma\sqrt{\frac{\sigma^2}{d^2} - 1}\right\}.$$

Note that for some σ, such as $\sigma = \sqrt{d(b-a)/2}$ in Figure 4.6, it holds that $\hat{\tau} \geqslant \underline{\tau}$, that is, (4.5) is lower than Cantelli's bound for all $t \in [\mu, \overline{\tau}]$. To visually clarify all boundaries discussed above, Figure 4.7 only shows Cantelli's bound for $\sigma = 0.27$ and marks τ_2, $\hat{\tau}$, $\underline{\tau}$ and $\overline{\tau}$. It should be noted that this value of σ is very close to the minimum of 0.25, and hence Cantelli's bound compares more favorably than is generally to be expected. In the appendix we provide a similar comparison with the variance-based bound of [64], and also perform numerical experiments with the other bounds derived in the previous subsection.

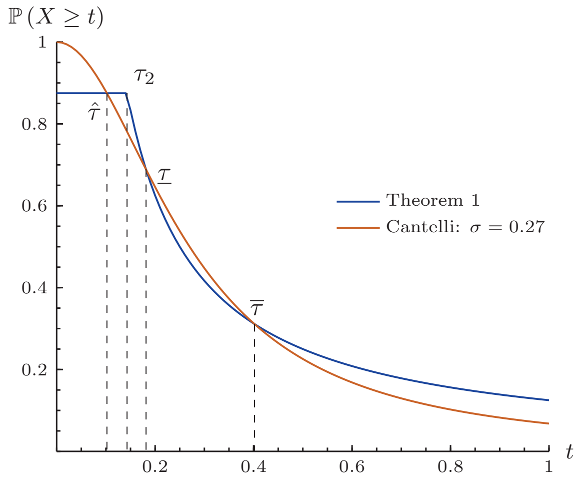


Figure 4.7: An illustration of $\hat{\tau}$, τ_2, $\underline{\tau}$ and $\overline{\tau}$ for the parameter values $a = -1$, $\mu = 0$, $b = 1$ and $d = 0.25$

4.3. Distribution-free analysis of OR models

We now turn to three OR applications: the newsvendor problem, stop-loss reinsurance and radiotherapy optimization. These three models can be subjected to distribution-free analyses that make use of the novel tight bounds. The common theme is that with ambiguity described in terms of mean, MAD and restricted support, distribution-free analysis often leads to valuable structural insights.

4.3.1. Newsvendor problem

The newsvendor problem serves to find the order quantity that maximizes the expected profit for a single period given a stochastic demand. Denote by q the order quantity (number of units) and by D the stochastic demand during a single selling period. Per unit, p denotes the selling price and c the purchase cost. Let $p > c$, and assume without loss of generality that unsold units have zero salvage value. The expected profit is $\mathbb{E}_{\mathbb{P}}[\pi(q, D)]$ with $D \sim \mathbb{P}$ and

$$\pi(q, D) = p \min\{q, D\} - cq.$$

The decision maker then chooses the optimal order quantity q^* that solves $\max_q \mathbb{E}_{\mathbb{P}}[\pi(q, D)]$. This solution is known to be the $\eta := (p - c)/p$ quantile (critical quantile) of the distribution of D, that is,

$$q^* = \min\{q \ : \ \mathbb{P}(D \leqslant q) \geqslant \eta\}. \tag{4.19}$$

In practice, however, the decision maker might only know partial information on the demand distribution. [191] pioneered distribution-free analysis of the newsvendor problem, when only the mean and the variance of the demand are known. Scarf obtained the optimal order quantity for the worst-case demand, turning the newsvendor into a maxmin decision maker that solves

$$\max_q \inf_{\mathbb{P}\in\mathscr{P}_{(\mu,\sigma)}} \mathbb{E}_{\mathbb{P}}\left[\pi(q,D)\right] \tag{4.20}$$

with $\mathscr{P}_{(\mu,\sigma)}$ the ambiguity set that contains all distributions with a given mean μ and variance σ^2, and solution

$$q^S = \begin{cases} 0, & \text{if } \eta < \frac{\sigma^2}{\mu^2+\sigma^2}, \\ \mu + \frac{\sigma}{2}\frac{2\eta-1}{\sqrt{\eta(1-\eta)}}, & \text{if } \eta \geqslant \frac{\sigma^2}{\mu^2+\sigma^2}. \end{cases} \tag{4.21}$$

We shall instead consider all demand distributions with given mean μ, MAD d and support $[0,b]$, and consider

$$\max_q \inf_{\mathbb{P}\in\mathscr{P}_{(\mu,d,b)}} \mathbb{E}_{\mathbb{P}}\left[\pi(q,D)\right]. \tag{4.22}$$

This is the counterpart of problem (4.20). [191] solved (4.20) directly, computing the lower bound $\inf_{\mathbb{P}\in\mathscr{P}_{(\mu,\sigma)}} \mathbb{E}_{\mathbb{P}}\left[\pi(q,D)\right]$ via a linear program. Instead, we do not solve (4.22) directly, but apply the tail probability bound from Theorem 4.1 to the first-order condition for q^* in (4.19). Clearly, the tight lower and upper bounds for q^* follow from substituting $\inf_{\mathbb{P}\in\mathscr{P}_{(\mu,b,d)}} \mathbb{P}(D > q)$ and $\sup_{\mathbb{P}\in\mathscr{P}_{(\mu,b,d)}} \mathbb{P}(D > q)$ into (4.19), respectively.

Proposition 4.7 **(Order quantity bounds under mean-MAD-range ambiguity).** *Suppose the newsvendor knows the mean μ, the mean absolute deviation d and the support's upper bound b of the demand distribution $\mathbb{P}(D \leqslant q)$. The optimal order quantity q^* that solves $\max_q \mathbb{E}_{\mathbb{P}}[\pi(q,D)]$ is then contained in the interval $[q^L, q^U]$ with*

$$[q^L, q^U] = \begin{cases} \left[0,\ \frac{2\mu(b-\mu)-bd}{2(b-\mu)(1-\eta)-d}\right], & \text{if } \eta < \frac{d}{2\mu}, \\ \left[\mu - \frac{d}{2\eta},\ \mu + \frac{d}{2(1-\eta)}\right], & \text{if } \frac{d}{2\mu} \leqslant \eta \leqslant 1 - \frac{d}{2(b-\mu)}, \\ \left[\frac{\mu-b(1-\eta)}{\eta-d/(2\mu)},\ b\right], & \text{if } \eta \geqslant 1 - \frac{d}{2(b-\mu)}, \end{cases} \tag{4.23}$$

where $\eta = (p-c)/p$ is the critical quantile of the distribution of D.

The proposition provides various handles for a robust policy that responds to the uncertainty captured in $\mathscr{P}_{(\mu,d,b)}$. The lower bound q^L follows from the worst-case demand distribution. Observe that q^L is larger than μ when the profit margin η exceeds $1 - d/2(b-\mu)$, and smaller than μ otherwise. This insight can be contrasted with q^S in (4.21) that also considers the worst-case scenario, but then in view of $\mathscr{P}_{(\mu,\sigma)}$ ambiguity. Scarf's q^S is larger than μ if $\eta > 1/2$ and smaller than μ otherwise. Hence, q^L quantifies the dependency on b, where q^S does not. In particular, when the profit margin η is fixed, the pessimistic newsvendor that uses q^L will only order above the mean when b does not exceed $\mu + d/2(1-\eta)$.

Table 4.1 shows that the support $[0,b]$ also influences the intervals $[q^L, q^U]$, in particular for low and high profit margins. We also recognize the three different regimes in Proposition 4.7 that correspond to low margins, average margins and high margins.

We mention two further works related to Proposition 4.7. [20] use general techniques for stochastic programs with limited information such as (4.22). For such stochastic programs the

Table 4.1: The intervals $[q^L, q^U]$ for mean-MAD ambiguity with $\mu = 5$, $d = 1.5$ and various profit margins η

η	$b = 10$	$b = 15$	$b = 20$	$b = \infty$
0.01	[0.00, 4.17]	[0.00, 4.23]	[0.00, 4.26]	[0.00, 4.29]
0.1	[0.00, 4.67]	[0.00, 4.70]	[0.00, 4.71]	[0.00, 4.72]
0.2	[1.25, 5.94]	[1.25, 5.94]	[1.25, 5.94]	[1.25, 5.94]
0.4	[3.13, 6.25]	[3.13, 6.25]	[3.13, 6.25]	[3.13, 6.25]
0.5	[3.50, 6.50]	[3.50, 6.50]	[3.50, 6.50]	[3.50, 6.50]
0.7	[3.93, 7.50]	[3.93, 7.50]	[3.93, 7.50]	[3.93, 7.50]
0.9	[5.33, 10.00]	[4.17, 12.50]	[4.17, 12.50]	[4.17, 12.50]
0.95	[5.63, 10.00]	[5.31, 15.00]	[5.00, 20.00]	[4.21, 20.00]
0.99	[5.83, 10.00]	[5.77, 15.00]	[5.71, 20.00]	[4.24, 80.00]

available information is often not sufficient to find the optimal solution. [20] develop a method to construct the minimal set that should contain the optimum. They also demonstrate this technique for the newsvendor model with given mean and MAD, but unbounded support, and obtain intervals that indeed arise from Proposition 4.7 for the limit $b \to \infty$:

$$[q^L, q^U] = \begin{cases} \left[0, \frac{\mu - d/2}{1-\eta}\right], & \text{if } \eta < \frac{d}{2\mu}, \\ \left[\mu - \frac{d}{2\eta}, \mu + \frac{d}{2(1-\eta)}\right], & \text{if } \eta \geqslant \frac{d}{2\mu}. \end{cases} \tag{4.24}$$

Natarajan et al. [163] introduce semi-variance as an extra piece of information about the skewness of the distribution. Together with the mean and variance, this results in a more restrictive ambiguity set (compared to Scarf), and therefore a less conservative (or sharper) estimation of q^*. In our case, we restrict the ambiguity set further with $\mathbb{P}(X \geqslant \mu) = \beta$ information that, like semi-variance, measures skewness. The following problem is the mean-MAD counterpart of the mean-variance-semivariance model discussed in the work of [163]:

$$\max_q \inf_{\mathbb{P} \in \mathscr{P}_{(\mu,b,d,\beta)}} \mathbb{E}_{\mathbb{P}}[\pi(q, D)], \tag{4.25}$$

where β adopts the role of the semivariance as a measure of skewness. We use Theorems 4.5 and 4.6 to bound the tail distribution of the demand D, and obtain sharper bounds for q^*.

Proposition 4.8 **(Order quantity bounds under mean-MAD-β ambiguity).** *Suppose the newsvendor knows that $\mathbb{P} \in \mathscr{P}_{(\mu,b,d)}$ and $\mathbb{P}(D \geqslant \mu) = \beta$. The optimal order quantity q^* that solves*

$\max_q \mathbb{E}_{\mathbb{P}}[\pi(q, D)]$ *is then contained in the interval* $[q^L, q^U]$ *with*

$$[q^L, q^U] = \begin{cases} \left[0, \frac{(1-\beta)\mu - d/2}{(1-\eta-\beta)}\right], & \text{if } \eta < \frac{d}{2\mu}, \\ \left[\mu - \frac{d}{2\eta}, \mu\right], & \text{if } \frac{d}{2\mu} \leqslant \eta < 1-\beta, \\ \mu, & \text{if } \eta = 1-\beta, \\ \left[\mu, \mu + \frac{d}{2(1-\eta)}\right], & \text{if } (1-\beta) < \eta < 1 - \frac{d}{2(b-\mu)}, \\ \left[\frac{b(1-\eta)-\beta\mu-d/2}{(1-\eta-\beta)}, b\right], & \text{if } \eta \geqslant 1 - \frac{d}{2(b-\mu)}, \end{cases} \tag{4.26}$$

where $\eta = (p-c)/p$ *is the critical quantile of the distribution of* D.

This result provides a robust policy that protects against uncertainty contained in $\mathscr{P}_{(\mu,b,d,\beta)}$. Obviously, ordering the mean is optimal if $\eta = 1-\beta$. The lower bound q^L relates to the worst-case demand distribution. Similar to the case with mean-MAD-range information, q^L is larger than μ when the profit margin η exceeds $1 - d/2(b-\mu)$. Hence, skewness information does not determine whether the pessimistic newsvendor orders more than the mean, since the upper bound b again plays a decisive role. This can be contrasted with the results of [163], who show that, for $\mathscr{P}_{(\mu,\sigma,s)}$ ambiguity, the order quantity is greater than μ if $\eta > \frac{1}{2}(1+s)$, where s is the normalized semivariance.

Table 4.2 shows that the bounded support $[0, b]$ again influences the intervals for low and high profit margins. The new intervals are sharper than the ones found in Table 4.1. This is, of course, an obvious result of incorporating more distributional information.

Table 4.2: The intervals $[q^L, q^U]$ for mean-MAD-β ambiguity with $\mu = 5$, $d = 1.5$, $\beta = 0.5$ and various profit margins η

η	$b = 10$	$b = 15$	$b = 20$	$b = \infty$
0.01	[0.00, 3.57]	[0.00, 3.57]	[0.00, 3.57]	[0, 3.57]
0.1	[0.00, 4.38]	[0.00, 4.38]	[0.00, 4.38]	[0.00, 4.38]
0.2	[1.25, 5.00]	[1.25, 5.00]	[1.25, 5.00]	[1.25, 5.00]
0.4	[3.13, 5.00]	[3.13, 5.00]	[3.13, 5.00]	[3.13, 5.00]
0.5	5.00	5.00	5.00	5.00
0.7	[5.00, 7.50]	[5.00, 7.50]	[5.00, 7.50]	[5.00, 7.50]
0.9	[5.63, 10.00]	[5.00, 12.50]	[5.00, 12.50]	[5.00, 12.50]
0.95	[6.11, 10.00]	[5.56, 15.00]	[5.00, 20.00]	[5.00, 20.00]
0.99	[6.43, 10.00]	[6.33, 15.00]	[6.22, 20.00]	[5.00, 80.00]

Apart from modifying or narrowing the ambiguity set, conservatism can be alleviated by choosing alternate objective functions, for instance by replacing the profit function by a regret function (opportunity cost of not making the optimal decision) [174, 229], or by extending the profit function with a utility function $u(\cdot)$ for max-min analysis of $\mathbb{E}[u(\pi(q, D))]$ as in [97]. See [163] for an extensive review of many other studies on distribution-free newsvendor models.

The tight bounds developed in this chapter can be used for distribution-free analysis of more advanced models, including those modeling regret and utility mentioned above, the risk-averse newsvendor with stochastic price-dependent demand [52] and multi-product settings [55].

4.3.2. Stop-loss reinsurance

Reinsurance is a classical topic in the actuarial sciences and insurance mathematics and implies that an insurance company transfers part of its risk to a reinsurance company; see e.g., [8], [127]. Say an insurance company faces a total claim S that is the sum of n individual claims X_i, $i = 1, \dots, n$. The insurance company pays the claim up to a level z, and the reinsurance company covers the remainder. This gives rise to the so-called retention function $\psi(z, S) = \min\{S, z\}$ that represents the payment of the insurer. We provide an upper bound for the standard stop-loss retention function in Proposition 4.9.

Proposition 4.9. *The worst-case expected claim payment of the direct insurer as a function of the retention limit z is given by*

$$\sup_{\mathbb{P}\in\mathcal{P}_{(\mu,b,d)}} \mathbb{E}_{\mathbb{P}}[\psi(z,S)] = \begin{cases} z, & \text{if } z \in [0,\tau_1], \\ \mu - \frac{d(b-z)}{2(b-\mu)}, & \text{if } z \in [\tau_1,\mu], \\ z(1-\frac{d}{2\mu}), & \text{if } z \in [\mu,\tau_2], \\ \mu, & \text{if } z \in [\tau_2, b], \end{cases} \tag{4.27}$$

where

$$\tau_1 = \mu - \frac{d(b-\mu)}{2(b-\mu)-d}, \quad \tau_2 = \mu + \frac{d\mu}{2\mu - d}.$$

Proof. First, note that

$$\psi(z,S) = \min\{S,z\} = S - \max\{S-z,0\},$$

and hence

$$\sup_{\mathbb{P}\in\mathcal{P}_{(\mu,b,d)}} \mathbb{E}_{\mathbb{P}}[\psi(z,S)] = \mu - \inf_{\mathbb{P}\in\mathcal{P}_{(\mu,b,d)}} \mathbb{E}_{\mathbb{P}}[\max\{S-z,0\}]. \tag{4.28}$$

The second term is convex in S, and thus equivalent to [179]

$$\min_{\frac{d}{2(b-\mu)}\leqslant\theta\leqslant 1-\frac{d}{2\mu}} \left\{\theta\max\{\mu + \frac{d}{2\theta} - z, 0\} + (1-\theta)\max\{\mu - \frac{d}{2(1-\theta)} - z, 0\}\right\}, \tag{4.29}$$

a convex optimization problem with a piecewise linear objective function. The optimal value depends on the retention limit z. Solving problem (4.29) for $z \in [0, b]$ and subtracting the optimal value from μ results in the four cases mentioned in (4.27). □

The payment function of the reinsurance company puts forward a more challenging problem when the insurance coverage is limited. In this case, a relevant performance characteristic is to what extent the insurance company benefits from the reinsurance contract. This benefit is measured with the function

$$\phi(z,m,S) = \begin{cases} m, & \text{if } S \geqslant z+m, \\ S-z, & \text{if } z \leqslant S \leqslant z+m, \\ 0, & \text{if } S \leqslant z. \end{cases} \tag{4.30}$$

When the total claim S stays below the retention limit z, the insurance company covers the entire claim, but when S exceeds z the reinsurer pays the excess claim up to a maximum m. Thus, the reinsurance company does not compensate large claims that exceed the exit point $m+z$. Above this level the risk is retained by the insurance company. We obtain a novel bound by primal-dual arguments.

Proposition 4.10. *The expected insurer's benefit is bounded by* $\sup_{\mathbb{P}\in\mathscr{P}_{(\mu,b,d)}} \mathbb{E}_{\mathbb{P}}[\phi(z,m,S)] =$

$$\begin{cases} \min\{m, \frac{m}{m+z}(\mu - \frac{d(b-(m+z))}{2(b-\mu)})\}, & \text{if } z \leqslant z+m \leqslant \mu, \\ \min\{m(1-\frac{d}{2\mu}), z(\frac{d}{2\mu}-1)+\mu\}, & \text{if } z \leqslant \mu \leqslant z+m \leqslant b, \\ \min\{m(1-\frac{d}{2\mu}), \frac{dm}{2(m+z-\mu)}\}, & \text{if } \mu \leqslant z \leqslant z+m \leqslant b, \end{cases} \tag{4.31}$$

where the function $\phi(z,m,S)$ *degenerates to* $\max\{S-z,0\}$ *if* $z+m>b$. *In this case,*

$$\sup_{\mathbb{P}\in\mathscr{P}_{(\mu,b,d)}} \mathbb{E}_{\mathbb{P}}[\phi(z,m,S)] = \begin{cases} z(\frac{d}{2\mu}-1)+\mu, & \text{if } z \leqslant \mu, \\ \frac{d(b-z)}{2(b-\mu)}, & \text{if } z \geqslant \mu. \end{cases} \tag{4.32}$$

For the sake of conciseness, we only sketch the proof. The full details are highly similar to the derivations used in the proof of Theorem 4.1.

We will show via primal-dual reasoning that the stated stop-loss formulas are tight upper bounds. We consider the measurable function $\phi(z,m,s)$. For a random variable S with distribution $\mathbb{P} \in \mathscr{P}_{(\mu,b,d)}$, we solve

$$\begin{aligned} \sup_{\mathbb{P}\in\mathscr{M}^+} \quad & \int_s \phi(z,m,s)\mathrm{d}\mathbb{P}(s) \\ \text{s.t.} \quad & \int_s \mathrm{d}\mathbb{P}(s) = 1,\ \int_s s\,\mathrm{d}\mathbb{P}(s) = \mu,\ \int_s |s-\mu|\,\mathrm{d}\mathbb{P}(s) = d. \end{aligned} \tag{4.33}$$

Consider the dual of (4.33),

$$\begin{aligned} \inf_{\lambda_0,\lambda_1,\lambda_2} \quad & \lambda_0 + \lambda_1\mu + \lambda_2 d \\ \text{s.t.} \quad & \phi(z,m,s) \leqslant \lambda_0 + \lambda_1 s + \lambda_2|s-\mu|,\ \forall s \in [0,b]. \end{aligned} \tag{4.34}$$

Define $F(s) := \lambda_0+\lambda_1 s+\lambda_2|s-\mu|$. Then the inequality in (4.34) can be written as $\phi(z,m,s) \leqslant F(s)$, $\forall s$, i.e. $F(s)$ majorizes the "staircase" function $\phi(z,m,s)$. Note that $F(s)$ has a 'kink' at $s=\mu$. There are six candidate scenarios, which are displayed in Figure 4.8. When $m+z \leqslant \mu$, $F(s)=1$ and touches $\phi(z,m,s)$ in $\{0, m+z, b\}$ (scenario 1a), or $F(s)$ touches $\phi(z,m,s)$ in $[m+z, b]$ (scenario 1b). When $z \leqslant \mu \leqslant m+z$, $F(s)$ touches in $\{0\} \cup [\mu, m+z]$ (scenario 2a) or in $\{0\} \cup [m+z, b]$ (scenario 2b). Finally, if $\mu \leqslant z \leqslant m+z$, $F(s)$ coincides with $\phi(z,m,s)$ in $[0,\mu] \cup \{m+z\}$ (scenario 3a) or in $\{0\} \cup [m+z, b]$ (scenario 3b).

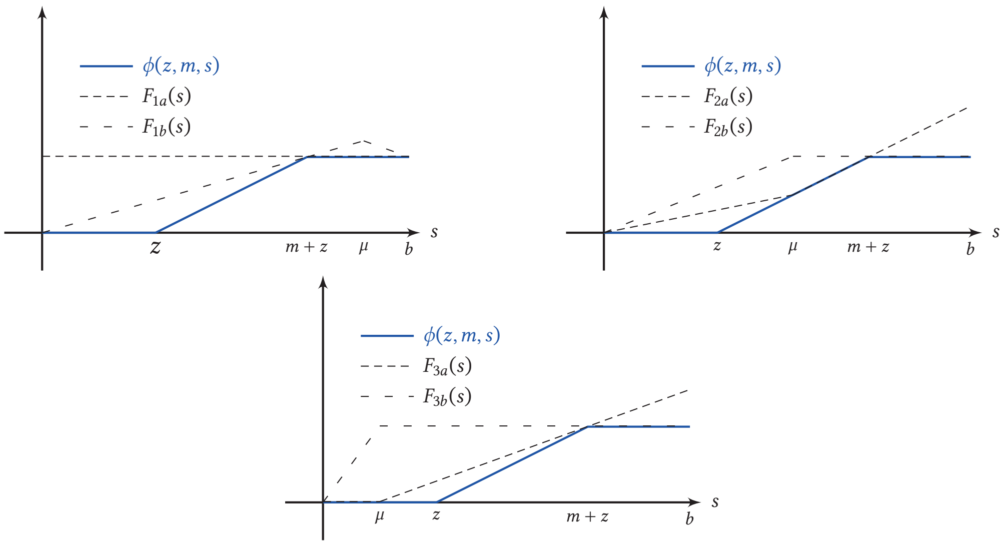


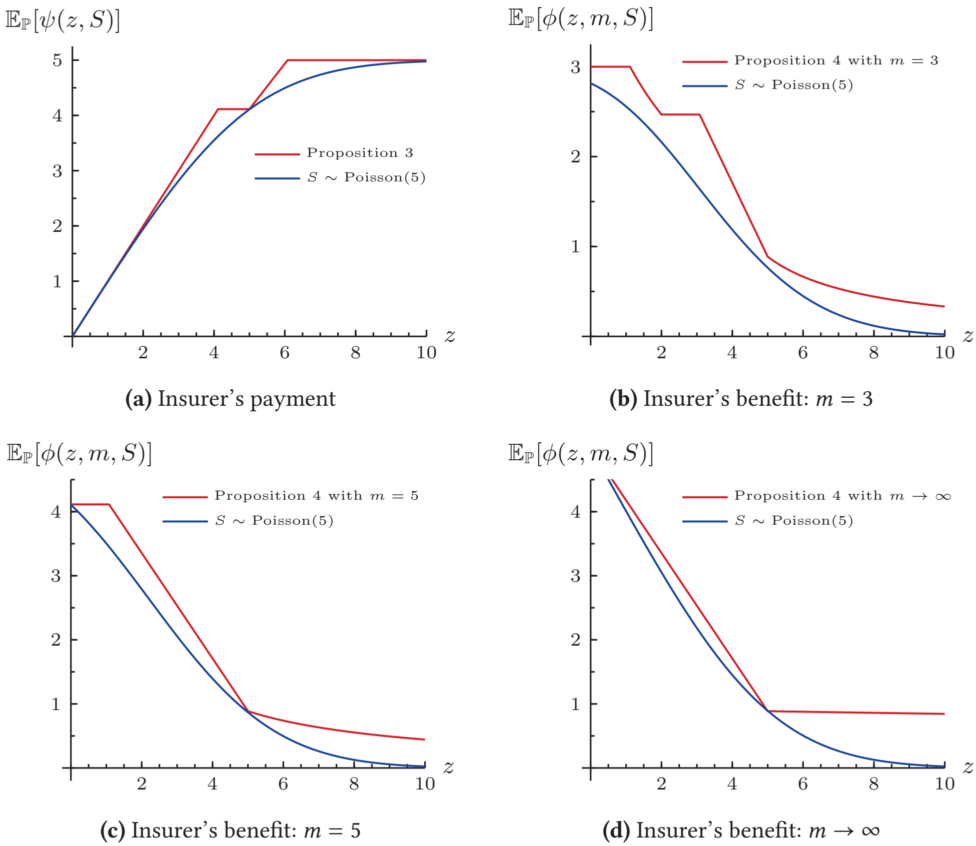
Figure 4.8: The different scenarios and majorizing functions

Scenario 1a implies that $F(0)=F(m+z)=F(\mu)=F(b)=m$, and hence $\lambda_0=m$, $\lambda_1=\lambda_2=0$ with objective value m. It is clear that the optimal primal objective value is also equal to m as the primal solution can only assign probability to values greater than or equal to $m+z$ (which is a consequence of complementary slackness).

Scenario 1b implies $F(0)=0$, $F(m+z)=F(b)=m$. It can be shown that allocating probability mass to the points $\{0, m+z, b\}$ in the primal problem (4.33) yields the same objective value, and hence the corresponding solutions are optimal.

Similarly, scenarios 2a, 2b, 3a and 3b imply values for at least three of $F(0)$, $F(\mu)$, $F(m+z)$ and $F(b)$, from which a dual and primal solution with equal objective value can be derived. The proof of the first part of the theorem is then completed by taking the minimum for each scenario. The second part is an immediate consequence of upper bound (8) in [179], which is a result that was already shown in [19].

An illustration of the bounds for the stop-loss payments is provided in Figure 4.9, where we display payments as functions of z with $\mu=5$, $d=1.77$, and $m=3$, $m=5$ and $m\to\infty$. We assume that the "true" total claim S follows a Poisson(5) distribution. Note the resemblance between the shape of the stop-loss bound in Figure 4.9b and of the mean-MAD tail probability bound in Theorem 4.1. The former bound, however, has an additional linear part with a negative



(a) Insurer's payment

(b) Insurer's benefit: $m = 3$

(c) Insurer's benefit: $m = 5$

(d) Insurer's benefit: $m \to \infty$

Figure 4.9: The bounds and "true" values of the expected claim payment of the insurance company $\mathbb{E}_{\mathbb{P}}[\psi(z,S)]$ and the insurer's benefit $\mathbb{E}_{\mathbb{P}}[\phi(z,m,S)]$ as functions of the retention limit z

slope for $\mu - m \leqslant z \leqslant \mu$. This linear segment is only present when m exceeds $d\mu/(2\mu - d)$; moreover, the bound approaches a linear function for $z \leqslant \mu$ when m is chosen sufficiently large. Additionally, letting $m \to \infty$, our example results in a bound equal to the constant $d/2$ if $z \geqslant \mu$, and thus the bound for the stop-loss payment of the reinsurer degenerates to a piecewise linear function consisting of two parts (a linear part with negative slope $d/2\mu - 1$ and a constant part equal to $d/2$).

These results complement the vast literature on tight bounds for expected claim payments. [58] considers bounded support and known first and second moment and obtains tight bounds using general results for moment problems. Other related works explore ways to sharpen the bounds using additional information. When modifying the ambiguity set by incorporating skewness information, imposing unimodality and symmetry conditions or using higher order moments, the gap between the upper and lower bounds narrows significantly; see [112], [65] and [120]. Note that mean-MAD information can easily be extended with additional parameters, such as the probability $\beta = \mathbb{P}(S \geqslant \mu)$ or the median.

4.3.3. Radiotherapy optimization

We consider a continuous optimization problem that arises in radiotherapy. Here, the biological effective radiation dose delivered to a tumor is to be maximized subject to a constraint on the biological effective dose delivered to the surrounding healthy tissue. Mathematically, the biological effective dose (BED) for a dose $\mathbf{x} \in \mathbb{R}^n$ delivered over n fractions is given by

$$B(\mathbf{x}) = \sum_{t=1}^{n} \left(x_t + \frac{1}{\rho} x_t^2 \right),$$

where ρ is the radiosensitivity parameter of the irradiated tissue. More specifically, it can be interpreted as the tissue's sensitivity to fractionation, where a low value indicates a high sensitivity to fractionation, i.e., the distribution of treatment over multiple fractions.

While there is an extensive body of research on the value of ρ for different tumor sites, it remains subject to significant uncertainty [124]. Moreover, since this value can differ from patient to patient, there is a very limited amount of data available and there is little evidence to suggest it follows some well known distribution. Throughout the rest of the example, we denote the sensitivity to fractionation by ρ_1 and ρ_2 for the tumor and the surrounding healthy tissue, respectively.

For illustrative purposes, we consider a setting in which it has been decided to deliver the treatment over two fractions, i.e., the optimization variables are limited to the dose in the first and second fraction. Moreover, we focus on the uncertainty of ρ_2, and thus model the restriction of sparing the healthy tissue through an ambiguous chance constraint. Mathematically, we wish to solve the following optimization problem [207]:

$$\max_{\mathbf{x} \in \mathbb{R}^2} \; x_1 + x_2 + \frac{1}{\rho_1} \left(x_1^2 + x_2^2 \right) \tag{4.35a}$$

$$\text{s.t. } \mathbb{P} \left(\sigma(x_1 + x_2) + \frac{1}{\rho_2} \sigma^2 \left(x_1^2 + x_2^2 \right) \leqslant t(\rho_2) \right) \geqslant 1 - \epsilon, \qquad \forall \mathbb{P} \in \mathscr{P}_{(\mu,b,d)}, \tag{4.35b}$$

$$x_1, x_2 \geqslant x_{min}, \tag{4.35c}$$

where σ is the generalized dose-sparing factor that denotes the fraction of the mean tumor dose that the healthy tissue receives on average, x_{min} is the minimum dose that must be delivered in each fraction, and $t(\rho_2)$ denotes the tolerance level of the healthy tissue and is given by

$$t(\rho_2) = \phi D \left(1 + \frac{\phi D}{T} \frac{1}{\rho_2} \right).$$

In other words, the healthy tissue is known to tolerate a total dose of D gray if it is delivered in T fractions under dose shape factor ϕ. This dose shape factor is a parameter that characterizes the spatial heterogeneity of a dose distribution [175].

The ambiguity of ρ is modeled through the mean-MAD ambiguity set, where the lower bound of the support is given by a instead of 0. In general, convex ambiguous chance constraints in which the uncertain parameter appears on the right-hand side can be reformulated as a tractable convex constraint.

Proposition 4.11. *Let $g : \mathbb{R}^n \mapsto \mathbb{R}$, $h \in \mathbb{R}$ and let Z be a random variable whose distribution lies in the ambiguity set*

$$\mathcal{P} = \{\mathbb{P} : \mathbb{P}[Z \in [-1, b]] = 1, \;\; \mathbb{E}[Z] = 0, \;\; \mathbb{E}[|Z|] = d\},$$

for some $d \in \left[0, \frac{2b}{1+b}\right]$. For any $\epsilon \in \left(0, \frac{1}{1+b}\right)$ and $\mathbf{x} \in \mathbb{R}^n$ it holds that

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{P}[g(\mathbf{x}) + Z \leqslant 0] \geqslant 1 - \epsilon, \tag{4.36}$$

if and only if

$$g(\mathbf{x}) + \min\left\{b, \frac{d}{2\epsilon}\right\} \leqslant 0. \tag{4.37}$$

Proof. We first rewrite (4.36) to

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}[Z > -g(\mathbf{x})] \leqslant \epsilon.$$

From Theorem 4.1 and the fact that $\epsilon < 1/(1 + b)$ we know that it must hold that $-g(\mathbf{x}) > \mathbb{E}[Z] = 0$. Given that requirement, we know by Theorem 4.1 that

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}[Z > -g(\mathbf{x})] = \begin{cases} \min\left\{\frac{d}{-2g(\mathbf{x})},\; 1 - \frac{d}{2}\right\} & \text{if } -g(\mathbf{x}) < 1 \\ 0 & \text{if } -g(\mathbf{x}) \geqslant 1. \end{cases}$$

From $d \in [0, \frac{2b}{1+b}]$ and $\epsilon \in (0, \frac{1}{1+b})$, it follows that $1 - d/2 > \epsilon$, and thus any feasible solution $\mathbf{x}$ must satisfy $-g(\mathbf{x}) \geqslant 1$ and/or $-d/2g(\mathbf{x}) \leqslant \epsilon$. The latter can be equivalently stated as

$$-g(\mathbf{x}) \geqslant \frac{d}{2\epsilon},$$

which can easily be combined with the former as

$$-g(\mathbf{x}) \geqslant \min\left\{1, \frac{d}{2\epsilon}\right\} \iff g(\mathbf{x}) + \min\left\{1, \frac{d}{2\epsilon}\right\} \leqslant 0.$$

Because $d/2\epsilon > 0$, we find that the requirement $-g(\mathbf{x}) > 0$ is redundant, and thus (4.36) is equivalent to (4.37). ☐

The ambiguous chance constraint (4.35b) is not naturally stated in the form (4.36). It can be rewritten, however, as

$$\mathbb{P}\left(\rho \cdot (\sigma(x_1 + x_2) - \phi D) > \frac{\phi^2 D^2}{T} - \sigma^2(x_1^2 + x_2^2)\right) \leqslant \epsilon, \qquad \forall \mathbb{P} \in \mathcal{P}_{(\mu,b,d)}, \tag{4.38}$$

where we note multiplication by ρ is allowed as its support is nonnegative. Leveraging the tail probability bound, we find for $\epsilon \in (0, \frac{\mu-a}{b-a})$ that (4.38) is equivalent to

$$\mu\sigma(x_1 + x_2) + \sigma^2(x_1^2 + x_2^2) + \frac{d}{2\epsilon}\left|\sigma(x_1 + x_2) - \phi D\right| \leqslant \mu\phi D + \frac{\phi^2 D^2}{T}. \tag{4.39}$$

The resulting optimization problem can be solved efficiently, as (4.39) can be equivalently stated as two conic quadratic inequalities through the introduction of an auxiliary variable.

We solve (4.35) for a specific, realistic set of parameters taken from [207], which are reported in Table 4.3. Figure 4.10 shows the feasible region and optimal solution of (4.35) for different

Table 4.3: Parameter values used for solving (4.35)

Parameter	Value
τ	4
σ	0.9
ϕ	2
D	27
T	5
x_{min}	1.5
a	3
b	6
μ	4
d	0.25

values of ϵ as well as the feasible region when we assume having the exact knowledge that $\rho = \mu$ and the feasible region when $\epsilon = 0.1$ and $\rho \sim \text{Beta}(6.6, 13.2)$, which is a member of the ambiguity set. Remarkable in this example is the similarity between the feasible region of the problem without any uncertainty, the specified Beta distribution and that of the ambiguous problem for $\epsilon = 0.1$ and $\epsilon = 0.05$. From the feasible region for $\epsilon = 0.01$, however, it is clear that requiring a low risk of violation results in a solution that is much worse in terms of tumor BED. Remarkable, also, is that the feasible region for the specified Beta distribution hardly changes with ϵ compared to the behavior under ambiguity. In fact, even for $\epsilon = 0.01$, the feasible region when we assume $\rho \sim \text{Beta}(6.6, 13.2)$ contains the ambiguous feasible region for $\epsilon = 0.01$. Note that the figure does illustrate how the shape of the feasible region changes with ϵ: the feasibility of unbalanced solutions, i.e., solutions that administer a different dose in the two fractions, is impacted much more severely than that of balanced solutions.

We mention two related works on ambiguous chance constraints. [101] present a tractable framework for joint ambiguous chance constraints under a few simplifying conditions. In particular, they assume a conic, hence unbounded, support, which is a key difference to our approach. Their approach is very powerful in settings for which an unbounded support makes sense, however, as they are able to elegantly deal with *joint* ambiguous chance constraints.

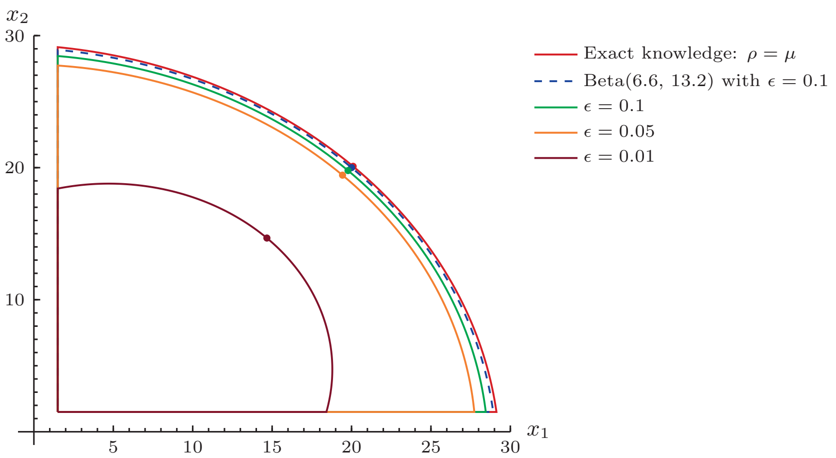


Figure 4.10: The feasible region and optimal solution of (4.35) for different values of ϵ as well as exact knowledge that $\rho = \mu$ (dots indicate the optimal solution)

[225], on the other hand, consider ambiguous chance constraints given a bounded support and moment information. Their assumptions on the ambiguity set do, however, exclude exact distributional information on nonlinear functions of the uncertain parameter, which we do assume in exact knowledge of the mean absolute deviation.

4.4. Conclusions and outlook

Tail probabilities are ubiquitous in probabilistic studies in many areas of science and application domains. Just as the original Chebyshev's inequality for mean-variance ambiguity, we expect our novel tail bounds for mean-MAD ambiguity to find many applications. In our search for tight bounds under limited information, we had to solve for the worst-case distribution and worst-case value of the expectation of the indicator function $\mathbb{1}\{X \geqslant t\}$. In this chapter the limited information was captured through ambiguity sets $\mathscr{P}_{(\mu,b,d)}$, $\mathscr{P}_{(\mu,b,d,\beta)}$ and $\mathscr{P}_{(m,b,d_m)}$, and it turned out that the combination of the non-convex indicator function with these ambiguity set gave rise to semi-infinite linear programs with easy, closed-form solutions.

In future work, we expect to find more such solvable classes, i.e. specific combinations of objective function (other than the indicator function) and ambiguity sets that together give rise to solvable linear programs and hence easy extremal distributions. In this way, one can try to sharpen the tail bounds by including more information (e.g. higher moments or percentiles), or to consider objective functions other than the tail probability. Our proof method based on solving the dual problem with piecewise-linear majorants is not tailor-made for the indicator function, and could potentially work for a much larger class of (measurable) objective functions, as shown in Section 4.3.2. Another direction we shall pursue is the application of the bounds to more complex, and possibly high-dimensional, robust optimization problems. To do so, we shall leverage the connection with the quickly evolving field of DRO, as illustrated by examples in

Section 4.3.3. Indeed, minimax and maximin decision problems arise naturally, and the bounds and proof techniques can help in advancing that field.

5
A generalized moment approach for conditional expectations

5.1. Introduction

Distribution-free performance analysis of stochastic models strives to obtain tight bounds for the expectation of an objective function of random variables, using only limited information about the underlying probability distributions. Traditionally, given the moment information about the random variables, these problems are modeled as generalized moment problems. The sharpest (i.e., "best possible") upper and/or lower bounds for these objective functions of random variables are found by solving semi-infinite optimization problems, where the optimization is taken over all admissible distributions of the random variables. In this chapter, we explore a new situation in which, besides the given moment sequence, we possess stronger information—that is, we have knowledge that a particular random event will occur, which pertains to the realizations of the random variables rather than their underlying probability distribution. We are interested in solving the moment problem with the knowledge of this random event, which may be based on assumptions, assertions, expert judgment or past observations. Alternatively, we may want to determine the worst-case behavior of a stochastic model if such a random event occurs. Conditional expectations offer a way to model this event information and provide us with the best estimate of the expected value of a function of random variables, knowing that

This chapter is based on the research paper [209].

the specified random event will occur. It is of interest to define a generalized moment framework for this new setting with conditional expectations, instead of standard expectations in which the random variables are not restricted to take on a subset of the values in the event set. However, random variables conditioned on a random event give rise to a different probabilistic concept, requiring a distinct type of analysis compared to the traditional theory on generalized moment problems.

Moment problems have been investigated by probability theorists since the end of the 19th century, see, e.g., [43, 153, 205]. Smith [201] revisited the generalized moment problem in his contemporary discussion, highlighting its various applications in decision theory. In his work, Smith briefly mentions the setting with prior information, but notes that the resulting expectation is no longer linear in the probability measure, thus presenting a more challenging problem that is not directly amenable to the techniques discussed in his work. As a solution, Smith [201] suggested a linearization technique, as discussed in [141] and [59], which converts the problem into a series of regular generalized moment problems. Along the lines of Shapiro [196], we define a novel variant of the generalized problem of moments and use semi-infinite programming theory to obtain tight bounds in a setting where the objective function is a fractional, rather than linear, function of the probability distribution. A comprehensive overview of general semi-infinite programming theory can be found in [113]. To the best of our knowledge, there are no general techniques available for obtaining tight bounds on conditional expectations. Although [151] provided some bounds for conditional expectations for traditional power moments, these are only for the expectation of a single random variable and are not necessarily tight. Moreover, [151] primarily used conditional information to sharpen Chebyshev-type tail probability bounds. In contrast, we provide a general framework for obtaining the best possible bounds for conditional expectations under limited information.

Extending the ideas described above to the DRO setting, we are interested in obtaining tight bounds for the expectation of some objective function of random variables, where the function also contains a decision vector, in order to protect ourselves against the worst nature has to offer. Conditional-moment information has been used in various works to describe ambiguity sets. In the minimax stochastic programming literature, Birge and Wets [34] incorporated such information into the constraints of generalized moment problems to bound the objective values of stochastic programming problems. de Klerk et al. [63] restricted the distributions in the ambiguity set to polynomial density functions. They demonstrated that these ambiguity sets are highly expressive because they can conveniently accommodate distributional information about conditional probabilities, conditional moments and marginal distributions. Chen et al. [53] introduced scenario-wise ambiguity sets that capture information with conditional expectation constraints based on generalized moments. Although these works use conditional moment information, their objectives aim to maximize the conventional expectation. Consequently, their approaches cannot directly handle the case in which the conditional expectation is the objective.

Here we should mention some work on distributionally robust optimization in which the objective function is fractional. When only support information is available, Gorissen [90] extended robust optimization formulations to fractional programming in which the objective function is a fraction of two functions of the uncertain parameters. Liu et al. [145] solved a DRO problem with moment constraints which consists of the maximization of ambiguous fractional functions representing reward-risk ratios. Ji and Lejeune [122] also investigated this class of fractional DRO problems using semi-infinite programming epigraphic formulations to solve the ambiguous reward-risk ratio problem and, additionally, design a data-driven formulation and solution framework using the Wasserstein ambiguity set. Using the conditional information, we can use the conditional-expectation bounds to enhance "worst-case" decision-making in a DRO framework. As for this setting, there does exist a separate thrust of research named contextual distributionally robust optimization, in which the prior knowledge on the realizations of the random variables is commonly referred to as side information. This area of research is a natural extension of the prescriptive stochastic programming paradigm. In this paradigm, the central object of interest is the joint distribution of the side information and the outcome random variables, which, if known, would result in more accurate estimations of the outcome variable when conditioning this distribution on the side information given. In practice, however, this joint distribution is usually not known precisely and is only estimated using a finite data sample. The ultimate goal of prescriptive stochastic programming is to develop an optimization methodology that uses the available side information to improve decision-making given only limited insights into the predictive power of the side information on the uncertain outcome parameters (see, e.g., [12, 26, 203]). The contextual DRO modeling paradigm assumes that, next to this side information, the joint distribution is contained in an ambiguity set that is defined using the limited information available. Research on this paradigm is still relatively scarce. We highlight a number of works. Esteban-Pérez and Morales [78] described ambiguity through a partial mass transportation problem, and exploited probability-trimming methods to solve the contextual DRO problem. In contrast, Nguyen et al. [169] worked directly with the optimal transport ambiguity set, and the authors succeeded in finding tractable conic reformulations for DRO problems with side information. Nguyen et al. [168] considered a Wasserstein-ball ambiguity set [157], which is centered on the empirical distribution that follows from the available data sample of the side information and the outcome parameters. Even though the Wasserstein-ball ambiguity set is a class of distributional ambiguity sets obtained through the theory of optimal transport, the models and results obtained in [168, 169] behave qualitatively differently due to special properties of the type-∞ Wasserstein distance, which is used to construct the ambiguity set. The literature described above mostly considers ambiguity sets that are defined through measures that define distances between probability distributions (such as the Wasserstein metric), whereas in this chapter, we will focus on ambiguity sets described by generalized moment information.

5.1.1. Contributions and outline

The main contributions of this chapter may be summarized as follows.

1. We expand the theory on semi-infinite programming and generalized moment problems, by deriving duality results for linear-fractional programming in the semi-infinite setting.

We extend several well-known results from the theory on generalized moment problems in order to bound conditional expectations, rather than standard expectations of functions of random variables. The fact that most results carry over to this setting with conditional expectations is particularly intriguing because the conditional expectation is a nonlinear function of the probability measure (thus not amenable to standard techniques based on semi-infinite linear programming).

2. We apply these novel results for the generalized conditional-bound problem to univariate functions of a simple random variable. Using primal-dual arguments, we obtain several closed-form bounds for different types of dispersion information. In addition to generalized moment information, we show that structural properties of the distribution can also be included. We further demonstrate our approach by resolving a minimax optimization problem, taken from the robust monopoly pricing literature. It is further asserted that most computations, in the univariate setting, are as tractable as with the linear expectations operator.

3. We use findings from robust uncertainty quantification and distributionally robust convex optimization to develop conic reformulations for the multivariate problem. We then apply these reformulations to contextual DRO, presenting a generalized moment framework for distributionally robust optimization with side information. The resulting framework is designed for conditional decision-making, incorporating both the side information and the distributional information contained within the ambiguity set. The computational tractability of the reformulations turns out to be closely related to that of distributionally robust convex optimization problems with support restrictions on the random variables.

The remainder of the chapter is organized as follows. In Section 5.2, we introduce the generalized moment bound problem for conditional expectations and elaborate on the duality approach. Section 5.3 discusses several examples of tight bounds for the conditional expected value of functions of a single random variable. In Section 5.4, the moment-based contextual DRO framework is presented. Most proofs of minor results are deferred to Appendix B.3. Finally, in Section 5.5, we conclude and provide several directions for future research.

5.2. A duality framework for generalized conditional-bound problems

We first describe the problem of bounding conditional expectations in Section 5.2.1 and subsequently derive the associated dual problem in Section 5.2.2. Then, in Section 5.2.3, we provide fundamental results that will be employed in later sections to obtain the desired sharp bounds.

5.2.1. Problem statement

We aim to find the best upper bound for the conditional expectation of a random vector X. Let us first introduce some notation. Let $\mathbb{E}$ denote the expectation operator, and $g(\cdot)$ denote an arbitrary measurable function of X. The random vector X is defined on the support $\Omega \subseteq \mathbb{R}^n$, which we assume is a closed set endowed with the Borel sigma algebra $\mathcal{B}_\Omega$. The random vector X is governed by a probability measure $\mathbb{P} : \mathcal{B}_\Omega \to [0, 1]$, such that for a measurable set $\mathcal{S} \in \mathcal{B}_\Omega$ we have $\mathbb{P}(\mathcal{S}) = \mathbb{P}(X \in \mathcal{S})$. Furthermore, $\mathbb{P}$ lies in some convex set of probability measures $\mathcal{P}$. Throughout this chapter, the terms "probability measure" and "probability distribution" are used interchangeably. We assume that $\Xi \in \mathcal{B}_\Omega$ is an arbitrary measurable event modeling the random event observed, pertaining to the realization of X. Let Ξ be a set with strictly positive measure $\mathbb{P}(X \in \Xi) > 0$ so that $\mathbb{E}[g(X) \,|\, \Xi]$ is well-defined and denotes the conditional expectation of X restricted to the values in the set Ξ. We now have the necessary notation to develop an adapted version of the generalized moment problem that incorporates random events or, using different terminology, side information. The central problem in this chapter can be formulated as follows:

$$\sup_{\mathbb{P} \in \mathcal{M}_+(\Omega)} \quad \mathbb{E}_\mathbb{P}[g(X) \,|\, X \in \Xi] \quad \text{subject to} \quad \mathbb{E}_\mathbb{P}[h_j(X)] = q_j \text{ for } j = 0, \dots, m, \tag{5.1}$$

where $\mathcal{M}_+(\Omega)$ denotes all nonnegative measures defined on the support Ω, and $g, h_0, \dots, h_m$ are real-valued, measurable functions that model the objective function and the available (generalized) moment information. The probability mass constraint is explicitly included as $h_0 \equiv 1$ and $q_0 = 1$. Let $\mathcal{P}$ denote the ambiguity set that contains the true probability distribution. For a given moment vector $\mathbf{q}$, define the set

$$\mathcal{P}(\mathbf{q}) := \left\{ \mathbb{P} \in \mathcal{P}_0(\Omega) : \int h_j(x) \mathrm{d}\mathbb{P}(x) = q_j, j = 0, \dots, m \right\}, \tag{5.2}$$

which contains all probability distributions that comply with the given moment sequence. Here, we typically assume $\mathbb{P}$ is an element of a set of probability measures $\mathcal{P}_0(\Omega)$ with support contained in Ω. Thus, the constraint $h_0 \equiv 1$ is implicitly assumed so as to normalize the measures in $\mathcal{M}_+(\Omega)$ to obtain probability distributions. As a closely related concept, we consider the cone of moments $\mathbf{q} \in \mathbb{R}^m$ that yield a nonempty ambiguity set $\mathcal{P}(\mathbf{q})$, which can be defined as

$$\mathcal{Q} := \{ \mathbf{q} \in \mathbb{R}^m : \exists \mathbb{P} \in \mathcal{M}_+(\Omega) \text{ such that } \mathbb{P} \in \mathcal{P}(\mathbf{q}) \}.$$

This set thus contains all moment vectors $\mathbf{q}$ for which (5.1) has a solution. That is, the moment constraints are consistent or, in other words, there exists a probability distribution feasible to the generalized moment problem. We henceforth assume that the moment constraints are consistent so that the ambiguity set $\mathcal{P}$ is nonempty, and therefore a feasible solution to problem (5.1) always exists.

Now that we have introduced the required notation and studied the constraints of (5.1), let us turn to the objective function. Instead of the regular expectation of a function of a random vector studied in generalized moment problems, $\mathbb{E}[g(X)]$, we now study

$$\mathbb{E}_\mathbb{P}[g(X) \,|\, X \in \Xi] = \int_\Xi g(x)\, \mathrm{d}\mathbb{Q}(x) = \frac{\mathbb{E}_\mathbb{P}[g(X) \mathbb{1}_\Xi(X)]}{\mathbb{E}_\mathbb{P}[\mathbb{1}_\Xi(X)]}, \tag{5.3}$$

in which $\mathbb{Q}$ denotes the conditional probability measure, given that it exists. Notice that the objective function is a fractional, and thus nonlinear, function of the probability distribution $\mathbb{P}$. As a result, (5.1) belongs to the class of distributionally robust fractional optimization problems, which are fundamentally more difficult to solve than the standard problem that simply maximizes the expectation of a function (see, e.g., [122, 145]).

5.2.2. An equivalent problem and its dual

Problem (5.1) can be formulated equivalently as a semi-infinite linear-fractional program (LFP). The semi-infinite LFP reformulation of (5.1) is given by

$$\begin{aligned} \sup_{\mathbb{P}(x)\geqslant 0} \quad & \frac{\int_\Omega g(x)\mathbb{1}_\Xi(x)\mathrm{d}\mathbb{P}(x)}{\int_\Omega \mathbb{1}_\Xi(x)\mathrm{d}\mathbb{P}(x)} \\ \text{s.t.} \quad & \int_\Omega h_j(x)\mathrm{d}\mathbb{P}(x) = q_j, \ \forall j = 0,\dots,m, \\ & \mathbb{P}(X \in \Xi) > 0, \end{aligned} \tag{5.4}$$

in which $\mathbb{1}_\Xi(x)$ equals 1 if $x \in \Xi$, and 0 otherwise. Here the optimization of the linear-fractional objective is taken over infinite-dimensional variables (i.e., the probability distribution). We further have a finite number of constraints that describe the moment information. The final constraint ensures that the conditional expectation is well defined by avoiding conditioning on a set of measure zero. In the finite setting, these linear-fractional programs can be reduced to linear programs through a Charnes-Cooper transformation [41, Theorem 2]. If we generalize this to infinite-dimensional spaces, the Charnes-Cooper transformation becomes

$$\frac{\mathrm{d}\mathbb{Q}(x)}{\mathrm{d}\mathbb{P}(x)} = \alpha, \text{ with } \alpha = \frac{1}{\int_\Omega \mathbb{1}_\Xi(x)\mathrm{d}\mathbb{P}(x)}. \tag{5.5}$$

In some sense, this generalized Charnes-Cooper transformation constitutes a change of measure, from the original probability measure $\mathbb{P}$ to its conditional counterpart $\mathbb{Q}$. The variable α is a scaling parameter that essentially models the normalization on the random event.

After transformation (5.5), problem (5.4) reduces to the semi-infinite linear program (LP)

$$\begin{aligned} \sup_{\alpha,\mathbb{Q}(x)\in\mathbb{R}_+} \quad & \int_\Omega g(x)\mathbb{1}_\Xi(x)\mathrm{d}\mathbb{Q}(x) \\ \text{s.t.} \quad & \int_\Omega h_j(x)\,\mathrm{d}\mathbb{Q}(x) = \alpha q_j, \quad \forall j = 0,\dots,m, \\ & \int_\Omega \mathbb{1}_\Xi(x)\mathrm{d}\mathbb{Q}(x) = 1, \end{aligned} \tag{5.6}$$

where all the right-hand sides of the constraints in (5.6) are scaled by α, and the last line ensures that $\mathbb{Q}$ is a proper (conditional) distribution when defined on its support Ξ. To determine the dual of (5.6), one can employ semi-infinite linear programming duality, as in Section 6 of [113].

It then follows from standard calculations that the Lagrangian dual of (5.6) is

$$\begin{aligned} \inf_{\lambda_0,\dots,\lambda_{m+1}} \quad & \lambda_{m+1} \\ \text{s.t.} \quad & \sum_{j=0}^{m} \lambda_j q_j \leqslant 0, \\ & \sum_{j=0}^{m} \lambda_j h_j(x) + \lambda_{m+1} \mathbb{1}_{\Xi}(x) \geqslant g(x)\mathbb{1}_{\Xi}(x), \ \forall x \in \Omega, \end{aligned} \tag{5.7}$$

where the dual variables $\lambda_0, \dots, \lambda_{m+1}$ are associated to each constraint in the primal (5.6). From this point on, we use the shorthand notation $h_{m+1}(x) := \mathbb{1}_{\Xi}(x)$. Notice that the dual problem has a finite number of decision variables, but an infinite number of constraints. By virtue of weak duality, an upper bound for (5.6) follows from a feasible solution to (5.7). The optimal dual solution to problem (5.7) yields a viable upper bound by weak duality, but whether this bound is sharp (i.e., whether strong duality holds) is still an open question, to which we will seek an answer in the next subsection.

5.2.3. Strong duality of the generalized conditional-bound problem

The purpose of this section is twofold: we demonstrate that (5.7) is strongly dual to (5.1), also extending this result to more general, convex sets of probability measures $\mathscr{P}_0$ that might model structural properties such as symmetry and unimodality, and we show that (5.1) can be reduced to a finite-dimensional problem in which one optimizes over a parametric family of distributions.

Since the objective function of (5.1) is nonlinear with respect to the expectation operator, it is not immediately clear how to pass down sufficient conditions regarding strong duality for generalized moment problems to the setting in which conditional expectations are considered. Therefore, in order to prove our main results, we use an alternative formulation of problem (5.4). The equivalence of this formulation is provided by the following lemma, which is an adaptation of Proposition 2.1 in [145].

Lemma 5.1. *Suppose that problem* (5.1) *has a finite optimal value. Then problem* (5.1) *is equivalent to*

$$\begin{aligned} \inf_{\tau\in\mathbb{R}} \quad & \tau \\ \text{s.t.} \quad & \sup_{\mathbb{P}\in\mathscr{P}} \mathbb{E}_{\mathbb{P}}\left[g(X)\mathbb{1}_{\Xi}(X) - \tau\mathbb{1}_{\Xi}(X)\right] \leqslant 0, \end{aligned} \tag{5.8}$$

Moreover, the optimal value of (5.8), τ^*, *is also finite.*

It turns out that (5.8) is significantly easier to work with than the original semi-infinite formulation (5.4) since the problem is linear with respect to $\mathbb{P}$, rather than linear-fractional. This can be seen from the constraints of (5.8) in which the expectation $\mathbb{E}_{\mathbb{P}}[\cdot]$ appears linearly, whereas in (5.3), we have a fraction of expectation operators. From the duality theory of moment problems, solving problem (5.8) turns out to be equivalent to solving the dual (5.7). As a consequence, solving (5.4) will also be equivalent to solving (5.7) due to the equivalence between (5.4) and (5.8). We next show strong duality holds. To this end, we make the following assumptions:

(A1) The function $g(x)$ is bounded on the support Ω.

(A2) There exists a positive number $\epsilon > 0$ such that

$$\inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\Xi}(X)] \geqslant \epsilon.$$

(A3) The Slater condition holds; that is, $\mathbf{q} \in \mathrm{int}(\mathcal{Q}_{\mathcal{P}})$, where "int" denotes the interior of a set.

Assumption (A1) is standard and could be relaxed. It is satisfied, for example, when Ω is a compact set and $g(x)$ is upper-semicontinuous, by virtue of Weierstrass' Extreme Value Theorem. This assumption is used to guarantee that the optimal value of (5.4) is finite. Assumption (A2) provides a sufficient condition for the conditional expectation to be well-defined, as it avoids conditioning on a set of measure zero. Assumption (A3) constitutes a Slater-type condition on the moments of X, which is standard for generalized moment problems; see, for example, Proposition 3.4 in [196]. Under these regularity conditions, we show strong duality holds for a general, convex set of probability distributions $\mathcal{P}_0$, possibly endowed with structural properties (e.g., symmetry and unimodality). We will focus on structural ambiguity sets $\mathcal{P}_0(\Omega)$ that possess a mixture representation. In other words, we assume $\mathcal{P}_0$ can be "generated" by a convenient class of distributions, say $\mathcal{T}$, such that every distribution $\mathbb{P} \in \mathcal{P}_0$ can be written as a mixture (i.e., an infinite convex combination) of the extremal distributions (i.e., the extreme points of the convex set $\mathcal{P}_0$) that constitute $\mathcal{T}$. For every Borel set $\mathcal{S} \in \mathcal{B}_\Omega$, it should thus hold that

$$\mathbb{P}(X \in \mathcal{S}) = \int_{\mathcal{T}} \mathbb{T}_{\mathbf{t}}(X \in \mathcal{S})\mathrm{d}\mathbb{M}(\mathbf{t})$$

where $\mathbb{T}_{\mathbf{t}} \in \mathcal{T} \subseteq \mathcal{P}_0$, is a parametrized representation of the family of extremal distributions of $\mathcal{P}_0$, and $\mathbb{M}$ represents the mixture distribution that generates $\mathbb{P}$ from the extremal distributions in $\mathcal{T}$. This finite-dimensional parameterization of the family of extremal distributions will prove useful when determining the optimal bounds. For a thorough discussion on these structural ambiguity sets in the context of DRO, we refer to the work of [177]. We use these general ambiguity sets with structural properties when formulating our main results.

Finally, before formulating our main result, we introduce some final technical notation from conic duality theory for generalized moment problems (see, e.g., [177, 196]). Denote by $\mathcal{A} = \mathrm{co}(\mathcal{P}_0)$ the cone of measures $\mathcal{A}$ generated by the set of probability distributions $\mathcal{P}_0$. Define its dual cone as $\mathcal{A}^* := \left\{h \in \mathcal{H} : \int_\Omega h(x)\,\mathrm{d}\mathbb{P}(\mathrm{x}) \geqslant 0,\ \forall \mathbb{P} \in \mathcal{A}\right\}$, where $\mathcal{H}$ is the linear space of functions formed by combinations of $g, h_0, \dots, h_m$, and the spaces of functions and measures are paired by the integral operator. We now have the necessary preliminaries to demonstrate strong (conic) duality. Lemma 5.1, in conjunction with assumptions (A1)–(A3), pave the way for us to formulate the main results.

Theorem 5.2 **(Strong conic duality).** *Suppose that assumptions* (A1)–(A3) *hold. Then, the op-*

timal value of the primal problem (5.1) *is finite and equals that of its dual problem*

$$\begin{aligned} \inf_{\lambda_0,\ldots,\lambda_{m+1}} \quad & \lambda_{m+1} \\ \text{s.t.} \quad & \sum_{j=0}^{m} \lambda_j q_j \leqslant 0, \\ & \sum_{j=0}^{m+1} \lambda_j h_j(x) - g(x)\mathbb{1}_\Xi(x) \in \mathscr{A}^*, \end{aligned} \tag{5.9}$$

in which $\mathscr{A}^*$ *is the dual cone of* $\mathscr{A}$.

Proof. The assumptions ensure that, for all $\mathbb{P} \in \mathscr{P}$,

$$\begin{aligned} \left| \frac{\mathbb{E}_\mathbb{P}[g(X)\mathbb{1}_\Xi(X)]}{\mathbb{E}_\mathbb{P}[\mathbb{1}_\Xi(X)]} \right| &\leqslant \frac{1}{\epsilon} |\mathbb{E}_\mathbb{P}[g(X)]| \\ &\leqslant \sup_{\mathbb{P}\in\mathscr{P}} \frac{1}{\epsilon} |\mathbb{E}_\mathbb{P}[g(X)]| \\ &\leqslant \frac{1}{\epsilon} \sup_{x\in\Omega} |g(x)| < \infty. \end{aligned}$$

Hence, instead of (5.1), it is equivalent to consider (5.8). Since the constraints of this problem are linear in the probability distribution $\mathbb{P}$, standard conic duality for generalized moment problems suffices, which holds under the Slater-type condition, as in [196]. Notice that if we substitute τ with λ_3, the dual of (5.8) is equivalent to (5.9). For $\mathscr{A} = \mathscr{M}_+$, the strong conic dual problem (5.7) reduces to the semi-infinite LP (5.7). The result for general cones of measures follows from conic duality arguments as in Theorem 3.1 of [177]. ☐

As a consequence, duality enables us to reduce the primal problem, which has infinite-dimensional variables, to a dual problem with $m + 1$ variables, but with an infinite number of constraints. These constraints are indexed by the probability measures, i.e., the constraints should hold $\forall\mathbb{P} \in \mathscr{P}_0$. This indexation might turn out to be difficult. However, this difficulty can be greatly reduced if we instead use the generating set $\mathscr{T}$, as shown in [177]. In this case, the dual cone can be reduced to $\mathscr{A}^* = \left\{h \in \mathscr{H} : \int_\Omega h(x)\, \mathrm{d}\mathbb{P}(x) \geqslant 0,\ \forall\mathbb{P} \in \mathscr{T}\right\}$, and hence, the indexing now only runs over the set of extreme points of $\mathscr{P}_0$.

Another classical result is that the semi-infinite LP that models a generalized moment problem can be reduced to an equivalent finite-dimensional problem that yields the same optimal value. The Richter-Rogosinski theorem (see, e.g., [98, 186, 201]) states that there exists an extremal distribution for the semi-infinite LP with at most $m + 1$ support points. Analogous to the basic solutions for conventional semi-infinite linear programming, we define a basic distribution as a convex combination of extremal distributions of $\mathscr{P}_0$. As there are $m + 1$ moment functions, these basic distributions consist of at most $m+1$ extreme points (i.e., elements of $\mathscr{T}$). We let $\mathscr{D}(\mathbf{q})$ denote the set of basic distributions that comply with the given moment sequence $\mathbf{q}$. We can then state the following result, which is an adaptation of the fundamental theorem for convex classes of distributions, but now for the generalized conditional-bound problem.

Theorem 5.3 **(Fundamental theorem for conditional expectations).** *Consider problem* (5.1). *Under assumptions* (A1)–(A3),

$$\sup_{\mathbb{P}\in\mathscr{P}(\mathbf{q})} \mathbb{E}_{\mathbb{P}}[g(X)\,|\,X\in\Xi] = \sup_{\mathbb{P}\in\mathscr{D}(\mathbf{q})} \mathbb{E}_{\mathbb{P}}[g(X)\,|\,X\in\Xi].$$

Moreover, if the optimal value is attained, then there exists an optimal basic distribution, a convex combination of $m+1$ probability distributions from the generating set $\mathscr{T}$, that achieves this value.

Proof. Again, under the stated assumptions, it suffices to consider the equivalent problem (5.8). Since the constraints of (5.8) are linear in $\mathbb{P}$, standard generalized moment problem results apply. Hence, Theorem 3.2 in [177] yields the result. □

This theorem states that, even if the bound is not achieved, it is sufficient to consider only the basic feasible distributions in $\mathscr{D}(\mathbf{q})$ to determine the supremum. This result holds for general convex classes of distributions $\mathscr{P}$ with the optimal distributions taken as convex combinations of the extremal distributions in $\mathscr{T}$. We further remark that both theorems also hold in the setting in which the supremum of the conditional expectation grows infinitely large. In this case, a maximizing sequence of probability measures exists (also taken from the set of basic distributions) for which the conditional expectation diverges. Combined, the concept of weak duality described in Section 5.2.2 and the reduction to the basic distributions generated by $\mathscr{T}$ as proposed in Theorem 5.3 provide an effective way of solving problem (5.1), as will be demonstrated in the next section.

5.3. Tight bounds for conditional expectations

In this section, we study several easy examples for the case with $n=1$; that is, X is a random variable conditioned on itself. First, in Section 5.3.1, we seek the best possible bounds for conditional expectation $\mathbb{E}[X\,|\,X\geqslant t]$ when mean-dispersion information, and possible structural properties of the underlying distribution, are given. We find the tight bounds using primal-dual arguments. In Section 5.3.2, we demonstrate this also works for arbitrary choices of $g(x)$, using an example from the robust pricing literature. Finally, Section 5.3.3 shows that, in the univariate setting, tight bounds can be obtained by solving semidefinite programming problems.

5.3.1. Simple examples for mean-dispersion information

For the sake of exposition, we concentrate our efforts on the event $\Xi=\{X\geqslant t\}$. We thus seek to bound the conditional expectation

$$\mathbb{E}_{\mathbb{P}}[g(X)\,|\,X\geqslant t] = \frac{\int_\Omega g(x)\mathbb{1}_\Xi(x)\mathrm{d}\mathbb{P}(x)}{\int_\Omega \mathbb{1}_\Xi(x)\mathrm{d}\mathbb{P}(x)},$$

in which $\mathbb{1}_\Xi(x)$ is the indicator function modeling the occurrence of the event $\{X\geqslant t\}$ and $\mathbb{P}$ is the underlying probability distribution of which we assume that it lies in the mean-variance ambiguity set, $\mathscr{P}_{(\mu,\sigma)}$, which contains all distributions that comply with the available mean-variance information. Then, the problem of interest can be stated as

$$\begin{aligned} \sup_{\mathbb{P}(x)\geqslant 0} \quad & \frac{\int_\Omega g(x)\mathbb{1}_\Xi(x)\mathrm{d}\mathbb{P}(x)}{\int_\Omega \mathbb{1}_\Xi(x)\mathrm{d}\mathbb{P}(x)} \\ \text{s.t.} \quad & \int_\Omega \mathrm{d}\mathbb{P}(x) = 1, \int_\Omega x\,\mathrm{d}\mathbb{P}(x) = \mu, \int_\Omega x^2\,\mathrm{d}\mathbb{P}(x) = (\sigma^2+\mu^2), \end{aligned} \tag{5.10}$$

which is a semi-infinite LFP. Through the generalized Charnes-Cooper transformation, introduced in Section 5.2.1, it is possible to write (5.10) as

$$\begin{aligned} \sup_{\alpha,\mathbb{Q}(x)\geqslant 0} \quad & \int_\Omega g(x)\mathbb{1}_\Xi(x)\mathrm{d}\mathbb{Q}(x) \\ \text{s.t.} \quad & \int_\Omega \mathrm{d}\mathbb{Q}(x) = \alpha, \int_\Omega x\,\mathrm{d}\mathbb{Q}(x) = \alpha\mu, \int_\Omega x^2\,\mathrm{d}\mathbb{Q}(x) = \alpha(\sigma^2+\mu^2), \int_\Omega \mathbb{1}_\Xi(x)\mathrm{d}\mathbb{Q}(x) = 1, \end{aligned} \tag{5.11}$$

where α is a decision variable denoting the "worst-case" probability of X exceeding t. The semi-infinite linear programming dual of (5.11) is given by

$$\begin{aligned} \inf_{\lambda_0,\lambda_1,\lambda_2,\lambda_3} \quad & \lambda_3 \\ \text{s.t.} \quad & \lambda_0 + \lambda_1\mu + \lambda_2(\sigma^2+\mu^2) \leqslant 0, \\ & \lambda_0 + \lambda_1 x + \lambda_2 x^2 + \lambda_3\mathbb{1}_\Xi(x) \geqslant g(x)\mathbb{1}_\Xi(x),\ \forall x \in \mathbb{R}, \end{aligned} \tag{5.12}$$

which, for this specific choice of random event Ξ, results in

$$\begin{aligned} \inf_{\lambda_0,\lambda_1,\lambda_2,\lambda_3} \quad & \lambda_3 \\ \text{s.t.} \quad & \lambda_0 + \lambda_1\mu + \lambda_2(\sigma^2+\mu^2) \leqslant 0, \\ & \lambda_0 + \lambda_1 x + \lambda_2 x^2 \geqslant 0, && \forall x < t, \\ & \lambda_0 + \lambda_1 x + \lambda_2 x^2 + \lambda_3 \geqslant g(x), && \forall x \geqslant t. \end{aligned}$$

Consider the standard conditional expectation (i.e., consider the function $g : x \mapsto x$). We next try to find feasible solutions to the dual problem, and prove optimality by finding primal solutions with matching objective values. Let $t < \mu$. The dual problem of $\sup_{\mathbb{P}\in\mathscr{P}_{(\mu,\sigma)}} \mathbb{E}[X \mid X \geqslant t]$ is given by

$$\begin{aligned} \inf_{\lambda_0,\lambda_1,\lambda_2,\lambda_3} \quad & \lambda_3 \\ \text{s.t.} \quad & \lambda_0 + \lambda_1\mu + \lambda_2(\sigma^2+\mu^2) \leqslant 0, \\ & \lambda_0 + \lambda_1 x + \lambda_2 x^2 \geqslant 0, && \forall x < t, \\ & \lambda_0 + \lambda_1 x + \lambda_2 x^2 \geqslant x - \lambda_3, && \forall x \geqslant t. \end{aligned} \tag{5.13}$$

Denote the left-hand sides of the constraints by $M(x) := \lambda_0 + \lambda_1 x + \lambda_2 x^2$. The function $M(x)$ is dual feasible when it is greater than or equal to 0 for $x < t$ and greater than or equal to $x - \lambda_3$ for $t \leqslant x \leqslant b$. We construct two candidate dual solutions, the quadratic functions $M_1(\cdot)$ and

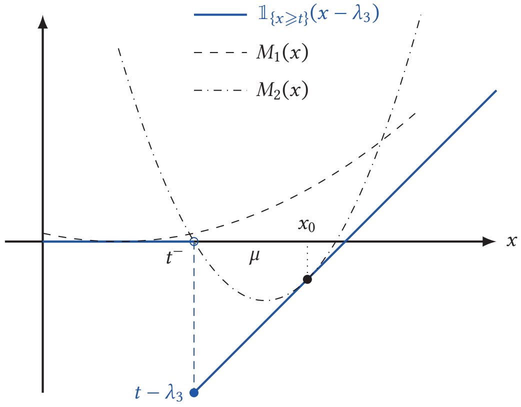


Figure 5.1: $M_1(x)$ and $M_2(x)$

$M_2(\cdot)$ (see Figure 5.1), with the former's minimum occurring below t, while for the latter, its minimum occurs above the threshold t.

The first dual function, $M_1(\cdot)$, does not admit a feasible solution. Since $\sigma > 0$, from $\lambda_0 + \lambda_1\mu + \lambda_2(\sigma^2 + \mu^2) \leqslant 0$ it follows that $M(\mu) = \lambda_0 + \lambda_1\mu + \lambda_2\mu^2 < 0$, but for the minimizer $x^* = -\lambda_1/(2\lambda_2)$, $M(x^*) \geqslant 0$. Thus, this case is infeasible.

The second parabola, $M_2(\cdot)$, does admit a feasible solution. This function coincides with the function $g(\cdot)$ at t and some point x_0. Since $M_2(t) = 0$, $M_2(x_0) = x_0 - \lambda_3$ and $M_2'(x_0) = 1$,

$$\lambda_0 = -\frac{t(\lambda_3(t - 2x_0) + x_0^2)}{(t - x_0)^2}, \; \lambda_1 = \frac{t^2 + x_0(x_0 - 2\lambda_3)}{(t - x_0)^2}, \; \lambda_2 = \frac{\lambda_3 - t}{(t - x_0)^2}. \tag{5.14}$$

Hence, this yields an optimization problem in two variables, $\min_{x_0,\lambda_3} \lambda_3$, with the additional constraint $\lambda_0 + \lambda_1\mu + \lambda_2(\sigma^2 + \mu^2) \leqslant 0$, in which we substitute (5.14). If this constraint is tight, the dual problem can be reduced to

$$\min_{\lambda_3, x_0} \lambda_3 \equiv \min_{x_0} \frac{(t^2 + x_0^2)\mu - t(x_0^2 + \mu^2 + \sigma^2)}{(t - \mu)(t - 2x_0 + \mu) - \sigma^2}.$$

Optimizing over x_0, it then follows that

$$x_0^* = \mu + \frac{\sigma^2}{\mu - t}, \; \lambda_3^* = \mu + \frac{\sigma^2}{\mu - t},$$

with λ_3^* a feasible upper bound for $\mathbb{E}[X|X \geqslant t]$. To prove this bound is optimal, we construct a distribution that (asymptotically) achieves λ_3^*. From complementary slackness (see, e.g., [201]), we deduce that the candidate distribution has all of its probability mass on two points: t^- and x_0^*. Solving the system of moment constraints in (5.10) yields the probabilities

$$p_{t^-} = \frac{\sigma^2}{\sigma^2 + (\mu - t)^2}, \; p_{x_0^*} = \frac{(\mu - t)^2}{\sigma^2 + (\mu - t)^2}$$

Indeed, for this two-point distribution,

$$\mathbb{E}[X \mid X \geqslant t] = \frac{x_0^* \cdot p_{x_0^*}}{\mathbb{P}(X \geqslant t)} = \frac{x_0^* \cdot p_{x_0^*}}{p_{x_0^*}} = x_0^* = \mu + \frac{\sigma^2}{\mu - t},$$

ensuring the upper bound is tight. Hence, by weak duality,

$$\sup_{\mathbb{P} \in \mathcal{P}_{(\mu,\sigma)}} \mathbb{E}[X \mid X \geqslant t] = \mu + \frac{\sigma^2}{\mu - t}. \tag{5.15}$$

Notice that this bound is not actually attained, but it is achieved in the limit; that is, for $t_k = t - \frac{1}{k}$, as $k \to \infty$, this construction of the extremal distribution indeed gives the desired result. For $t \geqslant \mu$, it can be shown that

$$\sup_{\mathbb{P} \in \mathcal{P}_{(\mu,\sigma)}} \mathbb{E}[X \mid X \geqslant t] = \infty,$$

since we can construct the following sequence of (maximizing) measures:

$$\mathbb{P}_k = \frac{1}{k^2\sigma^2 + 1} \delta_{\mu + k\sigma^2} + \frac{\sigma^2}{\sigma^2 + \frac{1}{k^2}} \delta_{\mu - \frac{1}{k}}.$$

Letting $k \to \infty$ then results in $\mathbb{E}_{\mathbb{P}_k}[X \mid X \geqslant t] \to \infty$. Taken together, we obtain the following result.

Proposition 5.4. *For a real-valued random variable X with distribution $\mathbb{P} \in \mathcal{P}_{(\mu,\sigma)}$, it holds that*

$$\sup_{\mathbb{P} \in \mathcal{P}_{(\mu,\sigma)}} \mathbb{E}_{\mathbb{P}}[X \mid X \geqslant t] = \begin{cases} \mu + \frac{\sigma^2}{\mu - t}, & \text{for } t < \mu, \\ \infty, & \text{for } t \geqslant \mu. \end{cases} \tag{5.16}$$

A number of interesting observations can be drawn from this result. First, note that the maximizing sequence for the second case, $\{\mathbb{P}_k\}$, converges weakly to δ_μ, which is *not* included in the ambiguity set. Notice also that the solution to the second case becomes "uninformative," as it diverges for values of $t \geqslant \mu$. Degenerate behavior like this also holds for different ambiguity sets, as we will see in later examples. This result confirms tightness of the Mallows and Richter bound for conditional expectations under mean-variance information, stated in [151]. Further, notice that the worst-case distribution that achieves the upper bound matches the extremal distribution that yields the Cantelli lower bound for the tail probability, as also shown in [88]. The authors of the latter work provide a constructive proof of tightness using the extremal distribution that achieves the Cantelli bound. Despite Proposition 5.4 being a known result, this is the first instance in which it has been proven through a duality argument that provides immediate insight into the extremal distributions.

The method discussed above can be applied to different types of dispersion information, not only the traditional variance. For example, assume that instead of the variance, we consider the mean absolute deviation from the mean (MAD), $d := \mathbb{E}|X - \mu|$, as the measure of dispersion. Let $\mathcal{P}_{(\mu,d)}$ denote the mean-MAD ambiguity set, with the additional constraint that the support of X is (a subset of) the interval $[a, b]$. We can then prove the following result.

Proposition 5.5. *For a real-valued random variable X with distribution $\mathbb{P} \in \mathcal{P}_{(\mu,d)}$, it holds that*

$$\sup_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}[X \mid X \geqslant t] = \begin{cases} \mu + \frac{d(\mu-t)}{2(\mu-t)-d}, & \text{for } t < \mu - \frac{d(b-\mu)}{2(b-\mu)-d}, \\ b, & \text{for } t \geqslant \mu - \frac{d(b-\mu)}{2(b-\mu)-d}. \end{cases} \tag{5.17}$$

Again, we see that the second case becomes uninformative, as it simply reduces to the robust solution (i.e., it agrees with the upper bound of the support). It is worth noting here the interplay between the size of the ambiguity set and the set that describes the random event/side information. Despite having only a limited number of distributions to choose from, if the realizations of the random variable are limited to too small an interval, the bounds become overly conservative, as an extremal distribution can be constructed for which the moment constraints are satisfied, yet the support point on Ξ can be made arbitrarily large, bounded, of course, by the upper bound of the support. For the mean-MAD ambiguity set, the extremal distribution also agrees with the distribution attaining the lower bound on the corresponding tail probability [188].

Proposition 5.5 can be generalized further. Assume that the dispersion information is modeled through the expectation of a convex function $d(\cdot)$ of the random variable X, defined as $\bar{\sigma} := \mathbb{E}[d(X)]$. We can state the following result for such arbitrary convex dispersion measures.

Proposition 5.6. *Suppose that there exists a solution x_0^* to the equation*

$$\frac{\bar{\sigma}t - \mu d(t)}{td(x_0) - x_0 d(t)} v + \frac{\mu d(x_0) - \bar{\sigma}x_0}{td(x_0) - x_0 d(t)} = 1,$$

such that the corresponding two-point distribution, with support $\{t, x_0^*\}$, is feasible. Then, for a real-valued random variable X with distribution $\mathbb{P} \in \mathcal{P}_{(\mu,\bar{\sigma})}$, it holds that

$$\sup_{\mathbb{P}\in\mathcal{P}_{(\mu,\bar{\sigma})}} \mathbb{E}_{\mathbb{P}}[X \mid X \geqslant t] = x_0^*. \tag{5.18}$$

Proposition 5.6 covers a wide range of dispersion measures, not limited to only variance and MAD. It also incorporates asymmetric measures of dispersion, such as semivariance, semimean deviations, and partial moments. More generally, it encompasses all dispersion measures that are modeled using piecewise convex functions. Since the p-norms on $\mathbb{R}$ are convex, these are naturally included in the category of convex dispersion measures as well. Another notable function that falls into this class is the Huber-loss function, which has been extensively studied in the field of robust statistics.

As explained earlier, using only moment information often leads to overly conservative bounds and pathological worst-case distributions. We require additional assumptions about the distribution's shape to sharpen the bounds and avoid the pathological two-point distributions that constitute the worst-case scenario in the previous examples. We next study two such structural properties, i.e., symmetry and unimodality. The random variable X is said to admit a symmetric distribution about a point m if $\mathbb{P}(X \in [m-x, m]) = \mathbb{P}(X \in [m, m+x])$ for all $x \geqslant 0$. A random variable X has a unimodal distribution with mode m if its distribution function is a concave function on $(-\infty, m]$ and convex on (m, ∞). Both definitions are generalized so that they admit probability distributions that allow for point masses at m. We next consider a distribution that is symmetric about its mean μ with the values of the mean and variance given. Making use of primal-dual arguments, we obtain the following result.

Proposition 5.7. *For a real-valued random variable X with a symmetric distribution $\mathbb{P} \in \mathscr{P}^{\text{sym}}_{(\mu,\sigma)}$, it holds that*

$$\sup_{\mathbb{P}\in\mathscr{P}^{\text{sym}}_{(\mu,\sigma)}} \mathbb{E}_{\mathbb{P}}[X \mid X \geqslant t] = \begin{cases} \mu + \frac{(\mu-t)\sigma^2}{2(t-\mu)^2-\sigma^2}, & \text{for } t < \mu - \sigma, \\ \mu + \sigma, & \text{for } \mu - \sigma \leqslant t < \mu, \\ \infty, & \text{for } t \geqslant \mu. \end{cases} \tag{5.19}$$

Observe that the bounds are sharper than the bound in Proposition 5.4, but still vacuous for $t \geqslant \mu$. Combining the notions of symmetry and unimodality yields the following, even tighter, bounds:

Proposition 5.8. *For a real-valued random variable X with a symmetric, unimodal distribution $\mathbb{P} \in \mathscr{P}^{\text{uni}}_{(\mu,\sigma)}$, it holds that*

$$\sup_{\mathbb{P}\in\mathscr{P}^{\text{uni}}_{(\mu,\sigma)}} \mathbb{E}_{\mathbb{P}}[X \mid X \geqslant t] = \begin{cases} \frac{4\mu(x_0^*)^3-3\sigma^2(t+x_0^*-\mu)(t-x_0^*+\mu)}{4(x_0^*)^3-6\sigma^2(t+x_0^*-\mu)}, & \text{for } t < \mu - \frac{3\sqrt{3}\sigma}{5}, \\ \frac{1}{2}\left(\mu + t + \sqrt{3}\sigma\right), & \text{for } \mu - \frac{3\sqrt{3}\sigma}{5} < t < \mu, \\ \infty, & \text{for } t \geqslant \mu, \end{cases} \tag{5.20}$$

where x_0^* is the real-valued solution to the quartic equation

$$6\sigma^2 x_0^2 \left(3(t-\mu)^2 - x_0^2\right) + 9\sigma^4(\mu - t - x_0)^2 = 0,$$

which satisfies the condition $x_0^* \geqslant 3\sigma$.

Although Proposition 5.8 does not provide a closed-form solution, it does demonstrate the versatility of the primal-dual arguments used to derive it. It further highlights that structural properties can be addressed in conjunction with moment information, even for conditional-bound problems.

5.3.2. A robust pricing objective function

To demonstrate the primal-dual approach for an alternative objective function $g(x)$, we next turn our attention to a specific minimax problem. We consider the objective of the robust monopoly-pricing problem; see, e.g., [45, 75, 88]. This involves evaluating the revenue function

$$\Pi(p) := \mathbb{E}[p\mathbb{1}_{\{X\geqslant p\}}] = p\mathbb{P}(X \geqslant p),$$

which models the expected revenue that a seller of a single item receives when the price posted is equal to p, and the valuation of customers is distributed as X. As in [88], we will attempt to minimize the maximum relative regret by posting the minimax selling price p^*; that is, we solve

$$\min_p \max_{\mathbb{P}\in\mathcal{P}} \frac{\max_z \mathbb{E}_{\mathbb{P}}[\Pi(z)]}{\mathbb{E}_{\mathbb{P}}[\Pi(p)]}.$$

We use the "min" and "max" operators only to avoid notational clutter, as it does not imply that the optima are actually attained. Chen et al. [45] present various results for robust monopoly pricing and also consider this relative regret criterion. In [45] an almost identical objective function is considered, namely,

$$\min_p \max_{\mathbb{P}} \left\{ 1 - \frac{\mathbb{E}_{\mathbb{P}}[\Pi(p)]}{\max_z \mathbb{E}_{\mathbb{P}}[\Pi(z)]} \right\} = 1 - \max_p \min_{\mathbb{P}} \frac{\mathbb{E}_{\mathbb{P}}[\Pi(p)]}{\max_z \mathbb{E}_{\mathbb{P}}[\Pi(z)]}$$

We, however, work with the reciprocal

$$\min_p \max_{\mathbb{P}} \frac{\max_z \mathbb{E}_{\mathbb{P}}[\Pi(z)]}{\mathbb{E}_{\mathbb{P}}[\Pi(p)]} = \min_p \max_{\mathbb{P}} \max_z \frac{\mathbb{E}_{\mathbb{P}}[\Pi(z)]}{\mathbb{E}_{\mathbb{P}}[\Pi(p)]} = \min_p \max_z \max_{\mathbb{P}} \frac{\mathbb{E}_{\mathbb{P}}[\Pi(z)]}{\mathbb{E}_{\mathbb{P}}[\Pi(p)]},$$

where we swap the maximization operators to obtain the final equality. As in the proof of Theorem 4 in [45], it is imperative to solve the semi-infinite optimization problem

$$\max_{\mathbb{P}\in\mathcal{P}_{(\mu,\sigma)}} \frac{\mathbb{E}_{\mathbb{P}}[z\mathbb{1}\{X \geqslant z\}]}{p\mathbb{P}(X \geqslant p)}. \tag{5.21}$$

Notice that this problem effectively models the conditional expectation $\mathbb{E}[\frac{z}{p}\frac{\mathbb{1}\{X\geqslant z\}}{\mathbb{1}\{X\geqslant p\}} \,|\, X \in \Xi]$ with the random event $\Xi = \{X \geqslant p\}$. As [45] tries to optimize the dual problem directly, their proof requires several lengthy, tedious arguments to obtain the tight bounds. We will simplify their proof using primal-dual arguments as in the previous subsection. We assume that $p < \mu$, as otherwise, it is possible to construct an extremal distribution for which the relative regret ratio diverges; see [88]. For $z < p$, the expectation in the numerator will evaluate to 1. Hence,

$$\max_{\mathbb{P}\in\mathcal{P}_{(\mu,\sigma)}} \frac{\mathbb{E}_{\mathbb{P}}[\frac{z}{p}\mathbb{1}\{X \geqslant z\}]}{\mathbb{P}(X \geqslant p)} \leqslant \frac{z}{p} \max_{\mathbb{P}\in\mathcal{P}_{(\mu,\sigma)}} \frac{1}{\mathbb{P}(X \geqslant p)} = \frac{z}{p} \frac{1}{\min_{\mathbb{P}\in\mathcal{P}_{(\mu,\sigma)}} \mathbb{P}(X \geqslant p)}.$$

To bound the latter, we can use the one-sided version of Chebyshev's inequality (commonly known as Cantelli's inequality). It then follows that

$$\max_{\mathbb{P}\in\mathcal{P}_{(\mu,\sigma)}} \frac{\mathbb{E}_{\mathbb{P}}[\frac{z}{p}\mathbb{1}\{X \geqslant z\}]}{\mathbb{P}(X \geqslant p)} = \frac{z}{p} \frac{\sigma^2 + (\mu - p)^2}{(\mu - p)^2}.$$

Hence,

$$\max_{z\leqslant p} \max_{\mathbb{P}} \frac{\mathbb{E}_{\mathbb{P}}[\frac{z}{p}\mathbb{1}\{X \geqslant z\}]}{p\mathbb{P}(X \geqslant p)} = \frac{\sigma^2 + (\mu - p)^2}{(\mu - p)^2}.$$

For $z \geqslant p$, it holds that

$$\frac{\mathbb{E}_{\mathbb{P}}[\frac{z}{p}\mathbb{1}\{X \geqslant z\}]}{\mathbb{P}(X \geqslant p)} = \frac{z}{p}\frac{\mathbb{P}(X \geqslant z)}{\mathbb{P}(X \geqslant p)} = \frac{z}{p}\frac{\mathbb{P}(X \geqslant z \cap X \geqslant p)}{\mathbb{P}(X \geqslant p)} = \frac{z}{p}\mathbb{P}(X \geqslant z \,|\, X \geqslant p),$$

and thus, we require a bound for conditional probabilities. Using primal-dual arguments, we obtain the following result.

Proposition 5.9. *Suppose that $z \geqslant p$ and $p < \mu$. For a nonnegative random variable X with distribution $\mathbb{P} \in \mathscr{P}_{(\mu,\sigma)}$, it holds that*

$$\sup_{\mathbb{P}\in\mathscr{P}_{(\mu,\sigma)}} \mathbb{P}(X \geqslant z \,|\, X \geqslant p) = \begin{cases} \frac{\sigma^2}{(z-\mu)^2+\sigma^2}, & \text{for } z \geqslant \mu + \frac{2\sigma^2(\mu-p)}{(\mu-p)^2+\sigma^2}, \\ \left(\frac{(\mu-p)^2+\sigma^2}{\sigma^2+\mu^2-p^2+2z(\mu-p)}\right)^2, & \text{for } \frac{\sigma^2+\mu^2-p\mu}{\mu-p} \leqslant z \leqslant \mu + \frac{2\sigma^2(\mu-p)}{(\mu-p)^2+\sigma^2}, \\ 1, & \text{otherwise.} \end{cases} \tag{5.22}$$

Using these bounds, one can show that

$$\max_{z\geqslant p} \sup_{\mathbb{P}} \frac{\mathbb{E}_{\mathbb{P}}[\pi(z,X)]}{p\mathbb{P}(X \geqslant p)} = \frac{\sigma^2 + \mu(\mu-p)}{p(\mu-p)}.$$

Hence, combining the above results, the optimal price should solve

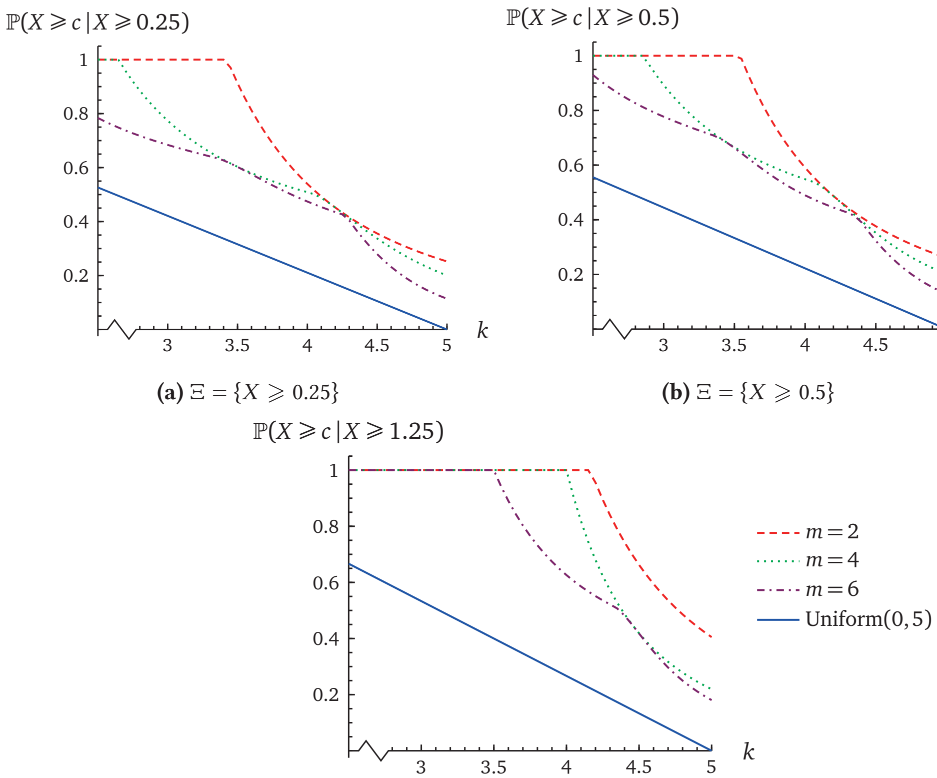
$$\min_{p<\mu} \max \left\{ \frac{\sigma^2}{p(\mu-p)} + \frac{\mu}{p}, \frac{\sigma^2}{(\mu-p)^2} + 1 \right\}.$$

It is shown by [88] that the optimal price is the value of p on which both branches of the max operator agree. Observe that by using primal-dual arguments, we have greatly reduced the number of necessary calculations to obtain the desired result. Contrary to [45], we work with the primal and dual problems concurrently so as to verify the optimality of the suggested solutions in a more effective manner. It goes without saying that the proposed duality approach in this section probably also works for other pricing problems with fractional objectives, such as e.g. the personalized pricing setting in [75].

5.3.3. Numerical bounds

We next show how the strong dual problem (5.7) can be reduced to a semidefinite programming problem for the univariate setting. Consequently, the tight bounds can always be obtained numerically as the solution to a (computationally tractable) optimization problem. The numerical experiments have been conducted in the Julia programming language, using the MOSEK solver [158] together with the Julia packages SumOfSquares.jl and PolyJuMP.jl [221].

Assume the objective function $g(x)$ is piecewise polynomial and the moment constraints are described by the traditional power moments. Then the dual problem can be reduced further to a semidefinite programming problem by applying standard DRO techniques discussed in, for example, [28, 29, 177]. In particular, the univariate moment problem reduces to solving a semidefinite program, provided that the dual-feasible set is semi-algebraic; that is, the dual constraint involves checking whether certain polynomial functions are nonnegative on intervals described by the event set Ξ and the support Ω. It is well known that a univariate polynomial is nonnegative if, and only if, it is a sum of squares. A classic result then states that the semi-infinite constraint in the dual-feasible set of (5.7), with the support Ω a possibly infinite interval, can be reduced to a set of linear matrix inequalities (LMIs) of polynomial size in the number of moments m (see, e.g., [28, 29, 167]). Since $g(x)$ is piecewise polynomial, the support of the dual constraints in (5.7) can be subdivided into subintervals so that these constraints can be reduced to a set of LMIs. Generalized moment information can be included if these moments are described by piecewise polynomials in the dual problem. Examples of piecewise polynomial objective functions are the indicator function $g(x) = \mathbb{1}_{[c,\infty)}(x)$ and the stop-loss function $g(x) = \max\{x - c, 0\}$. Both of these functions have several relevant applications in e.g. finance, insurance and inventory control.



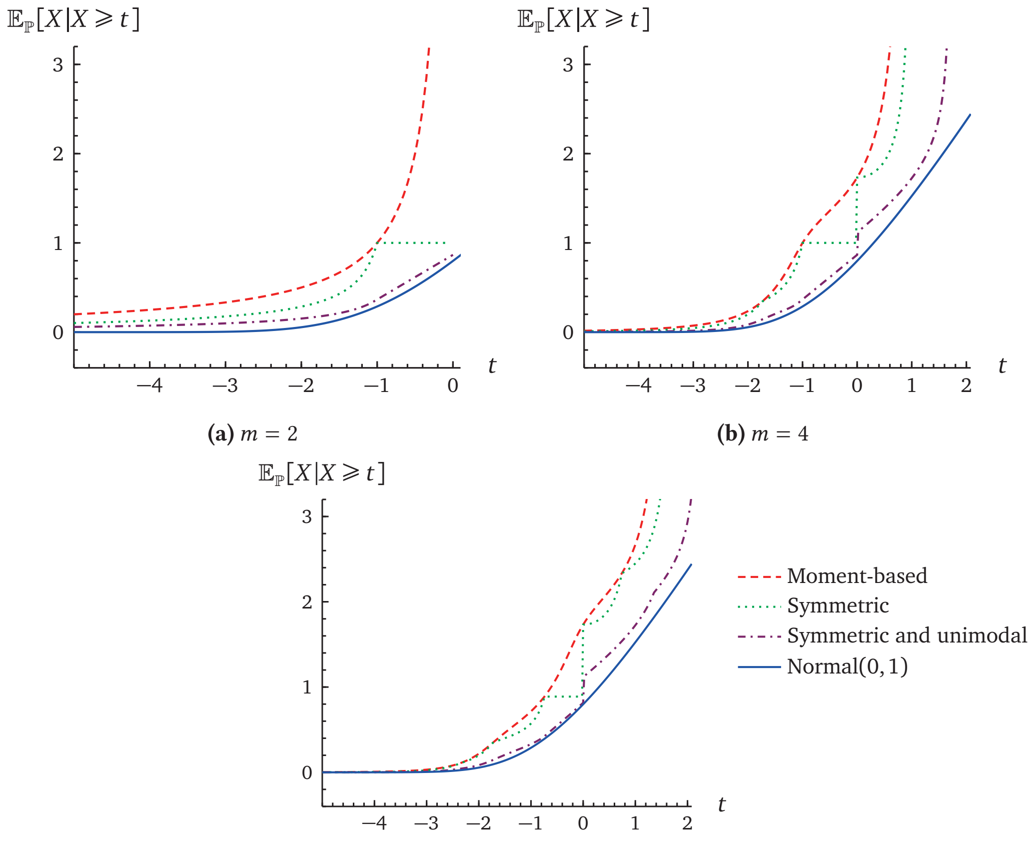
(a) $\Xi = \{X \geqslant 0.25\}$

(b) $\Xi = \{X \geqslant 0.5\}$

(c) $\Xi = \{X \geqslant 1.25\}$

Figure 5.2: Tight bounds for conditional tail probability for ambiguity sets matching the uniform distribution on $[0, 5]$

In Figure 5.2 we provide numerical bounds for the conditional probability $\mathbb{P}(X \geqslant c \,|\, X \geqslant t)$, which corresponds to the piecewise function $g(x) = \mathbb{1}_{[c,\infty)}(x)$. We assume that the ground truth is given by a uniform distribution with support $[0, 5]$. We determine the bounds for three types of ambiguity sets, in which the number of available moments varies between $m = 2, 4$, and 6. Obviously, the bounds become sharper when more moment information is included, but in addition, the uninformative solution becomes less prominent because the size of the ambiguity set reduces. Nevertheless, the uninformative solution becomes more apparent again as the size of Ξ reduces, as already noted in Section 5.3.1.



(a) $m = 2$

(b) $m = 4$

(c) $m = 6$

Figure 5.3: Tight bounds for conditional expectation for ambiguity sets matching the moments and properties of the standard normal distribution

Even if the ambiguity sets are augmented with structural properties, the resulting dual problem can still be reduced to a semidefinite optimization problem. Consider the setting in which $\mathcal{P}_0$ contains only symmetric distributions. Using the generator class consisting of symmetric pairs of Dirac measures, the dual problem (5.9) reduces to

$$\begin{aligned} \inf \quad & \lambda_3 \\ \text{s.t.} \quad & \sum_{j=0}^{m} \lambda_j q_j \leqslant 0, \\ & \sum_{j=0}^{m+1} \lambda_j \left(h_j(\mu - x) + h_j(\mu + x)\right) \geqslant \\ & \qquad g(\mu - x)\mathbb{1}_{\Xi}(\mu - x) + g(\mu + x)\mathbb{1}_{\Xi}(\mu + x),\ \forall x \geqslant 0. \end{aligned} \tag{5.23}$$

Likewise, if we consider symmetric, unimodal distributions (which can be generated by the convex combination of a Dirac measure δ_μ and uniform distributions, which we denote by $\delta_{[\mu-z,\mu+z]}$, $z > 0$), the dual problem becomes

$$\begin{aligned} \inf \quad & \lambda_3 \\ \text{s.t.} \quad & \sum_{j=0}^{m} \lambda_j q_j \leqslant 0, \\ & \sum_{j=0}^{m+1} \lambda_j \int_{\mu-x}^{\mu+x} h_j(z)\mathrm{d}z \geqslant \int_{\mu-x}^{\mu+x} g(z)\mathbb{1}_\Xi(z)\mathrm{d}z, \ \forall x \geqslant 0, \\ & \sum_{j=0}^{m+1} \lambda_j h_j(\mu) \geqslant g(\mu)\mathbb{1}_\Xi(\mu). \end{aligned} \tag{5.24}$$

Observe that the resulting integral transforms are still piecewise polynomial in x. As a consequence, we can reformulate both (5.23) and (5.24) as semidefinite programming problems.

Figure 5.3 shows the results for different structural assumptions. Notice that the addition of structural information significantly sharpens the bounds, independently of the available moment information. However, even though the bounds are sharpened with additional information, the bounds still diverge for $t \geqslant \mu$ when only mean-variance information is available. Still, it is obvious from the figures that this conservatism of the bounds can be mitigated significantly by adding additional moment information.

5.4. Distributionally robust optimization with side information

In this section, we lay the groundwork for a generalized moment-based framework for contextual distributionally robust stochastic optimization. Section 5.4.1 introduces moment-based contextual DRO and a class of ambiguity sets that lead to tractable problems. In Section 5.4.2, we provide examples of these ambiguity sets for which computationally tractable conic optimization problems can be derived.

5.4.1. Contextual DRO with mean-dispersion information

Given the side information in the form of the event $\mathbf{X} \in \Xi$, contextual DRO problems can be stated in general as

$$\inf_{\boldsymbol{\nu}\in\mathcal{V}} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_\mathbb{P}[f(\boldsymbol{\nu}, \mathbf{X}) \,|\, \mathbf{X} \in \Xi], \tag{5.25}$$

where f is some cost function to be minimized, and $\boldsymbol{\nu} \in \mathcal{V}$ denotes the decision vector with $\mathcal{V} \subseteq \mathbb{R}^p$ a closed, convex set. The probability distribution $\mathbb{P}$ is the joint measure governing $\mathbf{X}$. Let $\mathbf{Y} \in \mathcal{Y} \subseteq \mathbb{R}^{n_\text{y}}$ be a random vector that models the outcome variables that affect the decision problem directly, and let $\mathbf{Z} \in \mathcal{Z} \subseteq \mathbb{R}^{n_\text{z}}$ be the covariates (or features) that influence the outcome random variables. Assuming that the supports of $\mathbf{Y}$ and $\mathbf{Z}$ are independent, let $\mathbf{X} = (\mathbf{Y}, \mathbf{Z}) \in \mathcal{Y} \times \mathcal{Z} =: \mathcal{X}$. Henceforth the boldface lowercase characters represent the realizations of the random vectors. Furthermore, the expectation $\inf_{\nu} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[v(\boldsymbol{\nu}, \mathbf{X}) \,|\, \mathbf{X} \in \Xi]$ is conditioned primarily on the information given by the covariates $\mathbf{Z}$ with $\Xi_{\mathbf{z}} \subseteq \mathcal{Z}$ the information set built from the information on the covariates $\mathbf{Z}$. Therefore, the side information is described by $\Xi := \{\mathbf{x} = (\mathbf{y}, \mathbf{z}) \in \mathcal{X} \,:\, \mathbf{z} \in \Xi_{\mathbf{z}}\}$. This includes the case in which $\Xi_{\mathbf{z}}$ is represented by a singleton, which models a particular realization of the covariates. No conditional information is normally included about the outcome variables. Hence, in the remainder of the section, we occasionally use the notation $\mathbb{E}[f(\boldsymbol{\nu}, \mathbf{X}) \,|\, \mathbf{Z} \in \Xi_{\mathbf{z}}]$ for the conditional expectation given the side information.

We next introduce some additional technical notation tailored to this section. Boldfaced lowercase characters represent vectors, where the italic character x_k denotes the kth element of the vector $\mathbf{x}$ and $\mathbf{x}^\top$ denotes its transpose. Except for the random vectors described above, all boldface uppercase characters represent matrices. For a set $\mathcal{S}$, let $\mathrm{conv}(\mathcal{S})$, $\mathrm{cl}(\mathcal{S})$ and $\mathrm{int}(\mathcal{S})$ denote its convex hull, closure and interior, respectively. For a proper cone $\mathcal{K} \in \mathbb{R}^n$ (i.e., a closed, convex and pointed cone with nonempty interior), the general inequality $\mathbf{x} \preccurlyeq_{\mathcal{K}} \mathbf{u}$ is equivalent to the set constraint $\mathbf{u} - \mathbf{x} \in \mathcal{K}$, while the strict variant $\mathbf{x} \prec_{\mathcal{K}} \mathbf{u}$ expresses that $\mathbf{u} - \mathbf{x} \in \mathrm{int}(\mathcal{K})$. We use $\mathcal{K}^*$ to denote the dual cone of $\mathcal{K}$, given by $\mathcal{K}^* = \{\mathbf{u} \,:\, \mathbf{u}^\top\mathbf{x},\ \forall \mathbf{x} \in \mathcal{K}\}$ with $\mathbf{u}^\top\mathbf{x}$ the appropriate inner product. The set $\mathbb{S}^n_+$ represents the cone of symmetric positive semidefinite matrices in $\mathbb{R}^{n\times n}$. Finally, for matrices $\mathbf{A}, \mathbf{B}$, we use $\mathbf{A} \preccurlyeq \mathbf{B}$ to abbreviate the relation $\mathbf{A} \preccurlyeq_{\mathbb{S}^n_+} \mathbf{B}$.

In order to derive solvable reformulations of (5.28), we shall impose the following conditions:

(C1) The side information Ξ is a closed and convex set, and the support set $\mathcal{X} = \mathbb{R}^n$.

(C2) The dispersion is modelled by ($\mathcal{D}$-)convex epigraph functions (see [101, 224]).

(C3) The function $f(\boldsymbol{\nu}, \mathbf{x})$ can be represented as

$$f(\boldsymbol{\nu}, \mathbf{x}) = \max_{l\in\mathcal{L}}\{f_l(\boldsymbol{\nu}, \mathbf{x})\},$$

in which $\mathcal{L}$ is a set of indices and the auxiliary functions f_l are of the form

$$f_l(\boldsymbol{\nu}, \mathbf{x}) = \mathbf{s}_l(\boldsymbol{\nu})^\top \mathbf{x} + t_l(\boldsymbol{\nu})$$

where for all $l \in \mathcal{L}$, $\mathbf{s}_l(\cdot) \in \mathbb{R}^p$ and $t_l(\cdot) \in \mathbb{R}$ are some affine functions of $\boldsymbol{\nu}$.

(C4) The ambiguity set $\mathcal{P}$ satisfies the Slater condition.

The rationale behind the first part of condition (C1) is twofold: (i) it enables the use of robust optimization methods to reformulate the model, and (ii) it guarantees that the side event has a conic reformulation. There is no necessity to the second part of the condition (closedness and convexity suffice), as it is merely done to benefit the mathematical exposition in this section.

The second condition states that the dispersion function has an epigraph that can be described through convex cones. We denote the dispersion function by $\mathbf{d} : \mathbb{R}^n \mapsto \mathbb{R}^m$, and assume it admits a $\mathcal{D}$-epigraph that is conic representable, with $\mathcal{D} \subseteq \mathbb{R}^m$ a proper cone, meaning that the set $\{(\mathbf{x}, \mathbf{u}) \in \mathbb{R}^n \times \mathbb{R}^m \,:\, \mathbf{d}(\mathbf{x}) \preccurlyeq_{\mathcal{D}} \mathbf{u}\}$ can be described with conic inequalities, possibly using cones other than $\mathcal{D}$ and auxiliary variables. See [23] for a comprehensive introduction to conic representations. The epigraphic mean-dispersion ambiguity set can now be defined as

$$\mathcal{P}_{(\boldsymbol{\mu},\boldsymbol{\sigma})} = \{\mathbb{P} \in \mathcal{P}_0(\mathcal{X}) \;:\; \mathbb{E}_{\mathbb{P}}[\mathbf{X}] = \boldsymbol{\mu},\, \mathbb{E}_{\mathbb{P}}[\mathbf{d}(\mathbf{X})] \preccurlyeq_{\mathcal{D}} \boldsymbol{\sigma}\}, \tag{5.26}$$

where $\mathcal{P}_0$ is the set of probability distributions with support $\mathcal{X}$, the vector $\boldsymbol{\mu} \in \mathbb{R}^n$ represents the mean value of the random vector $\mathbf{X}$, and $\boldsymbol{\sigma} \in \mathbb{R}^m$ is an upper bound on the expected value of the dispersion measure $\mathbb{E}[\mathbf{d}(\mathbf{X})]$. Although the mean-dispersion ambiguity set $\mathcal{P}_{(\boldsymbol{\mu},\mathbf{d})}$ may seem simple, there are numerous practically relevant ambiguity sets that can be recovered by selecting appropriate dispersion functions $\mathbf{d}(\cdot)$. For example, setting $\mathbf{d} = \mathbf{0}$ yields the mean-support ambiguity set, while setting $\mathbf{d}(\mathbf{X}) = (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top$ enables modeling the correlation structures between the elements of the random vector $\mathbf{X}$. Other ($\mathcal{D}$-)convex dispersion measures include the Huber loss function, mean absolute deviations, any norm $\|\cdot\|$ on $\mathbb{R}^n$, and the other convex dispersion measures that were applicable to Proposition 5.6. We will elaborate on some of these dispersion measures in the next subsection.

The third condition states that the objective function f is a convex, piecewise-affine function of the uncertainty $\mathbf{x}$, for all $\boldsymbol{\nu} \in \mathcal{V}$, and the decision vector $\boldsymbol{\nu}$, for all $\mathbf{x} \in \mathcal{X}$. We will focus on piecewise-affine objective functions as these are more than sufficient to capture several interesting models. For example, they can capture max operators as in the newsvendor model, as well as the Conditional-Value-at-Risk, which is frequently used to optimize financial portfolios with risk-averse investors. The requirement on the objective function is not strictly necessary and can be relaxed to a much richer class of functions $f(\cdot,\cdot)$ that are convex in $\boldsymbol{\nu}$ for all $\mathbf{x}$, and concave in $\mathbf{x}$ for all admissible $\boldsymbol{\nu}$. For a detailed discussion on computational tractability, we refer the interested reader to [224].

The dual problem of the inner maximization problem of (5.25), in the multivariate case, is given by

$$\begin{aligned} \inf_{\lambda_0,\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\lambda_3} \quad & \lambda_3 \\ \text{s.t.} \quad & \lambda_0 + \boldsymbol{\lambda}_1^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_2^\top \boldsymbol{\sigma} \leqslant 0, \\ & \lambda_0 + \boldsymbol{\lambda}_1^\top \mathbf{x} + \boldsymbol{\lambda}_2^\top \mathbf{d}(\mathbf{x}) \geqslant (f(\boldsymbol{\nu},\mathbf{x}) - \lambda_3)\mathbb{1}_{\Xi}(\mathbf{x}),\ \forall \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{5.27}$$

with $\lambda_0, \lambda_3 \in \mathbb{R}$, $\boldsymbol{\lambda}_1 \in \mathbb{R}^n$ and $\boldsymbol{\lambda}_2 \in \mathbb{R}^m_+$.
We assume that condition (C4) holds so that $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{\nu},\mathbf{X})|\mathbf{X} \in \Xi]$ is strongly dual to (5.27). To be more specific, the Slater condition is given by $\boldsymbol{\mu} \in \mathrm{int}(\mathcal{X})$ and $\mathbf{d}(\boldsymbol{\mu}) \prec_{\mathcal{D}} \boldsymbol{\sigma}$. The semi-infinite constraint in (5.27) can be amended using standard robust optimization methods. This yields the following result.

Theorem 5.10 **(Contextual DRO with mean-dispersion information).** *Let $\mathbb{P}$ be a member of the mean-dispersion ambiguity set $\mathcal{P}_{(\boldsymbol{\mu},\boldsymbol{\sigma})}$. If conditions* (C1)–(C4) *hold, then the objective value* *of the contextual DRO problem* (5.25) *coincides with the optimal value of the semi-infinite LP*

$$
\begin{aligned}
\inf_{\boldsymbol{\nu},\lambda_0,\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\lambda_3} \quad & \lambda_3 \\
\text{s.t.} \quad & \lambda_0 + \boldsymbol{\lambda}_1^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_2^\top \boldsymbol{\sigma} \leqslant 0, \\
& \lambda_0 + \boldsymbol{\lambda}_1^\top \mathbf{x} + \boldsymbol{\lambda}_2^\top \mathbf{u} \geqslant 0, && \forall (\mathbf{x},\mathbf{u}) \in \overline{\mathscr{C}}, \\
& \lambda_0 + \boldsymbol{\lambda}_1^\top \mathbf{x} + \boldsymbol{\lambda}_2^\top \mathbf{u} + \lambda_3 \geqslant \mathbf{s}_l(\boldsymbol{\nu})^\top \mathbf{x} + t_l(\boldsymbol{\nu}), && \forall (\mathbf{x},\mathbf{u}) \in \mathscr{C}, \forall l \in \mathscr{L}, \\
& \boldsymbol{\nu} \in \mathscr{V},\ \lambda_0 \in \mathbb{R},\ \boldsymbol{\lambda}_1 \in \mathbb{R}^n,\ \boldsymbol{\lambda}_2 \in \mathbb{R}^m_+,\ \lambda_3 \in \mathbb{R},
\end{aligned}
\tag{5.28}
$$

in which

$$
\begin{aligned}
\mathscr{C} &:= \{(\mathbf{x},\mathbf{u}) \in \mathbb{R}^n \times \mathbb{R}^m \ : \ \mathbf{x} \in \Xi,\ \mathbf{d}(\mathbf{x}) \preccurlyeq_{\mathscr{D}} \mathbf{u}\}, \\
\overline{\mathscr{C}} &:= \operatorname{conv}\left(\{(\mathbf{x},\mathbf{u}) \in \mathbb{R}^n \times \mathbb{R}^m \ : \ \mathbf{x} \in \overline{\Xi},\ \mathbf{d}(\mathbf{x}) = \mathbf{u}\}\right),
\end{aligned}
\tag{5.29}
$$

with $\overline{\Xi} := \operatorname{cl}(\mathbb{R}^n \backslash \Xi)$. *Moreover, problem* (5.28) *admits a reformulation as a finite-dimensional conic optimization problem.*

Proof. The dual problem of the inner maximization problem is given by

$$
\begin{aligned}
\inf_{\lambda_0,\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\lambda_3} \quad & \lambda_3 \\
\text{subject to} \quad & \lambda_0 + \boldsymbol{\lambda}_1^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_2^\top \boldsymbol{\sigma} \leqslant 0, \\
& \lambda_0 + \boldsymbol{\lambda}_1^\top \mathbf{x} + \boldsymbol{\lambda}_2^\top \mathbf{d}(\mathbf{x}) \geqslant \mathbb{1}_{\Xi}(\mathbf{x})\big(f(\boldsymbol{\nu},\mathbf{x}) - \lambda_3\big),\ \forall \mathbf{x} \in \mathbb{R}^n.
\end{aligned}
\tag{5.30}
$$

By decomposing the semi-infinite constraints using the definition of the indicator function, we obtain two semi-infinite constraints,

$$
\begin{aligned}
& \lambda_0 + \boldsymbol{\lambda}_1^\top \mathbf{x} + \boldsymbol{\lambda}_2^\top \mathbf{d}(\mathbf{x}) \geqslant 0, && \forall \mathbf{x} \in \mathbb{R}^n \backslash \Xi, \\
& \lambda_0 + \boldsymbol{\lambda}_1^\top \mathbf{x} + \boldsymbol{\lambda}_2^\top \mathbf{d}(\mathbf{x}) + \lambda_3 \geqslant f(\boldsymbol{\nu},\mathbf{x}), && \forall \mathbf{x} \in \Xi,
\end{aligned}
\tag{5.31}
$$

in which $\mathbb{R}^n \backslash \Xi$ is the complement of Ξ. From a standard continuity argument, it follows that we are allowed to replace the complement with its closure, $\overline{\Xi}$. Since $f(\boldsymbol{\nu},\mathbf{x})$ is a convex, piecewise affine function by condition (C3), (5.31) is equivalent to

$$
\begin{aligned}
& \lambda_0 + \boldsymbol{\lambda}_1^\top \mathbf{x} + \boldsymbol{\lambda}_2^\top \mathbf{d}(\mathbf{x}) \geqslant 0, && \forall \mathbf{x} \in \overline{\Xi}, \\
& \lambda_0 + \boldsymbol{\lambda}_1^\top \mathbf{x} + \boldsymbol{\lambda}_2^\top \mathbf{d}(\mathbf{x}) + \lambda_3 \geqslant f_l(\boldsymbol{\nu},\mathbf{x}), && \forall \mathbf{x} \in \Xi,\ \forall l \in \mathscr{L}.
\end{aligned}
$$

Then, by lifting the nonlinearity in the uncertainty to the uncertainty set, we obtain the robust counterparts

$$
\begin{aligned}
& \lambda_0 + \boldsymbol{\lambda}_1^\top \mathbf{x} + \boldsymbol{\lambda}_2^\top \mathbf{u} \geqslant 0, && \forall (\mathbf{x},\mathbf{u}) \ : \ \mathbf{x} \in \overline{\Xi},\ \mathbf{d}(\mathbf{x}) = \mathbf{u}, \\
& \lambda_0 + \boldsymbol{\lambda}_1^\top \mathbf{x} + \boldsymbol{\lambda}_2^\top \mathbf{u} + \lambda_3 \geqslant f_l(\boldsymbol{\nu},\mathbf{x}), && \forall (\mathbf{x},\mathbf{u}) \ : \ \mathbf{x} \in \Xi,\ \mathbf{d}(\mathbf{x}) = \mathbf{u},\ \forall l \in \mathscr{L}.
\end{aligned}
\tag{5.32}
$$

As the constraints are linear in the uncertain parameters, we can equivalently use the convex hull of the uncertainty sets

$$
\begin{aligned}
& \lambda_0 + \boldsymbol{\lambda}_1^\top \mathbf{x} + \boldsymbol{\lambda}_2^\top \mathbf{u} \geqslant 0, && \forall (\mathbf{x},\mathbf{u}) \in \operatorname{conv}\left(\{(\mathbf{x},\mathbf{u}) \ : \ \mathbf{x} \in \overline{\Xi},\ \mathbf{d}(\mathbf{x}) = \mathbf{u}\}\right), \\
& \lambda_0 + \boldsymbol{\lambda}_1^\top \mathbf{x} + \boldsymbol{\lambda}_2^\top \mathbf{u} + \lambda_3 \geqslant f_l(\boldsymbol{\nu},\mathbf{x}), && \forall (\mathbf{x},\mathbf{u}) \in \operatorname{conv}\big(\{(\mathbf{x},\mathbf{u}) \ : \ \mathbf{x} \in \Xi, \mathbf{d}(\mathbf{x}) = \mathbf{u}\}\big),\ \forall l \in \mathscr{L}.
\end{aligned}
\tag{5.33}
$$

Here we note, for the first set of semi-infinite constraints, that although the convex hull defines a convex set, it might not admit a conic representable form involving only the "tractable" cones (i.e., the nonnegative orthant, second-order cone, exponential cone, power cone, positive semidefinite cone and their Cartesian products). From conditions (C1) and (C2), it immediately follows that the convex hull of the uncertainty set of the second set of robust counterparts is equivalent to $\mathscr{C}$. As the robust counterparts in (5.33) constitute an (infinite) intersection of halfspaces w.r.t. the dual variables $\lambda_0, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \lambda_3$ and the decision vector $\boldsymbol{\nu}$, it follows that the feasible set is convex and (5.28) is a convex optimization problem. This a fortiori implies that problem (5.28) can be phrased as a conic optimization problem. To demonstrate the second claim, in Appendix B.3 we rewrite (5.28) as a finite-dimensional conic program. This completes the proof. □

The difficulty now lies in reformulating the semi-infinite constraints (or robust counterparts) in the dual problem by constructing explicit expressions for the convex hulls $\mathscr{C}$ and $\overline{\mathscr{C}}$. To address this, robust optimization techniques for nonlinear types of uncertainty can be used; see, for example, [16, 228] for further details. It seems noteworthy to mention here that the distribution-free analysis of (5.28) shares many similarities with the literature based on uncertainty quantification [100, 101], and distributionally robust convex optimization [224]. However, in contrast to Theorem 5 in [224], we apply a lifting argument when solving the dual problem rather than during the construction of the ambiguity set. In the next section, we show that Theorem 5.10 leads to computationally tractable models for appropriate choices of $\mathscr{P}$ and Ξ.

5.4.2. Some examples for mean-dispersion information

For the sake of exposition, we limit our attention to two types of ambiguity sets, which are analogous to Propositions 5.4 and Proposition 5.5 in Section 5.3.1. Further, for the sake of simplicity, we assume that the event set is defined by a halfspace; that is,

$$\Xi_{\mathbf{z}} = \{\mathbf{z} \in \mathbb{R}^{n_z} \, : \, \mathbf{c}^\top \mathbf{z} \leqslant \bar{c}\},$$

with $\mathbf{c} \in \mathbb{R}^{n_z}$ and $\bar{c} \in \mathbb{R}$. Since $\mathscr{Y} = \mathbb{R}^{n_y}$, Ξ is unrestricted in the outcome space. We have chosen this specific setup so that, in the remainder of this section, we can obtain computationally tractable conic reformulations. In this context, "computationally tractable" means that our problems can be formulated as linear, conic-quadratic or semidefinite programs so that we are able to use mature, off-the-shelf solvers for conic optimization. The derivations of these conic programs are provided in Appendix B.3.

We first construct a Chebyshev-type ambiguity set [66, 215], which allows us to impose conditions on the covariance matrix of the random vector $\mathbf{X}$. Let $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu}$ denote the mean vector, and define the dispersion measure as $\mathbf{d}(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top$. We identify $\mathscr{D}$ with the cone of positive semidefinite matrices. The Chebyshev ambiguity set then consists of all distributions with mean $\boldsymbol{\mu} \in \mathbb{R}^n$ and covariance matrix bounded above by $\Sigma \in \mathbb{S}^n_+$. It can be defined as

$$\mathscr{P}_{(\boldsymbol{\mu},\Sigma)} = \left\{ \mathbb{P} \in \mathscr{P}_0\left(\mathbb{R}^n\right) \, : \, \mathbb{E}_{\mathbb{P}}[\mathbf{X}] = \boldsymbol{\mu}, \; \mathbb{E}_{\mathbb{P}}\left[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top\right] \preccurlyeq \Sigma \right\}.$$

When we consider this ambiguity set in conjunction with the half-space event set, the convex hulls in (5.28) can be described by LMIs. Hence, Theorem 5.10 yields the following result.

Corollary 5.11 **(Chebyshev ambiguity set).** *Suppose conditions* (C1)–(C4) *are satisfied. Let* $\Xi_{\mathbf{z}}$ *be defined by a halfspace. Then, for* $\mathscr{P} = \mathscr{P}_{(\mu,\Sigma)}$*, the contextual DRO problem* (5.28) *can be reformulated as a semidefinite optimization problem.*

Alternatively, the MAD can be used as dispersion measure [179]. Let $\mathbf{m} \in \mathbb{R}^n$ represent some center point, in our case the mean. Assume that we have bounds for the componentwise mean deviations $\mathbb{E}[|\mathbf{X} - \mathbf{m}|]$ and the pairwise mean deviations $\mathbb{E}[|(X_i \pm X_j) - (m_i \pm m_j)|]$, which are given by $\mathbf{f} \in \mathbb{R}^{n^2}$. This information results in the ambiguity set

$$\mathscr{P}_{(\mathbf{m},\mathbf{f})} = \{\mathbb{P} \in \mathscr{P}_0(\mathbb{R}^n) : \mathbb{E}_{\mathbb{P}}[\mathbf{X}] = \mathbf{m},\ \mathbb{E}[|X_i - m_i|] \leqslant f_{i,i}, \forall i,$$
$$\mathbb{E}[|(X_i \pm X_j) - (m_i \pm m_j)|] \leqslant f_{i,j}, \forall i \neq j\}.$$

For this ambiguity set, the convex hulls in (5.28) are representable by linear inequalities. Therefore, Theorem 5.10 leads to the following result.

Corollary 5.12 **(MAD ambiguity set).** *Suppose that conditions* (C1)–(C4) *hold. Let* $\Xi_{\mathbf{z}}$ *be defined by a halfspace. Then, for* $\mathscr{P} = \mathscr{P}_{(\mathbf{m},\mathbf{f})}$*, the contextual DRO problem* (5.28) *can be reformulated as a linear optimization problem.*

The precise mathematical models are relegated to Appendix B.3. Although we have obtained computationally tractable models, it is not immediately clear how the ambiguity sets and side information interact or under which conditions the contextual DRO problem reduces to its robust counterpart, $\inf_{\nu} \sup_{\mathbf{y}} f(\nu, \mathbf{y})$, as discussed in Section 5.3. Related to this, we would like to mention the nested ambiguity sets that were introduced in [224], which encompass distance-based ambiguity sets as a special type of generalized-moment ambiguity sets. The reason for this is that distance-based ambiguity sets can be defined by a finite number of (conditional) expectation constraints based on generalized moments [53]. The distance-based ambiguity sets are particularly interesting as they provide an explicit way to relate the "sizes" of ambiguity sets and event sets. This makes it possible to quantify when the contextual DRO problem becomes "uninformative." For an excellent discussion on the interplay between a distance-based ambiguity set based on optimal transport and the size of the event set, see [169]. As our framework applies to generalized moments, it is possible to extend Theorem 5.10 to include nested ambiguity sets. Furthermore, as discussed in Section 5.3, we can obtain tighter bounds by imposing structural properties on the base ambiguity set $\mathscr{P}_0$. However, both extensions would entail delving into many technical details. As these might detract from the main focus of this expository section, we leave them to the avid reader.

5.5. Conclusions and outlook

This chapter presents a novel framework for bounding conditional expectations that, in contrast to generalized moment-bound problems, can explicitly incorporate side information into the semi-infinite formulation. The key idea is to use a simple transformation to reduce the resulting semi-infinite fractional problem to a semi-infinite LP. The corresponding dual problem highly resembles that of a generalized moment problem, but includes an additional constraint that models the conditioning on this random event. Fortunately, this slight increase in complexity does not seem to affect significantly the computational tractability of the resulting models. The generalized conditional-bound framework can be used to obtain univariate bounds for different ambiguity sets and general objectives (for, e.g., pricing) through the use of primal-dual arguments. Moreover, it serves as the foundation for a moment-based contextual DRO framework that can be applied to stochastic optimization problems with side information.

We finally mention several potentially interesting avenues for further research. First, it seems of interest to find more applications for the univariate bounds, such as the robust monopoly-pricing problem. Second, alternative applications for the contextual DRO framework, discussed in Section 5.4, can be investigated. Moreover, it would be beneficial to expand our findings to nested ambiguity sets, as this class of ambiguity includes the distance-based ambiguity sets, which offer a more direct way to answer questions about when a solution becomes "uninformative," i.e., for which instances the DRO problem reduces to a robust optimization problem. Conducting a comprehensive complexity analysis for the nested ambiguity sets and different types of side information and objective function structures also seems a worthwhile topic to explore. In conclusion, our proposed framework provides a promising approach for bounding conditional expectations while incorporating side information. We anticipate that the suggested directions for future research will contribute to the development and applicability of this framework.

6

Robust knapsack ordering for a partially-informed newsvendor

6.1. Introduction

The newsvendor model is one of the cornerstones of inventory management, introduced by Arrow et al. [6] for finding the order quantity that minimizes expected costs in view of unknown demand and the trade-off between leftovers and lost sales. The newsvendor model finds many applications in e.g. perishable food, fashion and high-tech industries, particularly when the total time span of production and lead times exceeds the market lifetime of a product; see [159] and [80].

Manufacturers and retailers need to decide how to employ the available budget or resources when determining the optimal order quantities of different products. A budget constraint makes the problem multidimensional—as ordering more of one item leaves less budget for other items—and gives rise to a challenging optimization problem. Hadley and Whitin [95] solved this problem with Lagrangian optimization. Abdel-Malek et al. [4] and Lau and Lau [140] provided alternative solution methods, Erlebacher [77] established closed-form solutions for special demand distributions and Nahmias and Schmidt [161] developed heuristic solutions. All these works are for the full information setting, where the demand distributions for all items are fully specified. In this chapter we perform a distribution-free analysis of the multi-item newsvendor

This chapter is based on the research paper [35].

problem with budget constraint. This analysis does not rely on full specification of the demand distributions, but only requires for each item knowledge of the mean, the mean absolute deviation (MAD) and range. Given this partial demand information, we obtain a robust ordering policy by employing distributionally robust optimization methods.

The newsvendor model in this chapter seeks to minimize the expected costs as function of the order quantity. The cost function depends on the order quantity, but also on the demand, which is a random variable with some distribution. Given the demand distribution, the single-item newsvendor model finds the optimal order quantity that minimizes the expected costs. In traditional approaches, the demand distribution is fully specified, so that the expected costs can be calculated, and the optimal order quantity can be determined. A *robust version* of this problem assumes partial information, and only knows that the demand distribution belongs to some ambiguity set that contains all distributions that comply with this partial information. We adopt a minimax strategy that can be viewed as a game between the newsvendor and nature: the newsvendor first picks the order quantity after which nature chooses a demand distribution that maximizes the expected costs. The goal then becomes to solve this minimax problem.

The way we solve this minimax problem in this chapter fits in a much richer class of distributionally robust optimization (DRO) approaches that first calculate worst-case model performance, over the set of distributions satisfying some partial information, and then optimize against these worst-case circumstances. Such DRO techniques found applications in many domains including scheduling [135, 149], portfolio optimization [67, 178], pricing [45, 75, 134], complex networks [213], and inventory management [15, 83, 174, 191]. A classic distributionally robust approach is due to Scarf [191], who considered the single-item newsvendor problem with mean-variance demand information. Scarf was able to derive explicit expressions for the worst-case distribution, and solved the minimax problem to obtain the optimal order quantity. Whether a minimax problem is solvable depends on both the function to be optimized and the choice of ambiguity set. There are many ways to characterize a set of distributions. In DRO, one can define ambiguity by using distance-based metrics, such as total variation or Kullback-Leibler distance. Another popular class of ambiguity uses summary statistics. The ambiguity set studied in this chapter contains all distributions with known mean and MAD. The maximization part of the minimax problem can then be viewed as a semi-infinite linear optimization problem with three constraints, and an infinite number of variables (all distributions in the ambiguity set). In fact, such minimax problems are related to generalized moment bound problems, for which general theory says there exists an extremal distribution solving the maximization part with at most a number of support points equal to the number of moment constraints [186].

For the multi-item newsvendor model in this chapter, we solve the multi-dimensional minimax problem with a random vector that describes the demand for all items. Compared with tractable one-dimensional problems such as the single-item newsvendor model, applying DRO techniques to such problems with multiple random variables might present considerable challenges in terms of computational complexity. For example, given information on the mean and covariance of the demands, the distributionally robust multi-item newsvendor is significantly harder to solve than its single-item counterpart [100]. However, for the multi-item newsvendor model in conjunction with mean-MAD ambiguity, solving the minimax problem becomes tractable, and in fact has an elegant algorithmic solution. The key insight will prove to be that the worst-case demand distribution—the solution to the maximization part of the minimax problem—is identical for any order quantity. As a result, the minimax problem reduces to a known-distribution optimization problem. This known distribution is in fact, for each item, a unique three-point distribution. In turn, the minimization problem with this known (discrete) distribution can be solved using a reduction to a knapsack problem.

We next discuss some related literature on the newsvendor model. Gallego and Moon [84] considered the multi-item newsvendor model with budget constraint when the mean and variance of demand is known. Gallego and Moon [84] extended the ideas in Scarf [191] to obtain an optimization problem that can be solved with Lagrange multiplier techniques, similar to the full information setting with a known distribution. In contrast, our minimax analysis with mean-MAD-range information yields a knapsack ordering policy that generates a sorted list and prescribes to sort items successively according to that list, with order sizes equal to the minimal, mean or maximum demand. Other related works that consider the multi-item newsvendor model under partial information include Vairaktarakis [208], who assumed only the support of demand is known, and Ardestani-Jaafari and Delage [5] who assumed knowledge of partial moments and rephrase the robust optimization problem as a tractable linear program. Natarajan et al. [163] assumed knowledge of mean, variance and semivariance, for which the newsvendor model is solvable in the single-item setting using a semi-infinite linear program, but largely intractable in the multi-item setting. Natarajan et al. [163] therefore considered a relaxation that gives a semidefinite program (SDP) to find a lower bound (which is not tight). Hanasusanto et al. [100] considered mean and covariance knowledge. They prove that the distributionally robust problem is NP-hard but admits a semidefinite programming formulation with an exponential number of inequalities (that grows in the number of items). Xu et al. [227] and Natarajan and Teo [164] presented more tractable bounds for mean-covariance information. In this chapter, we assume only marginal information is available, since covariance information and other dependency structures are difficult to estimate, and fixing covariance information often leads to difficult optimization problems with non-intuitive solutions (policies). The knapsack ordering policy that we obtain in this chapter deals with the worst-case demand distributions among all demand distributions with a given mean, MAD and range, not conditioning on a specific dependency structure. This approach makes the knapsack ordering policy robust, but also suitable for scarce-data settings, as the mean, MAD and range are relatively easy to estimate.

6.1.1. Contributions and outline

The main contributions of this chapter are as follows:

1. Solution of minimax problem. We solve the minimax problem for mean-MAD ambiguity and a budget constraint. We first show that the worst-case scenarios arise when item demands follow specific three-point distributions that comply with the partial demand information. We minimize the associated worst-case costs to obtain a robust ordering policy as the solution to a knapsack problem. As opposed to existing methods for the

newsvendor model under full demand information, the knapsack problem leads to an effective closed-form ordering policy, also for scenarios with many items. As such, the present chapter further develops DRO theory that uses MAD information to formulate tractable minimax problems.

2. Budget consistency. The robust ordering policy only depends on the minimal, mean and maximal demand for each item. Hence, the worst-case distributions are independent of all other model parameters, which makes the robust ordering policy "budget consistent." When the budget is increased, the orders for the original budget remain unaltered, while only the additional budget is further divided over the items. Such budget consistency is useful because the optimization model needs to be solved only once. That is, for the initial budget value the decision maker can generate an ordered list of items as the solution to the knapsack problem, using only standard spreadsheet software, and this solution is valid for all budget levels. In contrast, most other exact and robust methods for the multi-item newsvendor model do not have this feature, which means that the decision maker has to recompute the optimal policy for each budget level.

3. Performance of ordering policy. Through a range of numerical examples we demonstrate the performance of the knapsack ordering. We draw comparisons with full information settings and other robust approaches that require partial demand information by assessing the so-called expected value of additional information (EVAI). Overall, the performance of the robust policy only deviates a few percent from the optimal performance with full information availability. We also quantify the value of MAD information by comparing the performance with the situations when only the mean and range of demand is known, and show that MAD indeed provides crucial information for providing good performance. In addition, we construct an ordering policy that attains the optimal value of a matching minimin problem which, in conjunction with the optimal value of the minimax problem, yields tight performance guarantees.

Section 6.2 introduces the single-item model and the multi-item model with budget, under the traditional assumption of full information about the demand distributions. In Section 6.3 we present our main results for the distributionally robust setting with partial information. Section 6.4 presents a detailed numerical study that demonstrates the robust policies. We present conclusions and several directions for future work in Section 6.5.

6.2. Classical newsvendor analysis

We introduce the newsvendor model and several well-known results in Section 6.2.1 for the single-item setting, and in Section 6.2.2 for the multi-item setting with budget constraint.

6.2.1. Classical single-item setting

Consider an item with purchase price c and selling pricing p. The decision maker places an order of size q. The demand for items is assumed to be the random variable D with distribution function $F_D(\cdot)$. Unsold items will be salvaged at the end of the period for salvage value s per item. The mark-up $m > 0$ represents the profit per sold item and satisfies $p = c(1+m)$ and the discount factor $d > 0$ captures the loss through $s = (1-d)c$.

The expected costs consist of two terms: opportunity costs of lost sales and overage costs in case of overstocking. This gives the cost function

$$G(q, D) = \begin{cases} (p-c)(D-q) & \text{if } q \leqslant D, \\ (c-s)(q-D) & \text{if } q > D. \end{cases} \tag{6.1}$$

The case $q \leqslant D$ amounts to lost sales and $q > D$ results in overstocking. The objective is to order the quantity q of items that minimizes the expected costs. Let $\mathbb{E}$ denote expectation, and define $\mu = \mathbb{E}[D]$ and $x^+ = \max(x, 0)$. Write the expected costs as

$$C(q) := \mathbb{E}[G(q, D)] = (c-s)q + (p-s)\mathbb{E}(D-q)^+ - (c-s)\mu = c\left(d(q-\mu) + (m+d)\mathbb{E}(D-q)^+\right). \tag{6.2}$$

To keep notation simple (and without loss of generality) set $c = 1$. Then, the optimal order quantity

$$q^* = \operatorname*{argmin}_{q \geqslant 0} C(q) \equiv \operatorname*{argmin}_{q \geqslant 0} dq + (m+d)\mathbb{E}(D-q)^+, \tag{6.3}$$

is given by

$$q^* = \inf\left\{q \ : \ F(q) \geqslant \frac{m}{m+d}\right\}. \tag{6.4}$$

A proof of (6.4) is provided in most standard textbooks on inventory management; see e.g. [95, 160, 200].

Scarf [191] introduced a distribution-free analysis for the single-item newsvendor model by assuming that the decision maker only knows the mean and variance of the demand. Define the ambiguity set containing all distributions with the same mean and variance as

$$\mathscr{P}_{(\mu,\sigma)} := \{\mathbb{P} \,|\, \mathbb{E}_{\mathbb{P}}(D) = \mu, \ \mathbb{E}_{\mathbb{P}}(D^2) = \sigma^2 + \mu^2\}.$$

Scarf [191] determined an upper bound on the cost function $C(q)$ by finding the worst-case distribution in the ambiguity set. To find the order quantity that protects against the ambiguity in $\mathscr{P}_{(\mu,\sigma)}$, the following *minimax* optimization problem is solved:

$$\min_q \max_{\mathbb{P} \in \mathscr{P}_{(\mu,\sigma)}} dq + (m+d)\mathbb{E}_{\mathbb{P}}(D-q)^+.$$

Since

$$\max_{\mathbb{P} \in \mathscr{P}_{(\mu,\sigma)}} \mathbb{E}_{\mathbb{P}}(D-q)^+ \leqslant \frac{\sqrt{\sigma^2 + (\mu-q)^2} + (\mu-q)}{2},$$

this minimax optimization problem becomes $\min_q \max_{\mathbb{P}} C^S(q)$ with

$$C^S(q) := dq + (m+d)\frac{\sqrt{\sigma^2 + (\mu-q)^2} + (\mu-q)}{2}. \tag{6.5}$$

and solution

$$q^S := \operatorname*{argmin}_q C^S(q) = \mu + \frac{\sigma}{2}\left(\sqrt{\frac{m}{d}} - \sqrt{\frac{d}{m}}\right). \tag{6.6}$$

The quantity q^S is known as Scarf's order quantity which prescribes to order more than the expected demand when $m > d$, and less than the expected demand when $d < m$.

6.2.2. Multi-item setting

Consider n different items and order q_i units for item i for a given period where $i = 1, \dots, n$. For item i, the unit purchasing and selling price are c_i and p_i respectively. Possible leftovers will be salvaged at the end of the period for unit salvage value s_i. We define the model in terms of the mark-up $m_i > 0$ and discount factor $d_i > 0$. The mark-up represents the profit per sold unit and the discount factor the loss, i.e. $p_i = c_i(1 + m_i)$ and $s_i = (1 - d_i)c_i$. The random demand for item i in one period is represented by the nonnegative random variable D_i, distributed according to $F_i(\cdot)$. As in the single-item setting, we minimize the expected costs. Define the multi-item cost function as

$$G(\mathbf{q}, \mathbf{D}) := \sum_{i=1}^{n} c_i \left(d_i(q_i - D_i) + (m_i + d_i)(D_i - q_i)^+\right). \tag{6.7}$$

We also introduce the budget constraint $\sum_{i=1}^{n} c_i q_i \leqslant B$ with B the available budget. The multi-item newsvendor model, with decision vector $\mathbf{q} = (q_1, \dots, q_n)$, is then given by

$$\begin{aligned} \min_{\mathbf{q}} \quad & C(\mathbf{q}) := \mathbb{E}[G(\mathbf{q}, \mathbf{D})] = \sum_{i=1}^{n} c_i \left(d_i(q_i - \mu_i) + (m_i + d_i)\mathbb{E}(D_i - q_i)^+\right) \\ \text{s.t.} \quad & \sum_{i=1}^{n} c_i q_i \leqslant B, \\ & q_i \geqslant 0, \quad i = 1, \dots, n. \end{aligned} \tag{6.8}$$

Its solution, referred to as the optimal ordering policy, will be denoted by $\mathbf{q}^*$. In the single-item setting the purchase costs had no influence on the objective function, but in the multi-item setting the optimal order quantity is affected by c_i. It is well known that model (6.3) is a convex optimization problem. In (6.8) we take the summation over n convex functions, which preserves convexity. Moreover, the constraints form a convex set, so that (6.8) is a convex optimization problem [36]. Gallego and Moon [84] built on Scarf's result to consider the multi-item setting, which gives the problem

$$\begin{aligned} \min_{\mathbf{q}} C^S(\mathbf{q}) &:= \sum_{i=1}^{n} c_i \left(d_i(q_i - \mu_i) + (m_i + d_i)\, \frac{\sqrt{\sigma_i^2 + (q_i - \mu_i)^2} - (q_i - \mu_i)}{2} \right) \\ \text{s.t.} \quad & \sum_{i=1}^{n} c_i q_i \leqslant B, \\ & q \geqslant 0. \end{aligned} \tag{6.9}$$

The optimal solution to problem (6.9) is referred to as $\mathbf{q}^S$. Applying Scarf's bound for each item individually results in (6.9). Similar to the full information setting with a known distribution, this optimization problem can be solved with Lagrange multiplier techniques.

6.3. Proposed robust approach

Section 6.3.1 presents the robust ordering policy for the single-item setting. This result serves as building block for the robust analysis of the multi-item setting in Section 6.3.2, which describes the optimal policy as the solution of an LP. In Section 6.3.3 we show that this LP can be viewed as a knapsack problem. All these results are based on a tight upper bound for the cost function. In Section 6.3.4 we derive a matching tight lower bound for the cost function.

6.3.1. Distribution-free ordering policy for single item

Let $\mathbb{P}$ denote a probability distribution, and write $\mathbb{E}_{\mathbb{P}}$ for $\mathbb{E}$ to emphasize that the expectation is taken with respect to the distribution $\mathbb{P}$ of D. The MAD for random demand D is defined as $\delta := \mathbb{E}_{\mathbb{P}}|D-\mu|$, where μ is the expected value of D. Similar to the variance, the MAD is a measure of dispersion or variability. We mention several properties of MAD in Appendix A. For the random variable D with mean μ, MAD δ, and (bounded) support $[a, b]$, where $0 \leqslant a \leqslant b < \infty$, the mean-MAD ambiguity set is defined as

$$\mathscr{P}_{(\mu,\delta)} := \{\mathbb{P} \,|\, \mathbb{E}_{\mathbb{P}}[D] = \mu,\ \mathbb{E}_{\mathbb{P}}|D-\mu| = \delta,\ \mathrm{supp}(D) \subseteq [a, b]\}.$$

We thus assume that the "true" distribution $\tilde{\mathbb{P}}$ of the random demand D is contained in this ambiguity set, that is, $\tilde{\mathbb{P}} \in \mathscr{P}_{(\mu,\delta)}$.

To obtain the robust order quantity, we solve

$$\min_q \max_{\mathbb{P}\in\mathscr{P}_{(\mu,\delta)}} dq + (m+d)\mathbb{E}_{\mathbb{P}}(D-q)^+,$$

for which we first consider $\max_{\mathbb{P}\in\mathscr{P}_{(\mu,\delta)}} \mathbb{E}_{\mathbb{P}}(D-q)^+$. To characterize this tight bound, we apply a general upper bound for convex functions of a random variable by Ben-Tal and Hochman [19]. The proof of the following result has been provided in Chapter 1.

Lemma 6.1. *The extremal distribution that solves* $\max_{\mathbb{P}\in\mathscr{P}_{(\mu,\delta)}} \mathbb{E}_{\mathbb{P}}(D-q)^+$ *is a three-point distribution on the values* a, μ *and* b *that does not depend on* q.

From the proof of Lemma 6.1, it follows that the worst-case probability distribution of D, the extremal distribution that solves $\max_{\mathbb{P}\in\mathscr{P}_{(\mu,\delta)}} \mathbb{E}_{\mathbb{P}}(D-q)^+$, is a three-point distribution defined as

$$\mathbb{P}(D=x) = \begin{cases} \dfrac{\delta}{2(\mu-a)}, & \text{for } x = a, \\ 1 - \dfrac{\delta}{2(\mu-a)} - \dfrac{\delta}{2(b-\mu)}, & \text{for } x = \mu, \\ \dfrac{\delta}{2(b-\mu)}, & \text{for } x = b. \end{cases} \tag{6.10}$$

Applying this worst-case distribution, the robust order quantity follows from solving $q^U = \text{argmin}_q C^U(q)$ with

$$\begin{aligned} C^U(q) &:= d(q-\mu) + \frac{\delta(m+d)}{2(\mu-a)}(a-q)^+ + (m+d)\left(1 - \frac{\delta}{2(\mu-a)} - \frac{\delta}{2(b-\mu)}\right)(\mu-q)^+ \\ &+ \frac{\delta(m+d)}{2(b-\mu)}(b-q)^+. \end{aligned} \tag{6.11}$$

The upper bound (6.11) coincides with the "true" cost function at the points a, μ and b. Clearly, for $q = a$ or b, it holds that $C^U(q) = C(q)$. When $q = \mu$, the cost function equals

$$C(\mu) = d(\mu-\mu) + (m+d)\mathbb{E}(D-\mu)^+ = \frac{\delta(m+d)}{2} = C^U(\mu),$$

since $\mathbb{E}(D-\mu)^+ = \mathbb{E}|D-\mu|/2$. By analyzing (6.11) one can obtain an explicit ordering rule for q^U. The objective function of (6.11) is composed of piecewise linear functions. By exploiting this structure, we can construct an explicit ordering policy. For scalars $\alpha_1, \dots, \alpha_m, \nu_1, \dots, \nu_m \in \mathbb{R}$, $f(x) = \max_{i=1,\dots,m}\{\alpha_i x + \nu_i\}$ denotes a convex, piecewise linear function. The function $C^U(q)$ in (6.11) admits a representation of the form

$$C^U(q) = d(q-\mu) + (m+d)\mathbb{E}(D-q) = m(\mu-q) =: f_0(q),$$

for $q \in [0, a)$ and

$$\begin{aligned} C^U(q) &= d(q-\mu) + (m+d)\left(1 - \frac{\delta}{2(\mu-a)} - \frac{\delta}{2(b-\mu)}\right)(\mu-q) + \frac{\delta(m+d)}{2(b-\mu)}(b-q) \\ &= q(\frac{\delta(m+d)}{2(\mu-a)} - m) + \nu_1 =: f_1(q), \end{aligned}$$

for $q \in [a, \mu)$, where ν_1 is some constant value. For $q \in [a, \mu)$, the mean-MAD objective function is defined by the linear function $f_1(q)$. For the interval $q \in [\mu, b]$, we obtain

$$C^U(q) = d(q-\mu) + \frac{\delta(m+d)}{2(b-\mu)}(b-q) = q\left(d - \frac{\delta(m+d)}{2(b-\mu)}\right) + \nu_2 =: f_2(q)$$

for some constant ν_2. The cost function is thus the pointwise maximum of the three linear functions $f_0(q)$, $f_1(q)$ and $f_2(q)$:

$$C^U(q) = \max\{f_0(q),\ f_1(q),\ f_2(q)\}.$$

Since $C^U(q) = \max_{j=0,1,2}\{\alpha_j q + \nu_j\}$ is a convex function, it holds that $\alpha_0 \leqslant \alpha_1 \leqslant \alpha_2$. Since we assume that $m > 0$, we know that $\alpha_0 < 0$. Therefore, from the derivatives α_1, α_2 of $C^U(q)$, we can derive an explicit order quantity by examining for which linear piece the slope turns positive. This allows us to state Theorem 6.2.

Theorem 6.2 **(Mean-MAD order quantity).** *The robust order quantity* $q^U \in \text{argmin}_q C^U(q)$ *is given by*

(a) *If* $m < \dfrac{\delta d}{2(\mu - a) - \delta}$, *then* $q^U = a$.

(b) *If* $\dfrac{\delta d}{2(\mu - a) - \delta} < m < \dfrac{d(2(b - \mu) - \delta)}{\delta}$, *then* $q^U = \mu$.

(c) *If* $\dfrac{d(2(b - \mu) - \delta)}{\delta} < m$, *then* $q^U = b$.

(d) *If* $m = \dfrac{\delta d}{2(\mu - a) - \delta}$ *and* $m = \dfrac{d(2(b - \mu) - \delta)}{\delta}$, *then* $q^U \in [a, \mu]$ *and* $q^U \in [\mu, b]$, *respectively.*

According to Theorem 6.2, the robust order quantity q^U for mean-MAD-range information consists of three predictable values (minimal, mean, maximum demand) that do not depend on the mark-up m and discount factor d, whereas the conditions that dictate how much to order do depend on them (in addition to the demand mean, MAD and range). Furthermore, based on the preceding observations, we can derive valuable performance guarantees. The costs associated with the optimal order quantity are always bounded from above by $\min\{C^U(a), C^U(\mu), C^U(b)\}$, but more importantly, the performance of the robust order quantity q^U (i.e., the mean value μ or either of the support values a, b) is actually equivalent to the true costs. This property sets it apart from Scarf's robust order quantity, which lacks such a guarantee.

6.3.2. Multiple items and budget constraint

A distribution-free analysis of the multi-item model requires a multivariate ambiguity set. As in the single-item case, the partial information is the mean μ_i, MAD δ_i and support $\text{supp}(D_i) = [a_i, b_i]$ for each random variable D_i, $i = 1, \dots, n$. The mean-MAD ambiguity set is defined as

$$\mathscr{P}_{(\mu,\delta)} := \{\mathbb{P} \mid \mathbb{E}_{\mathbb{P}}(D_i) = \mu_i,\ \mathbb{E}_{\mathbb{P}} |D_i - \mu_i| = \delta_i,\ \text{supp}(D_i) \subseteq [a_i, b_i],\ \forall i\}. \tag{6.12}$$

We henceforth assume that the distribution of the vector of random variables $\mathbf{D} = (D_1, \dots, D_n)$ belongs to this ambiguity set, i.e., $\mathbb{P} \in \mathscr{P}_{(\mu,\delta)}$. Since the objective function in (6.8) is separable, one can apply the single-item bound to each term $\mathbb{E}(D_i - q_i)^+$ in the summation individually. The following result, for the multi-item problem, is then a direct consequence of Lemma 6.1.

Lemma 6.3. *The extremal distribution that solves* $\max_{\mathbb{P} \in \mathscr{P}_{(\mu,\delta)}} \mathbb{E}_{\mathbb{P}}[G(\mathbf{q}, \mathbf{D})]$ *consists for each* D_i *of a three-point distribution with values* $\xi_1^{(i)} = a_i$, $\xi_2^{(i)} = \mu_i$, $\xi_3^{(i)} = b_i$ *and probabilities*

$$p_1^{(i)} = \frac{\delta_i}{2(\mu_i - a_i)}, \quad p_2^{(i)} = 1 - \frac{\delta_i}{2(\mu_i - a_i)} - \frac{\delta_i}{2(b_i - \mu_i)}, \quad p_3^{(i)} = \frac{\delta_i}{2(b_i - \mu_i)}. \tag{6.13}$$

For the multi-item newsvendor model based on mean-MAD ambiguity, we use Lemma 6.3 to solve the maximization part of

$$\min_{\mathbf{q}\,:\, \sum_i c_i q_i \leqslant B, q_i \geqslant 0} \ \max_{\mathbb{P} \in \mathscr{P}_{(\mu,\delta)}} \mathbb{E}_{\mathbb{P}}\Big[\sum_{i=1}^{n} c_i d_i (q_i - \mu_i) + c_i (m_i + d_i)(D_i - q_i)^+\Big], \tag{6.14}$$

and obtain

$$\begin{aligned} \min_{\mathbf{q}} \quad & \sum_{i=1}^{n} c_i \left(d_i(q_i - \mu_i) + (m_i + d_i) \left(p_1^{(i)} (a_i - q_i)^+ + p_2^{(i)} (\mu_i - q_i)^+ + p_3^{(i)} (b_i - q_i)^+ \right) \right) \\ \text{s.t.} \quad & \sum_{i=1}^{n} c_i q_i \leqslant B, \\ & q_i \geqslant 0, \quad i = 1, \dots, n. \end{aligned} \tag{6.15}$$

The objective function of (6.15) has a piecewise linear structure. Moreover, because of this result and since the constraints are linear, (6.15) can be cast as an LP. In particular, as explained below, the robust ordering policy $\mathbf{q}^U$ can be found by solving

$$\begin{aligned} \min_{\mathbf{q}} \quad & \sum_{i=1}^{n} \max_{j=0,1,2} \{\alpha_{i,j} q_i + \nu_{i,j}\} \\ \text{s.t.} \quad & \sum_{i=1}^{n} c_i q_i \leqslant B, \\ & q_i \geqslant 0, \quad i = 1, \dots, n, \end{aligned} \tag{6.16}$$

where

$$\alpha_{i,0} = -c_i m_i, \qquad \nu_{i,0} = c_i m_i \mu_i,$$
$$\alpha_{i,1} = c_i \left(\frac{\delta_i (m_i + d_i)}{2(\mu_i - a_i)} - m_i \right), \quad \nu_{i,1} = c_i (m_i + d_i) \left(\mu_i - \frac{\delta_i a_i}{2(\mu_i - a_i)} \right) - c_i d_i \mu_i,$$
$$\alpha_{i,2} = c_i \left(d_i - \frac{\delta_i (m_i + d_i)}{2(b_i - \mu_i)} \right), \quad \nu_{i,2} = \frac{c_i \delta_i (m_i + d_i) b_i}{2(b_i - \mu_i)} - c_i d_i \mu_i, \quad \text{for } i = 1, \dots, n.$$

Let $f_{i,j}(x) = \alpha_{i,j} x + \nu_{i,j}$ for $i = 1, \dots, n$ and $j = 0, 1, 2$. From the single-item case, we know that the objective, for each item i, can be written as $\max_{j=0,1,2}\{f_{i,j}(q_i)\}$ with $\alpha_{i,0} \leqslant \alpha_{i,1} \leqslant \alpha_{i,2}$, and thus the objective functions of (6.15) and (6.16) are equal, which makes the two models equivalent. Since we know from linear programming theory that convex, piecewise linear objective functions can be written as linear constraints, problem (6.16) admits an LP representation [36].

6.3.3. Knapsack algorithm

It turns out that problem (6.16) is intimately related to the continuous knapsack problem, thus making available efficient sorting-based algorithms to solve (6.16). We next describe an efficient algorithm that determines the robust ordering policy.

Define the linear funtion $f_{i,j}$ for each item i, and let $\alpha_{i,j}$ represent its derivative with respect to q_i, for items $i = 1, \dots, n$ and linear pieces $j = 0, 1, 2$. That is,

$$\frac{\mathrm{d} f_{i,j}(q_i)}{\mathrm{d} q_i} = \alpha_{i,j}.$$

For each item i, $f_{i,0}$, $f_{i,1}$ and $f_{i,2}$ represent the marginal effect on the value of (6.16) when we increase q_i to a_i, μ_i and b_i respectively. The parameter $\alpha_{i,j}$ represents the slope of these linear functions and an order quantity is increased only when $\alpha_{i,j} < 0$, because otherwise it will not reduce the expected costs. We consecutively allocate budget to the item that causes the largest relative decrease in expected costs; that is, item k with the smallest negative derivative $\alpha_{k,i}$ relative to its cost c_k. Define the set of all items as $N = \{1, \dots, n\}$. Since only order quantities that decrease the expected costs are considered, define the ordered set:

$$\mathcal{G} := \{(i, j) \,|\, \alpha_{i,j} < 0, i \in N, j \in \{0, 1, 2\}\}, \tag{6.17}$$

where the ordering is determined according to the value of $\alpha_{i,j}/c_i$. For $m = |\mathcal{G}|$, this ordering is represented by the sequence $(i_1, j_1), \dots, (i_m, j_m)$ for which it holds that $\alpha_{i_1,j_1}/c_{i_1} \leqslant \cdots \leqslant \alpha_{i_m,j_m}/c_{i_m}$. Here $\mathcal{G}$ contains tuples (i, j) for which i represents an item in the newsvendor model and j a linear piece of the piecewise function. As these functions are convex, the linear pieces appear for each item i in increasing order in the set $\mathcal{G}$. We can now state the knapsack algorithm for the distribution-free multi-item newsvendor model. This algorithm yields an optimal solution to (6.16), as asserted in the following theorem.

Theorem 6.4 **(Knapsack ordering policy).** *For given budget level $B \geqslant 0$, the robust ordering policy $\mathbf{q}^U$ that solves the multi-item newsvendor model* (6.16) *is determined by the following procedure:*

(i) *Initialize by setting $\mathbf{q} = (0, \dots, 0)$, and construct $\mathcal{G}$. Continue to* (ii).

(ii) *Select the first element $(i, j) \in \mathcal{G}$. If the set $\mathcal{G}$ is empty, the optimal solution is $\mathbf{q}^U = \mathbf{q}$. Otherwise, continue to* (iii).

(iii) *If $j = 0$, set $q_i = a_i$. If $j = 1$, set $q_i = \mu_i$. If $j = 2$, set $q_i = b_i$. Continue to* (iv).

(iv) *Determine whether the budget constraint $\sum_{i=1}^n c_i q_i \leqslant B$ is violated. If so, set q_i such that $c_i q_i = B - \sum_{k \in N | k \neq i} c_k q_k$, and the optimal solution is $\mathbf{q}^U = \mathbf{q}$. Otherwise, remove element (i, j) from $\mathcal{G}$ and return to step* (ii).

Proof. To prove that this algorithm produces an optimal solution, we construct a continuous knapsack problem that solves (6.16). In the following, (i_k, j_k) corresponds to the kth entry of the ordered sequence of items in $\mathcal{G}$, with $|\mathcal{G}| = m$. Define the following auxiliary model:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \sum_{k=1}^{m} p_k x_k \\ \text{s.t.} \quad & \sum_{k=1}^{m} c_k x_k \leqslant B, \\ & 0 \leqslant x_k \leqslant u_k \quad \forall k = 1, \dots, m, \end{aligned} \tag{6.18}$$

where

$$u_k = \begin{cases} a_{i_k}, & \text{for } j_k = 0 \\ \mu_{i_k} - a_{i_k}, & \text{for } j_k = 1 \\ b_{i_k} - \mu_{i_k}, & \text{for } j_k = 2 \end{cases}$$

and $p_k = \alpha_{i_k,j_k}$ and $c_k = c_{i_k}$. From the order of the sequence, it follows that $p_1/c_1 \leqslant \ldots \leqslant p_m/c_m$. Assume that $(x_1^*, \ldots, x_m^*)$ is an optimal solution to optimization problem (6.18). For $i \in N$, let $q_i^U = \sum_{k=1,\ldots,m|i=i_k} x_k^*$. Since $\alpha_{i,0} \leqslant \alpha_{i,1} \leqslant \alpha_{i,2}$, the pieces j_k appear in $\mathcal{G}$ in increasing order for each item i. Thus, in an optimal solution, u_{i_k,j_k} will only be attained if its predecessor u_{i_k,j_l} is also attained. By construction, $\mathbf{q}^U$ is feasible for (6.16). Moreover, the objective values of problems (6.16) and (6.18) only differ by a constant term, so both problems have the same optimal solution. For the continuous knapsack problem, a greedy allocation produces an optimal solution (see, e.g., [129]). Hence, $\mathbf{q}^U = (q_1^U, \ldots, q_n^U)$ is optimal for (6.16). □

Theorem 6.4 shows that there exists a ranking for the selection of items. Take an initial budget $B = 0$. If we increase the budget B by some small value, we first increase item i to a_i for the item that has the highest mark-up m_i. This makes sense intuitively because the product with the highest mark-up is most profitable and, since $q_i < a_i$, we have no risk of overstocking. We successively select the items with the greatest marginal benefit $\alpha_{i,j}/c_i$, and increase the order quantity consecutively to either a_i, μ_i or b_i. This procedure continues until we have spent the entire budget, or reached the uncapacitated optimum. Items that are ordered in the beginning of this procedure have the largest impact on the decrease in costs for the multi-item newsvendor model.

As the main complexity of the knapsack algorithm in Theorem 6.4 stems from sorting the set $\mathcal{G}$, the greedy approach is of computational complexity $O(n \log n)$. Moreover, the solution can be found in $O(n)$ time by first identifying the critical element (i_s, j_s) that will violate the budget constraint, as proposed in [11] for the continuous knapsack problem. One then compares each $\alpha_{i,j}/c_i$ with the ratio of the critical element to determine the optimal allocation of budget to the items. The optimal solution can also be found through the LP (6.16), which we solve with the simplex method. We remark that a single iteration of the simplex method takes $O(n^2)$ arithmetic operations [116], which exceeds the time requirement of the knapsack algorithm.

6.3.4. A matching lower bound

The robust analysis so far was based on finding a tight upper bound on the cost function when we know the mean, MAD and range of the demand distributions. When additional information is available, we can also construct a matching lower bound. We include the skewness information $\beta_i = \mathbb{P}(D_i \geqslant \mu_i)$ in the mean-MAD ambiguity set to obtain the tight lower bound. For the random variables $\mathbf{D} = (D_1, \ldots, D_n)$, define the ambiguity set as

$$\mathcal{P}_{(\mu,\delta,\beta)} := \{\mathbb{P} \,|\, \mathbb{P} \in \mathcal{P}_{(\mu,\delta)},\ \mathbb{P}(D_i \geqslant \mu_i) = \beta_i,\ i = 1, \ldots, n\}$$

with $\mathcal{P}_{(\mu,\delta,\beta)} \subseteq \mathcal{P}_{(\mu,\delta)}$. The proof of the following result is identical to that of Lemma 6.3, but now uses the tight lower bound for a convex function of random variables discussed in Ben-Tal and Hochman [19]. The proof for the univariate case has been provided in Chapter 1. This is sufficient since the univariate result can be applied to each term of the summation in $G(\mathbf{q}, \mathbf{D})$ separately, as with Lemma 6.3.

Lemma 6.5. *The extremal distribution that solves* $\min_{\mathbb{P}\in\mathcal{P}_{(\mu,\delta,\beta)}} \mathbb{E}_{\mathbb{P}}[G(\mathbf{q},\mathbf{D})]$ *consists for each* D_i *of a two-point distribution with values* $\mu_i + \frac{\delta_i}{2\beta_i}$, $\mu_i - \frac{\delta_i}{2(1-\beta_i)}$ *and probabilities* β_i, $1-\beta_i$, *respectively.*

Using this result, we obtain

$$\begin{aligned}
\min_{\mathbf{q}} \sum_{i=1}^{n} c_i & \left(d_i(q_i-\mu_i) + (m_i+d_i)\left(\beta_i(\mu_i + \frac{\delta_i}{2\beta_i} - q_i)^+ + (1-\beta_i)(\mu_i - \frac{\delta_i}{2(1-\beta_i)} - q_i)^+ \right)\right) \\
\text{s.t} \quad & \sum_{i=1}^{n} c_i q_i \leqslant B, \\
& q_i \geqslant 0, \quad \text{for } i = 1,\dots,n,
\end{aligned} \tag{6.19}$$

as a model to provide a lower bound for the multi-item newsvendor. Let $C^L(\mathbf{q})$ denote the objective function of (6.19). Since $C^L(\mathbf{q})$ also consists of piecewise linear functions, there exists an LP representation and knapsack algorithm for (6.19) analogous to the results for problem (6.15).

We can now solve (6.16) and (6.19) to obtain tight performance intervals for the multi-item newsvendor model, using recent DRO results (see Appendix A and [179]). For all feasible ordering policies $\mathbf{q}$ and $\mathbb{P} \in \mathcal{P}_{(\mu,\delta,\beta)}$, it holds that

$$C(\mathbf{q}) \in \left[C^L(\mathbf{q}), C^U(\mathbf{q})\right].$$

In addition, for the optimal solutions to the newsvendor problem and its distributionally robust counterparts,

$$C(\mathbf{q}^*) \in \left[C^L(\mathbf{q}^L), C^U(\mathbf{q}^U)\right].$$

One can find the tightest upper and lower bounds, based on mean-MAD ambiguity, for the multi-item newsvendor model by calculating the optimal solutions to models (6.15) and (6.19), respectively.

6.4. Numerical examples of robust ordering

We will now illustrate and visualize the robust ordering policies. To demonstrate the "budget-consistency" property, Section 6.4.1 applies the knapsack algorithm for a setting where the budget is increased. In Section 6.4.2 we contrast the performance of the knapsack policy for partial demand information against that of the optimal solution for the full information setting.

6.4.1. Numerical illustration of the "budget-consistency" property

We illustrate the knapsack algorithm and the process of allocating budget to different order quantities for items in the newsvendor model. Consider $n = 5$ identically distributed items with support $a = 10$, $b = 50$ and mean $\mu = 30$. From Figure 6.1, we can infer that item 1 is the most profitable. Low budget levels are allocated to this item such that we obtain $q_1 = \mu$. Item number 3 is the last item to which the budget is allocated. Hence, it is the least profitable item. Table 6.1 displays the ordered set $\mathcal{G}$.

Table 6.1: Table containing $\alpha_{i,j}/c_i$ and corresponding information of the ordered set $\mathcal{G}$

$\mathcal{G}$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\alpha_{i,j}/c_i$	-0.92	-0.75	-0.72	-0.49	-0.3	-0.15	-0.1	-0.08	-0.03	-0.01	0.14	0.42	0.45	0.7	0.7
Function piece	0	1	0	1	0	0	1	2	1	0	1	2	2	2	2
Item	1	1	2	2	4	5	4	1	5	3	3	5	2	4	3

From this table, we can indeed infer that item 1 has the smallest value for $\alpha_{i,0}/c_i$ and therefore is increased first.

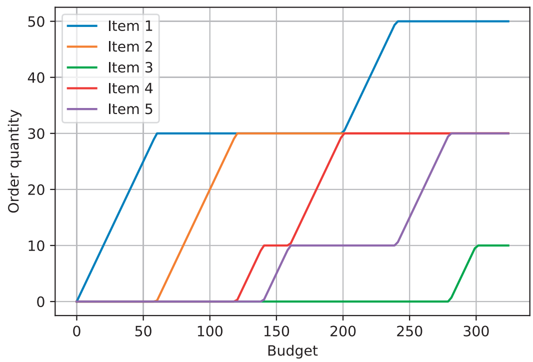
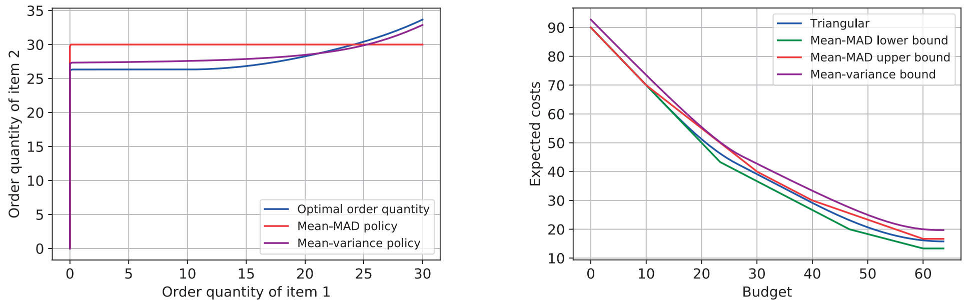


Figure 6.1: Development of the order quantities when the budget increases according to the knapsack algorithm

Figure 6.1 nicely illustrates that when the budget is increased, the orders for the original budget remain unaltered, while only the additional budget is further divided over the items. To further illustrate the "budget-consistency" property, consider the multi-item newsvendor model for which $n = 2$, $m_2 = 2$, the remaining cost parameters equal 1, and demand is identically distributed according to a symmetric triangle distribution supported on $[10, 50]$. In Figure 6.2 we plot the expected costs and order quantities for various budget levels. Figure 6.2a contains the allocation between both order quantities. For low budget values, one first increases the order quantity of item one, the most profitable item. Figure 6.2b shows the upper bound (6.15) and lower bound (6.19) that together lead to a tight performance interval for the expected costs. For the sake of comparison, we also show results for the partial demand information setting considered in Gallego and Moon [84], assuming that the mean and variance of demands are known. The results of Gallego and Moon [84] depend (non-trivially) on all model parameters, including the budget B. This lack of budget-consistency forces the decision maker to solve an optimization problem, i.e. (6.9), for each budget level separately, which explains the smooth curve in Figure 6.2a. Solving (6.9) yields the mean-variance alternative for the knapsack algorithm. In contrast, our knapsack algorithm generates a sorted ordering list that does not depend on B, and prescribes sorting items successively according to that list, with order sizes equal to the minimal, mean or maximum demand.



(a) Ordering policy **(b)** Newsvendor costs

Figure 6.2: Mean-variance and mean-MAD bounds and ordering policies for the newsvendor model.

We emphasize that these results are not meant to numerically compare the mean-MAD and mean-variance policies, because the displayed differences merely express different ways of dealing with ambiguity. Indeed, it is hard to compare both policies as the respective ambiguity sets can contain vastly different distributions. For instance, a finite variance excludes distributions with an infinite second moment, while finite MAD does not. For our purposes, MAD and variance are equally adequate descriptors of dispersion, and both are easily calibrated on data using basic statistical estimators. The crucial difference in the DRO context of this chapter is that MAD leads to a simple, budget-consistent ordering policy.

6.4.2. Expected value of additional information

We introduce as performance measure the expected value of additional information (EVAI), defined as

$$\mathrm{EVAI}(\mathbf{q}_B^U) = \frac{C(\mathbf{q}_B^U) - C(\mathbf{q}_B^*)}{C(\mathbf{q}_B^*)},$$

where $\mathbf{q}_B^U$ is the robust ordering policy and $\mathbf{q}_B^*$ is the optimal ordering policy when the joint demand distribution is known. We let B run from 0 to $\sum_{i=1}^n q_i^* =: B_{\text{opt}}$, and consider nine different demand distributions, listed in Table 6.2.

Table 6.2: Nine distributions used for multi-item performance analysis

Case		Case		Case	
1	Uniform$[10, 50]$	4	Beta$(1, 3)$ on $[0, 50]$	7	Triangular$(10, 50, 18)$
2	Uniform$[10, 100]$	5	Beta$(2, 2)$ on $[0, 50]$	8	Triangular$(10, 50, 30)$
3	Uniform$[10, 200]$	6	Beta$(3, 1)$ on $[0, 50]$	9	Triangular$(10, 50, 42)$

We consider $n = 25$ items. For each item i, let $c_i = d_i = 1$ and assume identically distributed demand. For example, in Case 2 the demand D_i for each item i follows the uniform distribution with parameters $a_i = 10$ and $b_i = 100$. Table 6.3 provides an overview for the mark-up, representing low, average and high margins.

Table 6.3: Mark-up values for all 25 items in the newsvendor model

Mark-up	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}
Low margin	0.1	0.14	0.18	0.21	0.25	0.29	0.33	0.36	0.4	0.44	0.48	0.51	0.55
Average margin	1	1.13	1.25	1.38	1.5	1.63	1.75	1.88	2	2.13	2.25	2.38	2.5
High margin	4	4.21	4.42	4.63	4.83	5.04	5.25	5.46	5.67	5.88	6.08	6.29	6.5
Mark-up	m_{14}	m_{15}	m_{16}	m_{17}	m_{18}	m_{19}	m_{20}	m_{21}	m_{22}	m_{23}	m_{24}	m_{25}	
Low margin	0.59	0.63	0.66	0.7	0.74	0.78	0.81	0.85	0.89	0.93	0.96	1	
Average margin	2.63	2.75	2.88	3	3.13	3.25	3.38	3.5	3.63	3.75	3.88	4	
High margin	6.71	6.92	7.12	7.33	7.54	7.75	7.96	8.17	8.37	8.58	8.79	9	

For the low margin regime, Figure 6.3 shows results for each of the nine cases, for both the robust ordering policy with mean-MAD-range information, and for the policy that uses the additional information $\beta_i = \mathbb{P}(D_i \geqslant \mu_i)$. For the former, the worst performance over all nine cases has a maximum deviation of approximately 23% compared to the optimal order quantity q_B^*. Overall, the performance of the robust policy only deviates a few percent from the optimal performance with full information availability. For the uniformly distributed cases (Cases 1-3), the performance decreases when the range increases. For beta distributed demand (Cases 4-6), right-tailed distributions perform worse than left-tailed distributions. This effect is also observed for the triangular distributions (Cases 7-9). The policy with additional information $\beta_i = \mathbb{P}(D_i \geqslant \mu_i)$ performs somewhat better in most cases.

Figure 6.4 depicts the results for the average profitability scenario. A quick glance reveals that these plots exhibit a different impression than the low profitability scenario. The performance of the mean-MAD policy stays below some threshold for most budget levels, but as the budget exceeds two-thirds of the maximum budget B_{opt}, the performance starts to decrease. By contrast, the mean-MAD-β EVAI actually starts decreasing when approaching the maximal budget.

Figure 6.5 shows similar results for high margins. The EVAI for the robust policy remains mostly below 10% for lower budget levels, but starts increasing rapidly when the budget approaches B_{opt} (i.e., when approaching the unconstrained model). When the budget is less restrictive, additional distributional information provides substantial value. In particular, since the policy uses skewness information β_i, it performs better (in expectation) for higher budget levels than the robust ordering policy.

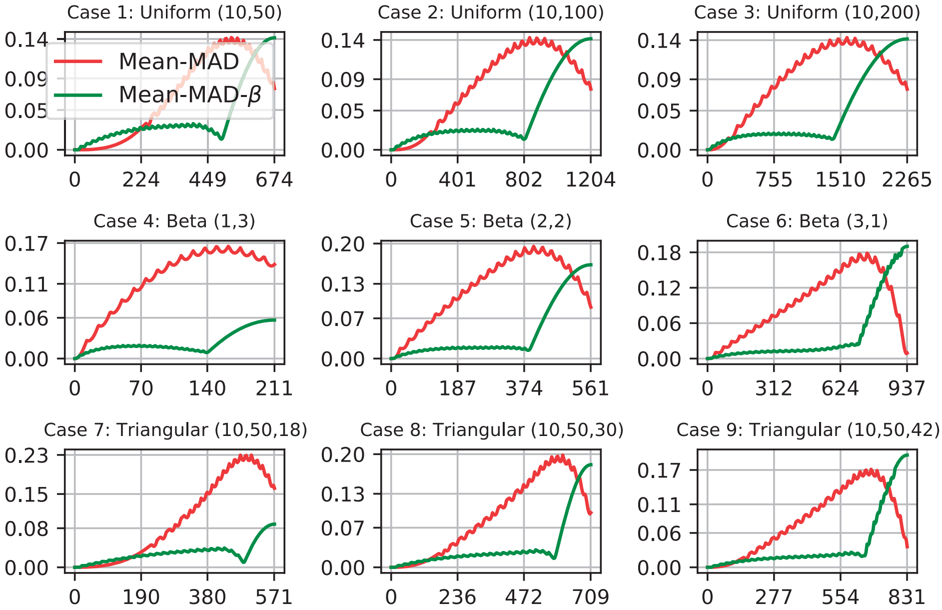


Figure 6.3: The results for the low margin setting, where the x-axis corresponds to B and the y-axis to the EVAI

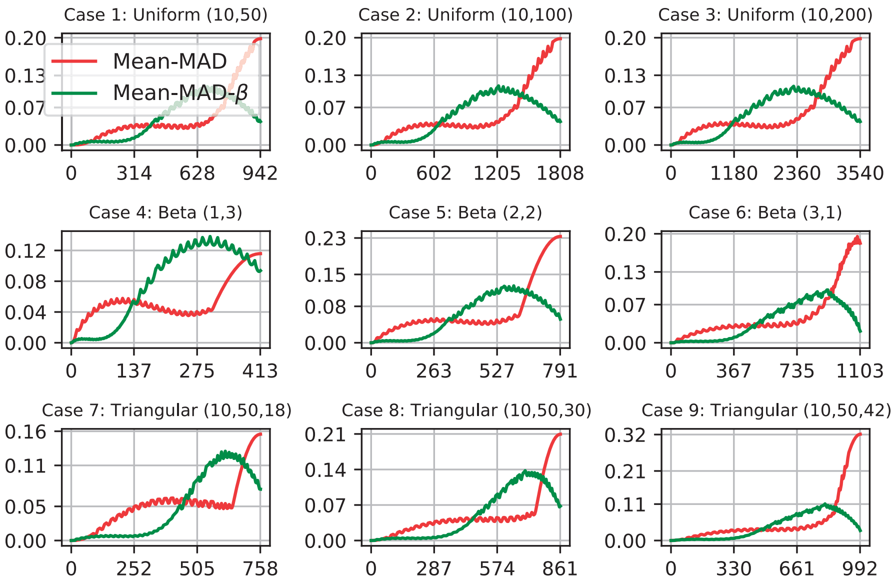


Figure 6.4: The results for the average margin setting, where the x-axis corresponds to B and the y-axis to the EVAI

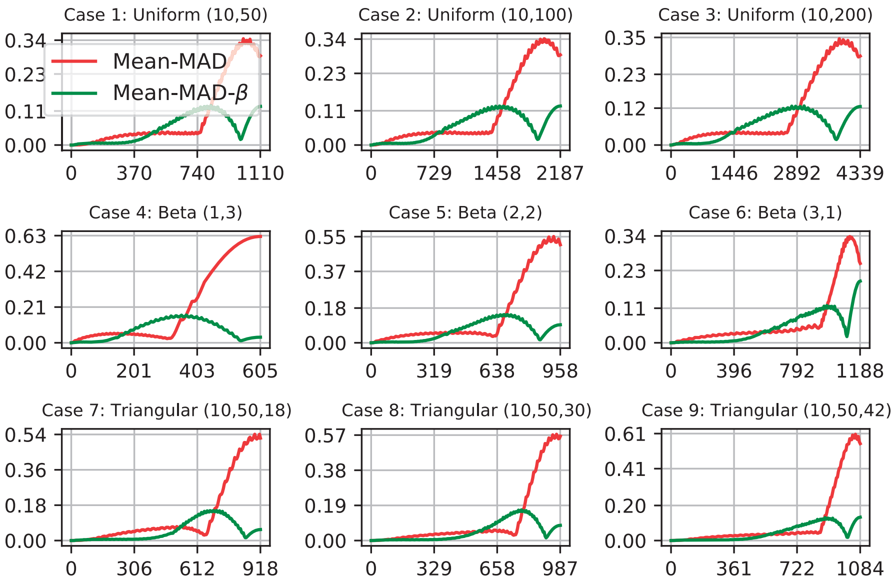


Figure 6.5: The results for the high margin setting, where the x-axis corresponds to B and the y-axis to the EVAI

We next quantify the value of MAD information by comparing the performance with the situations when only the mean and range of demand is known. For the low margin setting, Figure 6.6 shows the EVAI for the ordering policy with only mean-range information. Like the mean-MAD policy, this policy follows from a discrete distribution, in this case the extremal distribution on $\{a, b\}$ with probabilities $\frac{b-\mu}{b-a}$ and $\frac{\mu-a}{b-a}$ that attains the Edmundson-Madansky bound (see [19]). That is, instead of the worst-case three-point distribution, we take the expectation in (6.8) over this two-point distribution and find the robust mean-range ordering policy using the resulting LP. The plots clearly demonstrate that knowledge on dispersion in terms of MAD improves performance considerably.

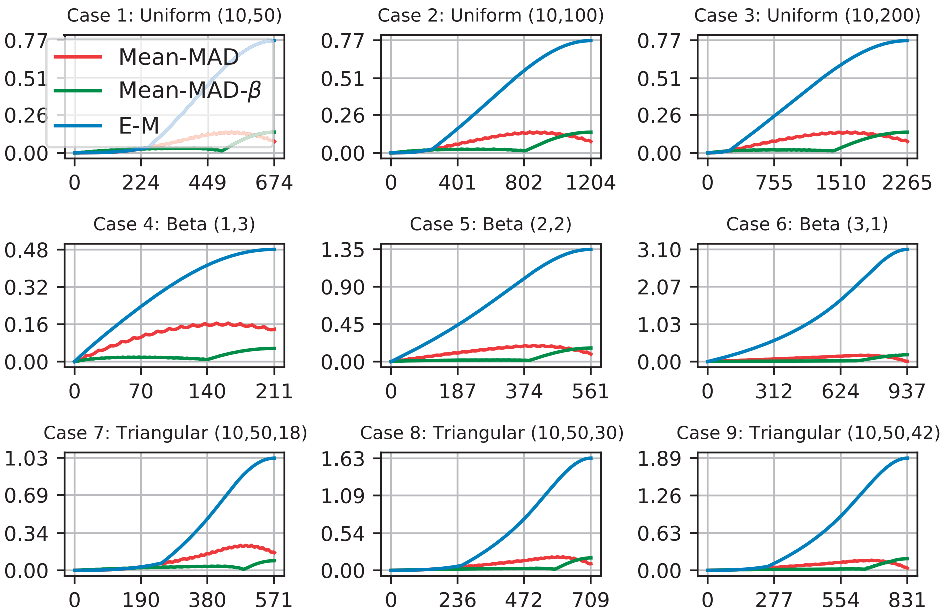


Figure 6.6: The results for the low margin setting, where E-M refers to the model with only mean information

6.5. Conclusions and outlook

This chapter establishes new ordering policies for the newsvendor with partial demand information (mean, MAD and range) with a budget constraint. The ordering policies follow from a minimax approach, where we search for the order quantities with minimal costs for the maximal (worst-case) cost function restricted to demand distributions that comply with the partial information.

The minimax analysis for the multi-item setting gives rise to a knapsack problem, and the solution of this knapsack problem in fact is the ordering policy. This policy prescribes to sort items based on their marginal effect on the total costs, reminiscent of the greedy algorithm that solves the continuous knapsack problem. The ordering policy only orders the minimum, mean or maximum demand for each item. Hence, the decision maker can rank the items based on their marginal effects, and then start ordering items according to this list until the budget is spent. The fact that the ranking list is easy to generate, and that the "order of ordering" does not depend on the budget, makes the policy transparent and easy to implement. Existing approaches for full and partial (such as mean-variance) knowledge of the demand distribution lack this property of "budget-consistency."

The minimax approach provides robustness, with an ordering policy that protects against all distributions that comply with the partial information. This approach avoids the need to estimate the demand distribution, which can be a daunting process in practice and is prone to errors. However, the minimax approach comes at the risk of being overly conservative. Through extensive numerical experiments we compared the robust policies for partial demand settings with the policies for full demand settings, and observed that the proposed policies perform well.

At the heart of our analysis lies the idea to set up the robust minimax analysis with MAD information. With MAD as dispersion measure we obtained a tractable optimization model, with a solution in terms of a robust ordering policy that satisfies the budget-consistency property. Using MAD to formulate solvable minimax problems can also be applied to other inventory models. We demonstrate this idea in [35] for three extended settings: the newsvendor with multiple contraints, the newsvendor with unreliable supply, and the risk-averse newsvendor. In all three cases, the minimax analysis leads to a tractable mathematical program, either a knapsack problem or a linear program.

7

Distributionally robust appointment scheduling that can deal with independent service times

7.1. Introduction

Unpredictability of arrival times and service times can lead to long waiting times and idleness, signals of poor quality of service and inefficient capacity usage. Ideally, a service system combines high quality and high efficiency. One way to achieve this dual goal is by having appointments scheduled in advance, so that customer waiting times are shortened and server utilization is increased. Appointment scheduling has long been recognized as a way to better regulate service processes. Spurred by increased availability of online reservation systems and teleservices, appointment scheduling rapidly gained further ground and became a driving factor in regulating modern economies.

Finding the best schedule of planned appointments is known in the queueing literature as the Appointment Scheduling Problem (ASP). The basic version of this problem involves a single server that serves a total of n customers over a period of length T. The service times are independent and identically distributed (i.i.d.) random variables and the objective function consists of the weighted sum of the expected waiting times and the expected idle times. The ASP then

This chapter is based on the research paper [211].

aims to determine the optimal schedule of arrival times that minimizes the objective function and strikes the right balance between wait and efficiency. Customers arrive punctually on the scheduled times (an assumption that can be easily relaxed) and wait in a queue before receiving service. Notice that queues are formed due to the stochastic service times. The ASP describes services that require instore or onsite visits, such as surgeries or outpatient clinics [40], but also remote care and teleservices. These are often service systems that operate close to maximum capacity, which can lead to long waiting times, urging the need to find a good if not optimal schedule.

Solving the ASP, and finding the optimal schedule, involves a stochastic optimization problem with an objective function that accounts for both the costs of idle times and the costs of waiting. This optimization problem requires as input the distributions of the uncertain sources, such as the service times (length of appointments) and arrival times. With historical data you could statistically estimate the service time distributions, but we take a different, more robust approach in this chapter. We work under the assumption that we do not know the distribution, but only have access to partial information in the form of summary statistics. More specifically, we assume that we know the mean, the mean absolute deviation (MAD) and the range of the service times. We then analyze the stochastic ASP with this partial information to find a robust optimal schedule that applies to all distributions that share the same partial information. We determine the robust schedules under partial information with a min-max analysis consisting of two steps. First we determine the worst-case (maximum) costs of the objective function under the conditions of the partial information. Then we determine the schedule by minimizing the maximum costs. As will become clear, the partial information consisting of mean, MAD and range presents a solvable maximization problem, which makes the minimization step particularly feasible and leads to optimal schedules with insightful structures. Just like the newsvendor problem in the previous chapter, the min-max approach in this chapter belongs to a much larger class of distributionally robust approaches that seek to calculate worst-case model performance, over the set of distributions satisfying some partial information.

Compared with tractable one-dimensional problems such as the newsvendor model, applying DRO techniques to problems with multiple *independent* random variables and distributions presents considerable, if not unsurmountable challenges. An early account of these specific challenges of multiple independent random variables can be found in [133], who searches for tight bounds for $\mathbb{E}[f(\mathbf{X})]$ with f a real-valued function and $\mathbf{X}$ a vector of random variables for which partial information is available, such as the first few moments. For the univariate case, Kingman shows that a standard optimization technique—related to the duality theory of general conic linear programs and moment problems [196]—yields tight bounds. For the multivariate case, Kingman explains that the same approach no longer works, because the dual problem can only be solved by allowing dependence between the random variables. That leads to less sharp bounds, or in Kingman's words: "It is not altogether surprising that this larger class of random variables allows a wider range of values of $\mathbb{E}[f(\mathbf{X})]$ to be attained. (...) These we must expect to be fairly weak, and the determination of refinements of these inequalities which are sharp is an unsolved problem, apparently of very considerable difficulty." It is indeed well known in the

DRO literature that multiple independent variables can cause major problems, and that allowing correlation often alleviates these problems. Another additional challenge comes with assuming that the random variables are not only independent, but also identically distributed, since this i.i.d. assumption introduces non-linearity to the optimization problem, often rendering conic duality theory unsuitable. For the classic ASP we are in fact facing this challenging setting with multiple i.i.d. random variables, as the objective function depends on all n independent service times.

7.1.1. Contributions and outline

We highlight the following contributions to the ASP literature:

1. We present a robust solution method for the classic ASP that is suitable for independent service times. For this we work with partial information that includes MAD instead of variance. This gives a solvable min-max problem, where the worst-case scenario leading to the maximum expected costs under partial information—the max of min-max—is given by independent three-point distributions for the service times. These three-point distributions do not depend on the number of customers, and hence remain the same for all problem sizes.

2. We also show that the min-max problem is computationally tractable, by performing the minimization after the maximization. The three-point distributions that follow from the maximization thus serve as input for the minimization, which can therefore be written as a standard stochastic program, solvable as linear program (LP) or as stochastic program with sample average approximation (SAA). To accurately approximate the optimal solution, SAA typically requires a number of samples that depends on n (i.e., the dimension of the problem) and the size of the feasible set. In our case, the required sample size for obtaining a near-optimal solution remains manageable, because we work with three-point distributions and therefore a relatively small feasible set.

3. Unlike other robust methods, our method leads to worst-case scenarios that respect the independence assumption. Most other robust studies do not impose independence, resulting in extreme worst-case scenarios, where long service times follow each other in rapid succession (due to the tolerated correlations). These scenarios in turn yield overly conservative schedules that build in most slack in the initial stages of the planning horizon. Our method gives realistic worst-case scenarios for the independent setting, and hence intuitive robust schedules that share universal features with the optimal schedules reported in the literature.

4. Finally, we show that our robust method does not only work for the classic ASP, but for many variants and extensions. We demonstrate our method for other settings with correlation structures and min-min (instead of min-max) analysis, but also for extended models that include features such as sequencing, no-shows, and risk-aversity. The latter extensions are typically difficult to solve in the partial-information setting. However,

when using mean-MAD information, the partial-information setting reduces to the full-information setting since we know the worst-case distribution, which is independent of the decision variables, and the extensions are therefore solvable as stochastic programs.

In what follows, we introduce a new perspective for solving the ASP under partial information by introducing a DRO approach that can deal with i.i.d. service times. We start with a detailed model description in Section 7.2, where we also discuss more ASP and DRO literature. Section 7.3 presents the new robust perspective for the ASP. We show how robust schedules can be determined as solutions to linear programs (LPs) and draw comparisons with other DRO approaches in the literature. Section 7.4 discusses several extensions of the classic ASP model, including correlated service times, no-shows and sequencing decisions when service times can have different distributions. We also discuss an alternative model for the ASP that uses as objective the conditional value at risk (CVaR) to address risk-aversity and unfairness. For all stochastic programs in Section 7.4, our robust method gives rise to LP reformulations and insightful robust schedules. We present some conclusions in Section 7.5.

7.2. Robust appointment scheduling

We now give a more precise specification of the appointment scheduling problem (ASP), with full and partial information, discuss more literature, and summarize our main contributions for robust appointment scheduling.

7.2.1. ASP with full information

Consider a total of n customers that need service from a single server during a period of length T. The goal is to find the appointment book (the arrival times of all n customers) that minimizes costs associated with waiting times, idle times and overtime. Let X_k denote the service time of customer k, and assume that all customers arrive precisely according to schedule. The scheduling problem is then to determine the interarrival times between customers, given the sequence in which they arrive. We use s_k to denote the interarrival time between job k and the next job, i.e., the length of the slot reserved for job k. The vector $\mathbf{s} = (s_1, s_2, \dots, s_n)$ will be referred to as the schedule. Job k is thus scheduled to arrive at time $\sum_{i=1}^{k-1} s_i$. Define the set of all possible schedules $\mathscr{S} = \{\mathbf{s} \in \mathbb{R}_+^n : s_1 + \dots + s_n \leqslant T\}$, where T is a positive upper time limit within which the schedules should be fit. Let W_k denote the delay of job k and let W_{n+1} denote the overtime. Delay costs are incurred at rate c_w. If the last job is completed after time T, overtime costs are incurred at rate c_o. Then $W_1 = 0$ and $W_k = (W_{k-1} + X_{k-1} - s_{k-1})^+,\ k = 2\dots, n+1$. It is then natural to seek a schedule $\mathbf{s} \in S$ that minimizes the total delay and overtime costs

$$f_n(\mathbf{s}, \mathbf{X}) = c_w \sum_{k=2}^{n} W_k + c_o W_{n+1}, \tag{7.1}$$

and the classic ASP can be formulated as

$$\min_{\mathbf{s}} \mathbb{E}_{\mathbb{P}}[f_n(\mathbf{s}, \mathbf{X})] \tag{7.2}$$

with the expectation taken over the distribution $\mathbb{P}$ of the service times $\mathbf{X} = (X_1, \dots, X_n)$.

The ASP thus tries to find an optimal design through an objective function that weighs conflicting performance measures, a tried and tested concept in OR. A prerequisite for successful execution is the availability of tractable expressions for the performance measures and thus the objective function. Special for the ASP is that the objective function consists of performance measures of a single-server queue, which in full generality renders an intractable optimization problem. With that in mind we divide the ASP literature into two methodological directions.

The first direction uses advanced queueing theory and approaches the ASP with explicit queue analysis and tractable objective functions [10, 93, 110, 126, 138, 144, 156, 217, 230, 231, 234]. This queueing approach requires further simplifying assumptions that facilitate exact analysis. Many works in this direction use specific service time distributions (such as exponential and phase-type distributions) and assume stationarity, hence approximating the ASP with finitely many appointments with an idealized model that considers an infinite number of appointments. Indeed, stationary, long-term queue analysis is simpler than non-stationary analysis. Since the stationary regime makes no distinction between customers, the approximate schedule is equidistant with slots of fixed size, which in many cases is far from optimal.

The second direction avoids advanced queueing theory and views the ASP as an optimization problem which only requires the distributions of the service times and the *queue dynamics*, i.e., the recursive relations that describe the waiting times and idles times as function of the appointment times and service times. The expected value problem (7.2) can be (approximately) solved using various methods, including SAA, quasi-gradient methods and sequential bounding approaches; see, e.g., [13, 33, 183, 198]. SAA leverages that, for a given realization of service times $\mathbf{x} = (x_1, \dots, x_n)$, the optimal schedule solves the LP

$$\begin{aligned} f_n(\mathbf{s}, \mathbf{x}) = \min_{\mathbf{w}} \ & c_w \sum_{k=2}^{n} w_k + c_o w_{n+1} \\ \text{s.t. } & w_2 \geqslant x_1 - s_1, \\ & w_{k+1} \geqslant w_k + x_k - s_k, \quad k = 2, \dots, n, \\ & w_k \geqslant 0, \quad k = 2, \dots, n+1. \end{aligned} \tag{7.3}$$

For more background on the numerical implications of SAA, see Shapiro et al. [198, Section 5.3]. These optimization methods can be computationally intensive and moreover require a precise description of the distribution $\mathbb{P}$, or at least a sufficient number of independent samples from this distribution.

7.2.2. Robust ASP with partial information

The computational problems of the full-information ASP, and the wish to design robust schedules that can deal with situations of partial information, have triggered earlier studies using min-max analysis and partial distributional information of the uncertain service times such as range, marginal moments and covariance. The distribution-free methods, proposed in these studies, lead to schedules that minimize the worst-case expected objective value among all possible distributions that comply with the partial information. The robust ASP can be formulated as

$$\min_{\mathbf{s}\in\mathscr{S}} \max_{\mathbb{P}\in\mathscr{P}} \mathbb{E}_{\mathbb{P}}[f_n(\mathbf{s},\mathbf{X})] \tag{7.4}$$

with $\mathscr{S}$ the feasible set of appointment schedules and ambiguity set $\mathscr{P}$ taken to be the family of all probability distributions consistent with the information known about the probability distribution of $\mathbf{X}$. This information can include, e.g., moments and covariance structures. Several papers solve some version of the robust ASP in (7.4) [30, 135, 149, 171, 180]. Kong et al. [135] formulated for known mean and covariance a robust min-max problem that can be approximately solved by semidefinite programming. Mak et al. [149] constructed for known marginal moments of the service time a tractable conic program. Both approaches cannot cope well with the case of i.i.d. service times, for the reasons explained earlier. That is, without restricting covariance and only specifying marginal moments, the worst-case probability distributions often correspond to highly correlated, unrealistic service times, see [149], while explicit specification of covariance leads to nonlinear hard-to-solve optimization models, see [135]. We will discuss these and other DRO papers [30, 171] in more detail later, but none of them can handle the setting of independent service times. To understand why the i.i.d. case stays out of reach, observe that independent service times make the expectation in (7.4) nonlinear and nonconcave (viewed as function of the probability distribution). Therefore, the associated semi-infinite program is no longer a convex optimization problem, which makes it hard, if not impossible, to find the worst-case distribution.

7.2.3. Features of optimal schedules

Several universal features of optimal schedules were reported in the literature. For i.i.d. service times with a known distribution, Wang [217], Denton and Gupta [68] and Robinson and Chen [183] observed independently that optimal schedules typically follow a "dome-shape" pattern, with successive time slots first increasing and later decreasing. This schedule allocates more time for customers in the middle and less time to customers at the beginning and the end of the day. This universal pattern fits with queueing theory intuition. Delays that arise early cause further delays, which pleads for longer slots in early stages. However, longer slots increase the chance of idleness. Shorter slots at the beginning and end of the day prevent idling and overtime, while longer slots in the middle of the day protect against long waiting times for many customers.

In most robust studies with only partially known service time distributions, it is not the dome shape but a "decreasing" schedule in which the slot lengths become smaller [123, 149]. This has to do with correlations between service times that are tolerated in the robust optimization models. The robust schedule is meant to guard against the worst-case scenario, which due to correlations, will be a scenario where long service times will follow each other at the beginning of the period. Because early delays are likely to transfer into delays for downstream jobs, this potentially creates major problems that can be mitigated by allocating more time for these first jobs. Such worst-case scenarios with positively correlated service times of consecutive customers inevitably lead to a strategy that avoids early delays, hence the decreasing schedule.

This decreasing schedule is also at odds with the famous Bailey's scheduling rule that lets the first two customers arrive at the start of the day, and schedules succeeding appointments at intervals equal to the expected service time [10, 135]. Bailey shows, using Monte-Carlo simulation, that this rule—which is an extreme case of an increasing (as opposed to decreasing) schedule—strikes a good balance between idle times and waiting times. With two customers present the server will not idle early on, while the waiting times remain acceptable. Allowing or banning correlations among service times thus leads to vastly different worst-case scenarios and min-max system designs. When anticipated, or statistically estimated, it makes sense to include correlations. However, there are also many situations where the independence assumption is more realistic and preferable. Moreover, the classic ASP with its origin in the single-server queue and inherent independence assumption serves as the benchmark model in the field. We therefore develop a robust optimization method for the ASP that can deal with independent service times. We will also explain why most other robust ASP studies cannot deal with the independence assumption.

7.3. Novel DRO approach for ASP

In Section 7.3.1 we discuss the ASP under partial information and present our min-max method for mean, MAD and range information of the service time distribution. In Section 7.3.2 we present a min-min method, complementing the upper bounds found in Section 7.3.1 with lower bounds. In Section 7.3.3 we discuss the numerical aspects of this approach, and present several structural properties of the robust schedules.

7.3.1. Solvable min-max problem

We now turn to the distributionally robust approaches for the ASP (7.4) and seek for the solution of the min-max problem for independent service times. Observe that

$$f_n(\mathbf{s},\mathbf{X}) = c_w \sum_{k=1}^{n-1} g_k(\mathbf{s},\mathbf{X}) + c_o g_n(\mathbf{s},\mathbf{X}) \tag{7.5}$$

with

$$g_k(\mathbf{s},\mathbf{X}) = \max\{X_k - s_k, \sum_{j=k-1}^{k}(X_j - s_j), \dots, \sum_{j=1}^{k}(X_j - s_j)\}, \quad k = 1, \dots, n, \tag{7.6}$$

which gives the following result:

Lemma 7.1. *The function $f_n(\mathbf{s},\mathbf{X})$ is jointly convex in $\mathbf{s}$ and $\mathbf{X}$.*

Because the function $f_n(\mathbf{s},\cdot)$ is convex, we choose to work with the mean-MAD ambiguity set

$$\mathcal{P}_{(\mu,d)} = \left\{\mathbb{P} : \ \mathrm{supp}(X_k) \subseteq [a_k, b_k],\ \mathbb{E}_{\mathbb{P}}[X_k] = \mu_k,\ \mathbb{E}_{\mathbb{P}}\left[|X_k - \mu_k|\right] = d_k,\ \forall k,\ X_k \perp\!\!\!\perp X_j,\ \forall k \neq j\right\}, \tag{7.7}$$

where $X_k \perp\!\!\!\perp X_j$, $\forall k \neq j$, denotes stochastic independence of the components $X_1, \dots, X_n$. This ambiguity set is known to generate, in conjunction with convex objective functions, explicit worst-case distributions and tight bounds, see [19, 179]. After applying these results to the maximization problem in (7.4), we obtain the following result:

Lemma 7.2. *The extremal distribution that solves* $\max_{\mathbb{P}\in\mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}[f_n(\mathbf{s},\mathbf{X})]$ *consists for each* X_k *of the three-point distribution on* $\{a_k, \mu_k, b_k\}$*.*

For notational convenience, we denote the three values of the extremal distribution of X_k as $\tau_1^{(k)} = a_k$, $\tau_2^{(k)} = \mu_k$, $\tau_3^{(k)} = b_k$ and the associated probabilities as

$$p_1^{(k)} = \frac{d_k}{2(\mu_k - a_k)}, \quad p_2^{(k)} = 1 - \frac{d_k}{2(\mu_k - a_k)} - \frac{d_k}{2(b_k - \mu_k)}, \quad p_3^{(k)} = \frac{d_k}{2(b_k - \mu_k)}. \tag{7.8}$$

Since this extremal distribution is independent of $\mathbf{s}$, we can substitute the 3^n terms (all values of the n independent three-point distributions of $X_1, \dots, X_n$), and still maintain a convex function in $\mathbf{s}$. Therefore, the minimization problem over $\mathbf{s}$ and hence the robust ASP with independent service times and mean-MAD ambiguity is equivalent with

$$\min_{\mathbf{s}\in\mathscr{S}} \max_{\mathbb{P}\in\mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}[f_n(\mathbf{s},\mathbf{X})] = \min_{\mathbf{s}\in\mathscr{S}} \sum_{\boldsymbol{\alpha}\in\{1,2,3\}^n} \prod_{i=1}^{n} p_{\alpha_i}^{(i)} f_n(\mathbf{s}, \tau_{\alpha_1}^{(1)}, \dots, \tau_{\alpha_n}^{(n)}). \tag{7.9}$$

By describing the maximum operator in (7.6) in terms of linear terms, (7.9) can be formulated as a tractable linear optimization problem:

Proposition 7.3. *The optimization problem* $\min_{\mathbf{s}\in\mathscr{S}} \max_{\mathbb{P}\in\mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}[f_n(\mathbf{s},\mathbf{X})]$ *can be written as*

$$\begin{aligned}
\min_{\mathbf{s}\in\mathscr{S}} \quad & \sum_{\boldsymbol{\alpha}\in\{1,2,3\}^n} \Big(\prod_{i=1}^{n} p_{\alpha_i}^{(i)} \Big(c_w \sum_{k=1}^{n-1} g_{\boldsymbol{\alpha}}^{(k)} + c_o g_{\boldsymbol{\alpha}}^{(n)} \Big) \Big) \\
\text{s.t.} \quad & g_{\boldsymbol{\alpha}}^{(k)} \geqslant 0, \quad k = 1, \dots, n, \ \forall \boldsymbol{\alpha} \in \{1,2,3\}^n, \\
& g_{\boldsymbol{\alpha}}^{(k)} \geqslant \sum_{j=l}^{k} (\tau_{\alpha_j}^{(j)} - s_j), \quad k = 1, \dots, n, \ l = 1, \dots, k, \ \forall \boldsymbol{\alpha} \in \{1,2,3\}^n.
\end{aligned} \tag{7.10}$$

We solve the LP (7.10), and other LPs presented later, with the programming language Julia and Gurobi 9.1. Because the number of constraints grows exponentially in n, the problem size that we can deal with is limited (say $n \leqslant 15$), even when exploiting the problem structure with the L-shaped method (see, e.g., [68]). However, since the worst-case distribution of the service times is known explicitly, and consists of only three support points, SAA presents an effective way of computing the bounds for large problem sizes.

7.3.2. Solvable min-min problem

We now show that tight lower bounds for the ASP objective function can be obtained by adding skewness information $\beta = \mathbb{P}(X \geqslant \mu)$ and hence considering the ambiguity set

$$\mathscr{P}_{(\mu,d,\beta)} = \left\{ \mathbb{P} \ : \ \mathbb{P} \in \mathscr{P}_{(\mu_k,d_k)}, \ \mathbb{P}(X_k \geqslant \mu_k) = \beta_k, \ \forall k \right\}. \tag{7.11}$$

Based on this ambiguity set, [19] also derives a tight lower bound as well as the distribution that attains this bound for the expectation of a convex function of random variables. Hence, again relying on the convexity shown in Lemma 7.1, we derive a lower bound for the ASP:

Lemma 7.4. *The distribution that solves* $\min_{\mathbb{P}\in\mathscr{P}_{(\mu,d,\beta)}} \mathbb{E}_{\mathbb{P}}[f_n(\mathbf{s},\mathbf{X})]$ *consists for each* X_k *of the two-point distribution on* $\{\mu_k + \frac{d_k}{2\beta_k}, \mu_k - \frac{d_k}{2(1-\beta_k)}\}$.

Again for notational convenience, write

$$q_1^{(k)} = \beta_k, \quad q_2^{(k)} = 1-\beta_k, \quad v_1^{(k)} = \mu_k + \frac{d_k}{2\beta_k}, \quad v_2^{(k)} = \mu_k - \frac{d_k}{2(1-\beta_k)}. \tag{7.12}$$

We can substitute the 2^n terms (all values of the the n independent two-point distributions of $X_1,\dots,X_n$) to obtain the following result:

Proposition 7.5. *The optimization problem* $\min_{\mathbf{s}\in\mathscr{S}} \min_{\mathbb{P}\in\mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}[f_n(\mathbf{s},\mathbf{X})]$ *can be written as*

$$\begin{aligned}
\min_{\mathbf{s}\in\mathscr{S}} \quad & \sum_{\boldsymbol{\alpha}\in\{1,2\}^n} \Big(\prod_{i=1}^{n} q_{\alpha_i}^{(i)} \big(c_w \sum_{k=1}^{n-1} g_{\boldsymbol{\alpha}}^{(k)} + c_o g_{\boldsymbol{\alpha}}^{(n)} \big) \Big) \\
\text{s.t.} \quad & g_{\boldsymbol{\alpha}}^{(k)} \geqslant 0, \quad k = 1,\dots,n, \ \forall \boldsymbol{\alpha} \in \{1,2\}^n, \\
& g_{\boldsymbol{\alpha}}^{(k)} \geqslant \sum_{j=l}^{k} (v_{\alpha_j}^{(j)} - s_j), \quad k = 1,\dots,n, \ l = 1,\dots,k, \ \forall \boldsymbol{\alpha} \in \{1,2\}^n.
\end{aligned} \tag{7.13}$$

The upper and lower bounds give a closed interval for $\mathbb{E}_{\mathbb{P}}[f_n(\mathbf{s},\mathbf{X})], \ \forall \mathbb{P} \in \mathscr{P}_{(\mu,d,\beta)}$. Denote the optimal values of (7.10) and (7.13) by g_U and g_L, respectively. The upper and lower bounds then also provide closed intervals for the optimal values:

Corollary 7.6. *For all distributions* $\mathbb{P} \in \mathscr{P}_{(\mu,d,\beta)}$,

$$\min_{\mathbf{s}\in\mathscr{S}} \mathbb{E}_{\mathbb{P}}[f_n(\mathbf{s},\mathbf{X})] \in [g_L, g_U].$$

Using the skewness information, we thus obtain a sharp performance interval for the expected value of $f_n(\mathbf{s},\mathbf{X})$ instead of only an upper bound. Figure 7.1 shows the optimal values of the ASP problem under mean-MAD ambiguity, where $(c_w, c_o) = (1,2)$, $[a,b] = [1,5]$, $\mu = 3$, $T = n\mu$, $d = 1$ and $\beta = 1/2$. As a result of Corollary 7.6, the red and green lines together provide a tight upper and lower bound for all distributions in $\mathscr{P}_{(3,1,1/2)}$ with support on the interval $[1,5]$. As a point of reference, we also plot the exact costs in the case that the X_k represent $U(1,5)$ distributed service times, which is a member of this ambiguity set.

Although the above min-min approach provides valuable insights for robust ASP performance, we will discuss in the remainder of this chapter only the min-max perspective.

7.3.3. Optimal robust schedules

We now present some first numerical results regarding the structure of the optimal schedules, based on the LP formulation (7.10). From this subsection onward, the optimal schedules depicted in the figures are obtained by solving (7.10) with SAA using 100,000 samples.

Figure 7.2 shows the minimax schedules for several settings with 25 customers. The optimal schedule for the uniformly distributed service times is indeed dome-shaped, see [68], and turns

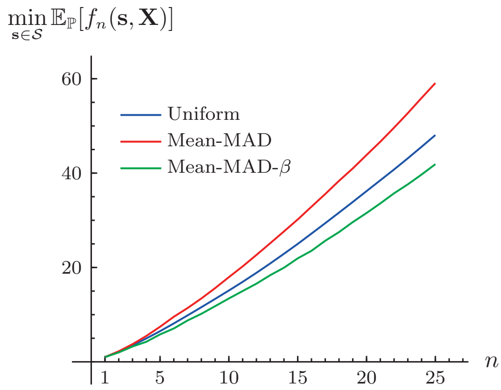
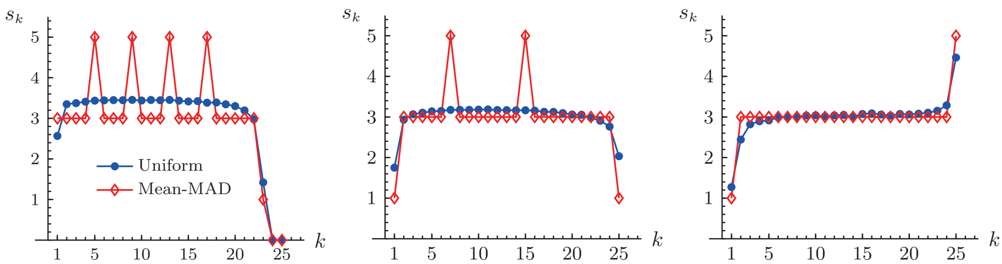


Figure 7.1: Tight bounds for the optimal objective values of the ASP (obtained by SAA using 100,000 samples)

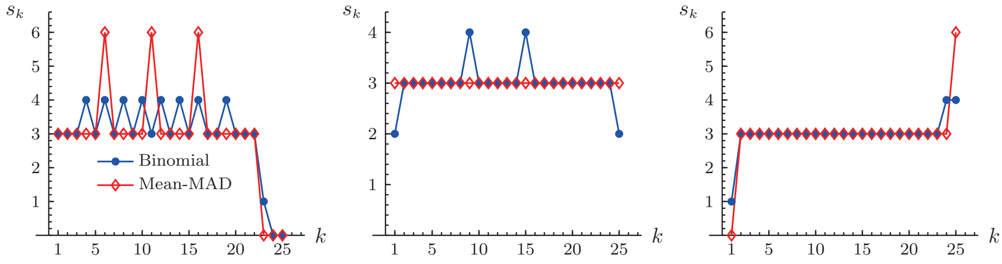


(a) $c_o = 1$ **(b)** $c_o = 10$ **(c)** $c_o = 40$

Figure 7.2: Optimal time allocation for $c_w = 1$ and $n = 25$ with schedules based on the worst-case three-point distribution and $U(1,5)$ distributed service times (obtained through SAA with 100,000 samples)

from dome-shaped to increasing when overtime is more heavily penalized. The robust schedule also resembles a dome shape, but of a different nature: the schedule installs longer slots in a periodic fashion. We see four such longer slots in Figure 7.2a, and two in Figure 7.2b. Since waiting times build up progressively, these longer slots can protect against long waiting times. Then, in Figures 7.2a and 7.2b, at some point during the day, the schedule becomes constant, and the final customers arrive relatively early since the schedules should be fitted within the time limit. The spiked structure of the robust schedules in Figures 7.2a and 7.2b can be viewed as a robust counterpart of the earlier observed dome-shaped schedules. In Figure 7.2c we see that for large overtime costs, the optimal schedule becomes nearly constant, except for the final customer for which additional time is reserved to avoid excessive overtime.

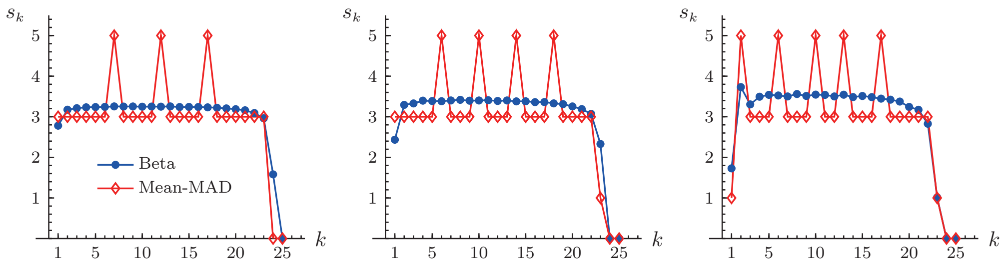
Also observe that the robust schedules in Figure 7.2a consist of integer-valued appointment slots, which indeed is the case when service times are discrete random variables; see Corollary 8.2 in Begen and Queyranne [14]. We see this also confirmed in Figure 7.3, where the reference distribution is the (discrete) Binomial distribution.



(a) $c_o = 1$ **(b)** $c_o = 10$ **(c)** $c_o = 40$

Figure 7.3: Optimal time allocation for $c_w = 1$ and $n = 25$ with schedules based on the worst-case three-point distribution and Binomial$(6, \frac{1}{2})$ distributed service times

Figure 7.3a again reveals the spiked patterns, but now turns for higher overtime costs into a flat line in Figure 7.3b. This flat line means that the robust schedule prescribes equal slots for all customers. Figure 7.3c shows a schedule similar to the one in Figure 7.2c.



(a) Beta$(4, 4)$ **(b)** Beta$(1, 1)$ **(c)** Beta$(\frac{1}{4}, \frac{1}{4})$

Figure 7.4: Optimal time allocation for $c_w = 1$ and $n = 25$ with schedules based on the worst-case three-point distribution and Beta(γ, θ) distributed service times

Figure 7.4 shows three robust schedules for ambiguity sets that include the Beta distribution with increasing MAD. Observe that with increasing MAD, so when scenarios become less predictable, more spikes are scheduled in the early stages.

To compare the performance of the robust schedules against some optimal schedules, we use Monte Carlo simulation to compute the total costs of the mean-MAD schedule for n appointments, where $n \in \{10, 15, 25\}$. Our numerical experiments are based on [149]. The service times are generated under two common distribution types, the beta and triangular distributions, and all service times are independent but not necessarily identically distributed. That is, for each customer k, we randomly set the parameters of the beta distribution, $\gamma_k, \theta_k \sim U(\frac{1}{2}, 2)$, and draw the mode of the triangular distribution from $U(10, 20)$. Let all distributions be supported on the interval $[0, 30]$. For each generated instance, the total session length is determined as $T = \sum_{k=1}^{n} \mu_k + R \cdot \sum_{k=1}^{n} d_k$, with $R \in \{-0.5, 0, 0.5\}$. Next, we generate 10,000 independent samples from these distributions and use SAA to approximate the schedule that minimizes (7.2), which assumes full knowledge of the distribution. We also use SAA and 10,000 independent samples from the three-point distributions (with matching mean and MAD) to compute the optimal mean-MAD schedule. In addition, we also solve a second-order conic program, as in Mak et al. [149], to compute the mean-variance schedule, which does not specify any correlation or dependency structure. To compare the out-of-sample performance of the three scheduling solutions, a Monte Carlo simulation is performed. We generate 1,000,000 scenarios from the underlying distributions that were used to generate the near-optimal solution. We then estimate the percentage difference between the mean performance of the distributionally robust solutions and the full-information solution. For each parameter setting and type of distribution, we consider 50 randomly generated instances. We then report the average percentage differences over these 50 instances for each case.

Table 7.1 provides a number of interesting insights. First, observe that the differences between the expected costs of the mean-MAD schedule and the near-optimal solution are not large, ranging from 1.2% to 13.6%. Second, congestion in the system, controlled by the parameter R, affects the performance differences the most, whereas the performance of the mean-MAD schedule is barely dependent on the number of customers n. In contrast, the performance of the mean-variance schedule deteriorates as n increases since it guards against scenarios where a large number of consecutive jobs have long durations simultaneously, which, in the setting with independence, leads to overly conservative schedules and hence high average costs.

Table 7.1: Percentage difference between the expected costs of the robust schedules and the full information schedule when scheduling 25 appointments

			Mean-MAD			Mean-variance		
Perf. measure	R	$X_k \sim \mathbb{P}$	10	15	25	10	15	25
mean	-0.25	Beta$(2, \theta)$	1.6%	1.5%	1.2%	8.0%	11.5%	17.8%
		Beta$(\gamma, 2)$	2.4%	2.2%	2.7%	7.2%	11.6%	20.6%
		Triangular	3.5%	4.9%	4.6%	7.9%	12.1%	19.9%
	0	Beta$(2, \theta)$	2.2%	2.2%	1.9%	10.9%	16.4%	27.9%
		Beta$(\gamma, 2)$	3.5%	4.4%	4.3%	10.3%	18.4%	31.3%
		Triangular	8.6%	6.9%	7.2%	10.8%	18.9%	30.4%
	0.25	Beta$(2, \theta)$	4.4%	3.5%	4.0%	16.4%	24.1%	42.0%
		Beta$(\gamma, 2)$	7.2%	6.3%	6.9%	17.0%	25.5%	48.7%
		Triangular	12.6%	10.8%	13.6%	17.4%	24.7%	46.5%

The numerical results in Table 7.1 demonstrate that the robust approach presents sharp performance guarantees, when compared to the (near-)optimal schedules. The performance gaps do not vary much when assuming different distributions. We conclude that the robust schedules are valuable in practice, and certainly not overly conservative, despite their ability to protect against all scenarios possible given the partial information.

7.4. Broader application of robust scheduling method

We now discuss extensions and variations of the standard ASP to which the novel robust approach developed in Section 7.3 can be applied. We first discuss three extensions: Section 7.4.1 discusses the model with sequencing decisions, Section 7.4.2 relaxes the assumption that service times are independent, and Section 7.4.3 considers the model where customers no longer arrive punctually at the scheduled times (or do not arrive at all). We then introduce in Section 7.4.4 the risk-averse ASP in which the conditional value at risk (CVaR) is the objective function, with the goal of reducing delay unfairness. In Section 7.4.5 we present a dynamic version of the ASP, which is solved with techniques from adjustable robust optimization. Finally, we discuss in Section 7.4.6 the stationary analysis of the ASP and show that our min-max approach can also deal with this setting that can serve as approximation when the number of appointments grows large.

7.4.1. Sequencing

In this section we discuss the problem of jointly determining the schedule and arrival sequence of customers, which we refer to as the appointment sequencing problem. We thus generalize the ASP model such that it includes the sequencing decision. Denote the sequencing decision by the variable $\mathbf{y} = (y_{km})_{k=1,\dots,n;m=1,\dots,n}$, where y_{km} is set to 1 if customer m is assigned to slot k. To ensure that the sequence of jobs is feasible, we make sure each slot is assigned to a customer (i.e., the assignment variables y_{km} add up to 1 for each k), and since every customer should be treated, the assignment variables amount to 1 for each m. Denote the feasible set by $\mathscr{Y}$. The distributionally robust appointment sequencing problem, under mean-MAD ambiguity, then reduces to a stochastic mixed-integer programming problem.

Proposition 7.7. *The optimization problem* $\min_{\mathbf{y}\in\mathscr{Y},\mathbf{s}\in\mathscr{S}} \max_{\mathbb{P}\in\mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}[f_n(\mathbf{s},\mathbf{X})]$ *can be written as*

$$\begin{aligned}
\min_{\mathbf{y}\in\{0,1\}^{n\times n},\, \mathbf{s}\in\mathscr{S}} \quad & \sum_{\boldsymbol{\alpha}\in\{1,2,3\}^n} \Big(\prod_{i=1}^{n} p^{(i)}_{\alpha_i} \Big(c_w \sum_{k=1}^{n-1} g^{(k)}_{\boldsymbol{\alpha}} + c_o g^{(n)}_{\boldsymbol{\alpha}} \Big) \Big) \\
\text{s.t. } & g^{(k)}_{\boldsymbol{\alpha}} \geqslant 0, \quad k=1,\dots,n,\ \forall \boldsymbol{\alpha}\in\{1,2,3\}^n, \\
& g^{(k)}_{\boldsymbol{\alpha}} \geqslant \sum_{j=l}^{k} \Big(\sum_{m=1}^{n} \tau^{(m)}_{\alpha_m} y_{jm} - s_k \Big), \quad k=1,\dots,n,\ l=1,\dots,k,\ \forall \boldsymbol{\alpha}\in\{1,2,3\}^n, \\
& \sum_{k=1}^{n} y_{km} = 1, \quad m=1,\dots,n, \\
& \sum_{m=1}^{n} y_{km} = 1, \quad k=1,\dots,n.
\end{aligned} \tag{7.14}$$

The sequencing decision does not affect convexity of $f_n(\mathbf{s},\mathbf{X})$ in $\mathbf{X}$, and therefore the three-point distribution again constitutes the extremal distribution. To solve problem (7.14), we can again resort to SAA.

We next show that, although a popular heuristic in practice, sequencing jobs by increasing order of variance (OV), or MAD, is not necessarily optimal for the distribution-free model with independent service times. Before doing so, we discuss some further literature. Since the sequencing decision makes the problem nonconvex, the appointment sequencing problem is significantly more difficult. For the model with only two jobs, [69] proves the optimal sequence is determined by ordering jobs by increasing variance of job duration. In [152] it is shown that the appointment scheduling problem, formulated as a stochastic mixed-integer program, is NP-hard, even when the number of job duration scenarios is finite. By exploiting a connection with serial supply-chain inventory models, a mixed-integer second-order cone programming approximation is developed in [148]. When only the mean and variance are known, [149] shows OV is optimal for their distributionally robust model. Using an analytical model, [136] provides insights into when the OV sequencing rule obtains superior performance. The OV rule is shown to be asymptotically optimal in [62], that is, as $n \to \infty$.

Contrary to the mean-variance discussed in [149], the optimal mean-MAD sequence is not necessarily obtained by ordering the jobs by increasing magnitude of the dispersion measure d. To demonstrate this, consider the following four instances with $n = 6$ jobs. Case 1 contains the following choices for the parameters: $\mu_k = 10$, $d_k = 1+0.5(k-1)$, $a_k = 0$ and $b_k = 15$ $k = 1, \dots, n$. For case 2, we have $\mu_k = 13.5 - k$, $d_k = 1 + 0.5(k-1)$, $a_k = 0$, $b_k = 15$ $k = 1, \dots, n$, and for case 3, let $\mu_k = 13.5 - k$, $d_k = 1 + 0.5(k-1)$, $a_1 = a_2 = a_3 = 0$, $a_4 = a_5 = a_6 = 5$, $b_1 = b_2 = b_3 = 25$ and $b_4 = b_5 = b_6 = 15$. For case 4, the parameters are given by $\mu_k = 13.5 - k$, $d_k = 2$, $a_k = 0$, $b_k = 15$ $k = 1, \dots, n$. In addition, we consider two values for the overtime costs: $c_o = 2$ and $c_o = 20$.

Table 7.2: Optimal job sequences for $c_o = 2$

n	1	2	3	4	5	6
case 1	1	2	3	4	5	6
s_k	10.0	10.0	10.0	10.0	10.0	10.0
case 2	1	2	3	4	5	6
s_k	12.5	11.5	13.0	9.5	8.5	5.0
case 3	1	2	4	5	3	6
s_k	12.5	11.5	9.5	11.0	10.5	5.0
case 4	4	5	6	3	2	1
s_k	9.5	8.5	7.5	10.5	11.5	12.5

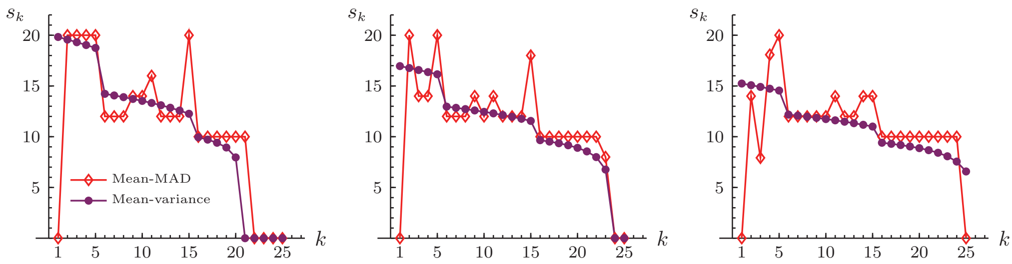
Table 7.3: Optimal job sequences for $c_o = 20$

n	1	2	3	4	5	6
case 1	1	2	3	4	5	6
s_k	10.0	10.0	5.0	10.0	10.0	15.0
case 2	1	2	6	5	4	3
s_k	12.5	11.5	2.5	8.5	10.0	15.0
case 3	1	6	2	3	5	4
s_k	12.5	5.0	11.5	10.5	5.5	15.0
case 4	6	5	4	1	2	3
s_k	5.0	7.5	10.5	12.5	11.5	13.0

Tables 7.2 and 7.3 provide the optimal sequencing and scheduling decisions for the three instances above, for both choices of c_o. When overtime costs are low, i.e. $c_o = 2$, the optimal sequence follows the increasing MAD pattern for cases 1 and 2. However, the optimal sequence for case 2 changes when the overtime costs are high, and instead places customers with higher expected service times at the end of the planning period, even though their service times are less variable. This ensures that larger jobs, with less variability, are dealt with at the end of the day to mitigate the risk of excessive overtime. Moreover, the third case shows that the range of the service time also affects the sequencing decision. Thus next to variability in terms of

MAD, the width of the support also brings about changes in the optimal sequence. We thus conclude that under the independence assumption the optimal job sequence does not admit a straightforward structure. However, we do expect the increasing MAD sequence to perform well as a heuristic, like the performance of OV in [69, 152].

In addition, we consider an example inspired by the numerical example in [135, Section 6.2]. Assume, in total, 25 customers are scheduled to arrive, divided into three distinct customer classes. The first class has the highest and most variable service requirements and consists of $n_1 = 5$ customers. The other two customer classes both consist of $n_2 = n_3 = 10$ customers and have lower mean and MAD. The customer classes are scheduled in decreasing order of mean and MAD. For the first customer class with $n_1 = 5$, $\mu_k = 14$ and $d_k = \sigma_k = 8$. For the second class consisting of $n_2 = 10$ customers, $\mu_k = 12$ and $d_k = \sigma_k = 4$, and for the final class with $n_3 = 10$ customers, the parameters $\mu_k = 10$ and $d_k = \sigma_k = 2$. The total period length $T = 276$. Figure 7.5 compares the mean-MAD schedule, which assumes independence, to the mean-variance schedule that allows all dependency structures. The mean-MAD schedule exhibits the "Bailey's rule + break" pattern, which was first observed in [135]. In contrast, the mean-variance schedule allots gradually decreasing slot lengths, and instead of inserting a break when switching to a different customer class, it schedules a tighter slot during the transition since the next customer class has lower service requirements.



(a) $c_o = 1$ **(b)** $c_o = 10$ **(c)** $c_o = 20$

Figure 7.5: Optimal time allocation for different customer classes with varying mean and MAD, and a fixed range $[0, 20]$

7.4.2. Correlations

As mentioned earlier, the assumption of independent service times is challenging and greatly influences the structure of optimal schedules. For the sake of comparison, we now consider the relaxation that allows for correlated service times. We explain why this setting is mathematically more tractable, but does lead to entirely different schedules.

Consider the ambiguity set

$$\mathscr{P}^{\text{corr}}_{(\mu,d)} = \{\mathbb{P} \ : \ \text{supp}(X_k) \subseteq [a_k, b_k], \ \mathbb{E}_{\mathbb{P}}[X_k] = \mu_k, \ \mathbb{E}_{\mathbb{P}}\left[|X_k - \mu_k|\right] = d_k, \ \forall k\}, \tag{7.15}$$

which is larger than $\mathscr{P}_{(\mu,d)}$ since correlations are allowed. In Appendix B.4, we extend results for marginal moments discussed in [149] to the setting of generalized moments (such as MAD) to obtain the following result:

Proposition 7.8. *Assume* $\mu_k \in (a_k, b_k)$, $d_k \in (0, \frac{2(b_k-\mu_k)(\mu_k-a_k)}{(b_k-a_k)})$, $\forall k$. *Then, the optimization problem* $\min_{\mathbf{s}\in\mathscr{S}} \max_{\mathbb{P}\in\mathscr{P}^{\mathrm{corr}}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}[f_n(\mathbf{s},\mathbf{X})]$ *can be written as*

$$\begin{aligned}
\min_{\mathbf{s},\zeta,\xi,\lambda^{(1)},\lambda^{(2)}} \quad & \sum_{k=1}^{n} \left(\zeta_k + \mu_k\lambda_k^{(1)} + d_k\lambda_k^{(2)}\right) \\
\text{s.t.} \quad & \sum_{i=k}^{\min\{n,j\}} \zeta_i \geqslant \sum_{i=k}^{\min\{n,j\}} \left(\xi_{ij} - s_i\pi_{ij}\right), && 1 \leqslant k \leqslant n,\ k \leqslant j \leqslant n+1, \\
& \xi_{ij} \geqslant \left(\pi_{ij} - \lambda_i^{(1)}\right) b_i - \lambda_i^{(2)}(b_i - \mu_i), && 1 \leqslant i \leqslant n,\ i \leqslant j \leqslant n+1, \\
& \xi_{ij} \geqslant \left(\pi_{ij} - \lambda_i^{(1)}\right) \mu_i && 1 \leqslant i \leqslant n,\ i \leqslant j \leqslant n+1, \\
& \xi_{ij} \geqslant \left(\pi_{ij} - \lambda_i^{(1)}\right) a_i - \lambda_i^{(2)}(\mu_i - a_i), && 1 \leqslant i \leqslant n,\ i \leqslant j \leqslant n+1, \\
& \mathbf{s} \in \mathscr{S},
\end{aligned} \tag{7.16}$$

where

$$\pi_{ij} = \begin{cases} c_w(j-i), & \text{for } 1 \leqslant i \leqslant j \leqslant n, \\ c_o + c_w(n-i), & \text{for } 1 \leqslant i \leqslant n,\ j = n+1. \end{cases} \tag{7.17}$$

Notice that the complexity of the LP (7.16) dropped compared to (7.10), with the number of constraints reduced to $O(n^2)$. Hence, this model remains tractable for larger values of n. However, since we can no longer deal with i.i.d. service times, (7.16) produces vastly different schedules.

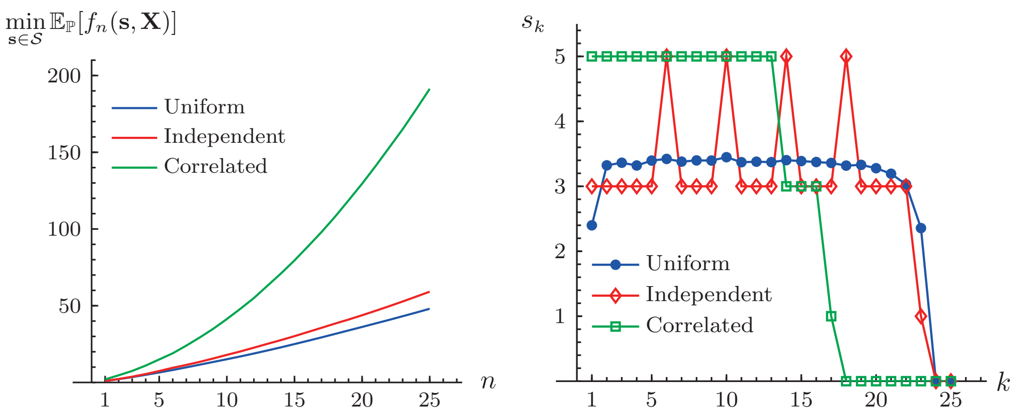


Figure 7.6: Tight bounds and optimal solutions for the ASP

As a reference consider $U(1,5)$ distributed service times, $c_w = 1$ and $c_o = 2$. Next to the "true" optimal value, we compare the mean-MAD upper bound for the original model with i.i.d. service times and ambiguity set $\mathscr{P}_{(\mu,d)}$, and the relaxation with possibly correlated service times and ambiguity set $\mathscr{P}^{\text{corr}}_{(\mu,d)}$. Figure 7.6 depicts the optimal values and the corresponding schedules. Observe that the schedule that allows correlated worst-case distributions provides much looser bounds compared to the tight bounds that follow from (7.10). This is as expected, since dropping the independence condition enlarges the solution space with distributions that tolerate correlations between the service times. As a consequence, the optimal schedules lose the dome structure, and instead prescribe monotonically decreasing slot lengths as reported in earlier DRO studies [149, 180].

7.4.3. No-shows

Let I_k be a Bernoulli(q_k) indicator variable that equals 1 when customer k shows up, and 0 otherwise. The delay of job k now satisfies

$$W_k = (W_{k-1} + I_{k-1}X_{k-1} - s_{k-1})^+. \tag{7.18}$$

Since expression (7.18) remains convex in the uncertain parameters, we obtain a similar result as in previous sections.

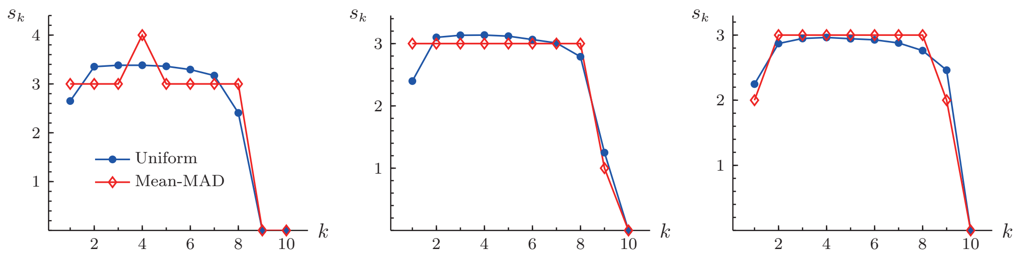
Proposition 7.9. *Let all I_k, Z_k be independent, with $I_k \sim$ Bernoulli(q_k). Then, the optimization problem* $\min_{\mathbf{s}\in\mathscr{S}} \min_{\mathbb{P}\in\mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}[f_n(\mathbf{s},\mathbf{X})]$ *can be written as*

$$\begin{aligned}
&\min_{\mathbf{s}\in\mathscr{S}} \sum_{\boldsymbol{\alpha}\in\{1,2,3\}^n\times\{0,1\}^n} \Big(\prod_{i=1}^{n} p^{(i)}_{\alpha_i} \big(c_w \sum_{k=1}^{n-1} g^{(k)}_{\boldsymbol{\alpha}} + c_o g^{(n)}_{\boldsymbol{\alpha}}\big)\Big) \\
&\text{s.t. } g^{(k)}_{\boldsymbol{\alpha}} \geqslant 0, \quad k = 1,\dots,n,\ \forall \boldsymbol{\alpha} \in \{1,2,3\}^n \times \{0,1\}^n, \\
&\qquad g^{(k)}_{\boldsymbol{\alpha}} \geqslant \sum_{j=l}^{k} (q^{(j)}_{\alpha_j}\tau^{(j)}_{\alpha_j} - s_j), \quad k = 1,\dots,n,\ l = 1,\dots,k,\ \forall \boldsymbol{\alpha} \in \{1,2,3\}^n \times \{0,1\}^n.
\end{aligned} \tag{7.19}$$

Problem (7.19) considers all possible scenarios for X_k and I_k. Here we still charge waiting time costs regardless of whether the customer actually shows up, as in Jiang et al. [123]. Alternatively, one might waive the waiting time costs when a customer does not show up by modifying the objective function, thus modeling the waiting time costs as $c_w I_k W_k$.

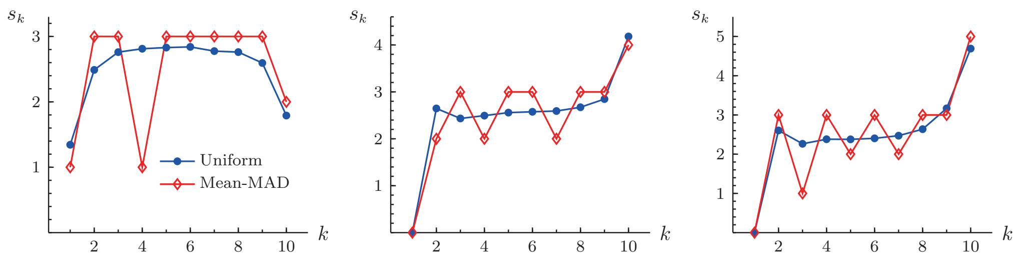
The LP in (7.19) presents a new DRO perspective for ASP with no shows. Other DRO studies are [123] with ambiguity for the no-show variables, and [137] in which this model is extended by introducing time-dependent no-show probabilities and deriving a conic programming formulation. We now compare our robust approach for partial information to the full information ASP with no shows, an extension of work in [68] covered in [76].

We now perform some numerical experiments motivated in part by the examples in Section 5.1 of Erdogan and Denton [76]. As a reference, we also consider the setting with all service times independent and identically distributed as $U(1,5)$. Figures 7.7 and 7.8 provide the optimal schedules for the ASP with no shows. Figure 7.7 shows that, when no shows occur with probability $1 - q_k = 0.4$, the optimal schedules allot less time to the customers in the middle of the session to hedge against possible idling of the server in case of a no show. In addition, for large overtime costs $c_o = 20$, the first two customers are double booked as with Bailey's rule.



(a) $q_k = 1,\ c_o = 1$ **(b)** $1 - q_k = 0.2,\ c_o = 1$ **(c)** $1 - q_k = 0.4,\ c_o = 1$

Figure 7.7: Optimal time allocation for schedules based on known customer no-show probabilities with $n = 10,\ T = 25,\ X_k \sim U(1,5)$



(a) $q_k = 1,\ c_o = 20$ **(b)** $1 - q_k = 0.2,\ c_o = 20$ **(c)** $1 - q_k = 0.4,\ c_o = 20$

Figure 7.8: Optimal time allocation for schedules based on known customer no-show probabilities with $n = 10,\ T = 25,\ X_k \sim U(1,5)$

7.4.4. Risk aversion and delay unfairness

Consider a risk-averse version of the ASP where one makes decisions based on CVaR. First introduced as a performance measure for financial risk management, CVaR denotes the average value of the costs exceeding the ϱth quantile of the cost distribution. For $f_n(\mathbf{s}, \mathbf{X})$, CVaR follows from solving a convex minimization problem [184, 185]. That is,

$$\mathrm{CVaR}_\varrho[f_n(\mathbf{s}, \mathbf{X})] = \min_{\vartheta \in \mathbb{R}} \left\{ \vartheta + \frac{1}{1-\varrho} \mathbb{E}(f_n(\mathbf{s}, \mathbf{X}) - \vartheta)^+ \right\}.$$

However, as the exact distribution is unknown, we consider the partial information setting as in [235], and solve

$$\min_{\mathbf{s} \in \mathscr{S}} \max_{\mathbb{P} \in \mathscr{P}_{(\mu,d)}} \min_{\vartheta \in \mathbb{R}} \left\{ \vartheta + \frac{1}{1-\varrho} \mathbb{E}_{\mathbb{P}}(f_n(\mathbf{s}, \mathbf{X}) - \vartheta)^+ \right\} = \min_{\mathbf{s} \in \mathscr{S}, \vartheta \in \mathbb{R}} \left\{ \vartheta + \frac{1}{1-\varrho} \max_{\mathbb{P} \in \mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}(f_n(\mathbf{s}, \mathbf{X}) - \vartheta)^+ \right\}, \tag{7.20}$$

where the identity holds by Sion's minimax theorem because the objective function of (7.20) is convex in ϑ, concave in $\mathbb{P}$, and $\mathscr{P}_{(\mu,d)}$ is weakly compact as $\mathrm{supp}(D)$ is compact. Since $(\cdot)^+$ preserves convexity, the three-point distribution maximizes $\mathbb{E}_{\mathbb{P}}(f_n(\mathbf{s}, \mathbf{X}) - \vartheta)^+$.

The risk-averse scheduler thus needs to solve the following problem:

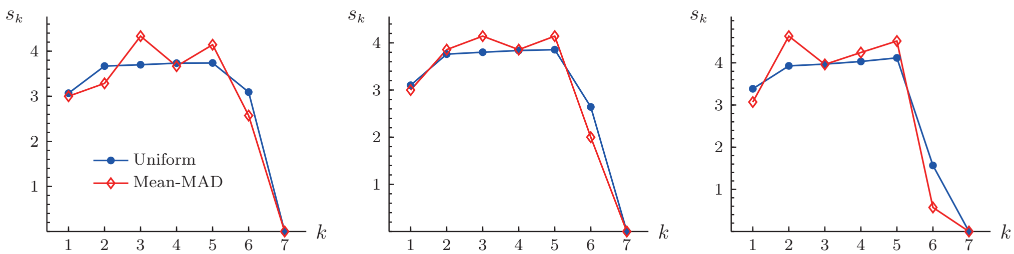
Proposition 7.10. *The optimization problem*

$$\min_{\mathbf{s}\in\mathscr{S}} \max_{\mathbb{P}\in\mathscr{P}_{(\mu,d)}} \min_{\theta\in\mathbb{R}} \left\{ \theta + \frac{1}{1-\rho}\mathbb{E}_{\mathbb{P}}(f_n(\mathbf{s},\mathbf{X}) - \vartheta)^+ \right\}$$

can be written as

$$\begin{aligned}
\min_{\mathbf{s}\in\mathscr{S},\boldsymbol{\eta},\vartheta}\ & \vartheta + \frac{1}{1-\varrho}\sum_{\boldsymbol{\alpha}\in\{1,2,3\}^n}\prod_{i=1}^{n} p_{\alpha_i}^{(i)}\eta_{\boldsymbol{\alpha}} \\
\text{s.t. } & \eta_{\boldsymbol{\alpha}} \geqslant c_w\sum_{k=1}^{n-1} g_{\boldsymbol{\alpha}}^{(k)} + c_o g_{\boldsymbol{\alpha}}^{(n)} - \vartheta, && \forall\boldsymbol{\alpha}\in\{1,2,3\}^n, \\
& \eta_{\boldsymbol{\alpha}} \geqslant 0, && \forall\boldsymbol{\alpha}\in\{1,2,3\}^n, \\
& g_{\boldsymbol{\alpha}}^{(k)} \geqslant 0, && k=1,\dots,n,\ \forall\boldsymbol{\alpha}\in\{1,2,3\}^n, \\
& g_{\boldsymbol{\alpha}}^{(k)} \geqslant \sum_{j=l}^{k}(\tau_{\alpha_j}^{(j)} - s_j), && k=1,\dots,n,\ l=1,\dots,k,\ \forall\boldsymbol{\alpha}\in\{1,2,3\}^n.
\end{aligned} \tag{7.21}$$

Figure 7.9 illustrates the optimal schedules when minimizing CVaR with the overtime costs set to $c_o = 2$. The schedule allots time to $n = 7$ customers, for which the service times are $U(1,5)$ distributed. The total time allowance is restricted to $T = 21$. For these parameter settings, the mean-MAD schedule closely resembles the optimal schedule under exact knowledge of the underlying distribution.



(a) $\varrho = 0.75$ **(b)** $\varrho = 0.85$ **(c)** $\varrho = 0.95$

Figure 7.9: Optimal schedules for CVAR with $c_w = 1$, $c_o = 2$, $n = 7$, $X_k \sim U(1,5)$

Instead of the total costs, the scheduler can consider customer unpleasantness in terms of delays exceeding certain thresholds. A suitable service quality measure for this purpose is the delay unfairness measure (DUM), as introduced in [180]. Let v_k be the tolerance threshold for customer k. Then, the delay unfairness for client k is found by solving

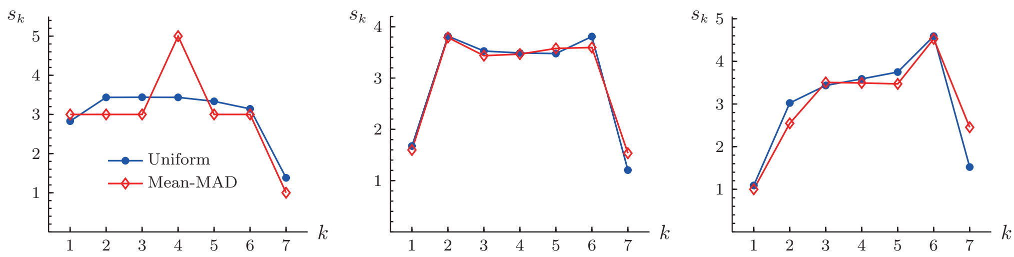
$$\inf\left\{1-\varrho \,\middle|\, \max_{\mathbb{P}\in\mathscr{P}_{(\mu,d)}} \text{CVaR}_{\varrho}[g_{k-1}(\mathbf{s},\mathbf{X})] \leqslant v_k,\ \rho\in[0,1)\right\}. \tag{7.22}$$

In this definition, CVaR denotes the worst-case expected delay conditioned on the $(1-\varrho)$ tail of the delay distribution. The quantity in (7.22) thus represents the smallest $(1-\varrho)$th quantile such that the worst-case average in the tail is not greater than the threshold level. Note that this definition also holds for the server overtime g_n. If the goal is to minimize the maximum delay unfairness (7.22), one solves

$$\begin{aligned} &\inf_{\varrho\in[0,1),\mathbf{s}\in\mathscr{S}} \ 1-\varrho \\ &\text{s.t. } \vartheta_k + \frac{1}{1-\varrho} \sup_{\mathbb{P}\in\mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}\left(g_{k-1}(\mathbf{s},\mathbf{X})-\vartheta_k\right)^+ \leqslant \nu_k, \quad k=1,\dots,n. \end{aligned} \tag{7.23}$$

This problem is nonlinear in ϱ. However, since the objective function and constraints are monotonic in ϱ, a simple bisection search will suffice to find the solution that minimizes $1-\varrho$. We note that the worst-case distribution is again the three-point distribution supported on a, μ and b. After substituting this distribution and fixing the value of ϱ, problem (7.23) reduces to an LP for which we only have to check feasibility.

In Figure 7.10 we provide the optimal schedules when minimizing the maximal delay unfairness. The parameter values are consistent with the setting of the CVaR example. Figure 7.10a shows the optimal schedule when minimizing total costs. In Figure 7.10b, we minimize the maximum DUM with a higher threshold value for the server overtime, and in Figure 7.10c, the threshold values are set to $\nu_k = 5$, $\forall k$. Since we consider another type of objective, the schedules exhibit a different kind of dome shape. Compared to the optimal schedule for cost minimization, when minimizing DUM, we allocate a much shorter slot to the first customer to reduce waiting times of later customers and the total server overtime.



(a) $\nu_k = \nu_{n+1} = \infty$ **(b)** $\nu_k = 3,\ \nu_{n+1} = 5$ **(c)** $\nu_k = \nu_{n+1} = 5$

Figure 7.10: Optimal time allocation based on DUM thresholds with $n = 7,\ T = 21,\ X_k \sim U(1,5)$

7.4.5. Dynamic appointment scheduling

This section considers a dynamic variant of the ASP in which the schedule can be adjusted for future customers at every arrival epoch. For simplicity, we assume that the required service time of a customer is sufficiently long so that at the arrival of customer k, the arrival time of customer $k+1$ can be scheduled. We next discuss an adjustable DRO approach in which we solve the dynamic ASP in a multistage fashion.

Before serving the kth customer, we observe the realizations $x_1,\dots,x_{k-1}$ of the random variables $X_1,\dots,X_{k-1}$. We can use this knowledge for choosing the optimal value of s_k by writing the appointment slot length as a function of the realizations: $s_k = s_k(x_1, \dots, x_{k-1})$. Since such functional dependencies typically give rise to NP-hard optimization problems, in robust optimization one often resorts to so-called linear decision rules (see, e.g., [18, 228]) of the form

$$s_k = z_{k0} + \sum_{i=1}^{k-1} z_{ki} x_i, \tag{7.24}$$

where z_{ki} becomes a decision variable in the new optimization problem. After substituting (7.24) into (7.5), the problem remains convex in the uncertain parameters. Hence, we can apply Lemma 7.2 to obtain the following linear program:

Proposition 7.11. *The optimization problem* $\min_{\mathbf{s}\in\mathcal{S}} \max_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}[f_n(\mathbf{s}, \mathbf{X})]$ *can be written as*

$$\begin{aligned}
\min_{\mathbf{s}\in\mathcal{S},\mathbf{z}} \quad & \sum_{\boldsymbol{\alpha}\in\{1,2,3\}^n} \Big(\prod_{i=1}^{n} p^{(i)}_{\alpha_i} \big(c_w \sum_{k=1}^{n-1} g^{(k)}_{\boldsymbol{\alpha}} + c_o g^{(n)}_{\boldsymbol{\alpha}} \big) \Big) \\
\text{s.t.} \quad & g^{(k)}_{\boldsymbol{\alpha}} \geqslant 0, \quad k = 1, \dots, n, \ \forall \boldsymbol{\alpha} \in \{1,2,3\}^n, \\
& g^{(k)}_{\boldsymbol{\alpha}} \geqslant \sum_{j=l}^{k} \left(\tau^{(j)}_{\alpha_j} - z_{j0} - \sum_{i=1}^{j-1} z_{ji} \tau^{(i)}_{\alpha_i} \right), \quad k = 1, \dots, n, \ l = 1, \dots, k, \ \forall \boldsymbol{\alpha} \in \{1,2,3\}^n.
\end{aligned} \tag{7.25}$$

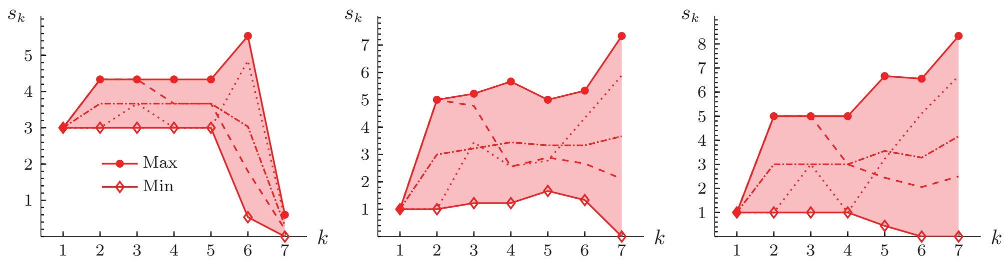
The linear program (7.25) is an approximation of the distributionally robust multistage problem and yields an upper bound since we only consider the subclass of linear functions for the decision s_k. Nevertheless, the solution of (7.25) still provides a robust schedule, which protects against distributional ambiguity, but additionally takes into account past realizations and thus makes adaptive decisions. To circumvent computational issues arising from the exponential number of constraints, we can again resort to SAA.

The model (7.25) is advantageous in practical settings where customers are already on standby, and thus benefit from regular schedule updates. Consider, for example, delivery services or surgery scheduling where the patient is already physically present in the hospital [147]. Another related line of research concerns settings in which, next to routine customers which are scheduled in advance, we have add-on customers that are fitted dynamically into the existing schedule as they call to request appointments, see [46, 76].

Figure 7.11 illustrates the adjustable robust schedules for $T = 21$, $n = 7$ and three different overtime costs. The figures show the range of values that the adjustable schedules assume and also display three possible traces of the schedule. For low overtime costs, the range of time allocation to the different customers remains narrow and almost no time is planned for the last customer. For higher values of c_o, the range of values widens and the slot length of the final customer strongly depends on the previous service time realizations.

7.4.6. Equidistant scheduling

As explained in Section 7.2, the ASP can be approximated by a version with infinitely many appointments and equidistant schedules with slots of fixed length. The ASP objective function can then be formulated in terms of the stationary waiting time W in a D/G/1 queue, which in



(a) $c_o = 1$ **(b)** $c_o = 10$ **(c)** $c_o = 20$

Figure 7.11: Optimal time allocation for $T = 21$, $c_w = 1$ and $n = 7$ with adjustable schedules based on the worst-case three-point distribution for $a = 1$, $b = 5$, $\mu = 3$ and $d = 1$

turn can be leveraged to determine the optimal slot length. Here W satisfies $W \stackrel{d}{=} \max\{W + X - s, 0\}$ with X the generic service time and s the slot length.

We assume that $s > \mu$, so that the long-run expected idle time per slot is $s - \mu$. The trade-off between stationary idle and waiting time then forms the ASP:

$$\min_{s \geqslant \mu} \ \omega(s - \mu) + (1 - \omega)\mathbb{E}[W]. \tag{7.26}$$

Here $\omega \in (0, 1)$ is a cost parameter. Using the infinite-series expression [202]

$$\mathbb{E}[W] = \sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E}[\max\{0, X_1 - s + \ldots + X_k - s\}], \tag{7.27}$$

one can write (7.26) in the form $\min_{s>\mu} \mathbb{E}_{\mathbb{P}}[f_\infty(s, \mathbf{X})]$ with the expectation taken over the distribution $\mathbb{P}$ of the service times $\mathbf{X} = (X_1, X_2, \ldots)$. Adopting the DRO method presented in this chapter, we then first determine tight bounds for $\mathbb{E}_{\mathbb{P}}[f_\infty(s, \mathbf{X})]$ under mean-MAD ambiguity of $\mathbf{X}$, and then solve for the optimal (robust) slot length:

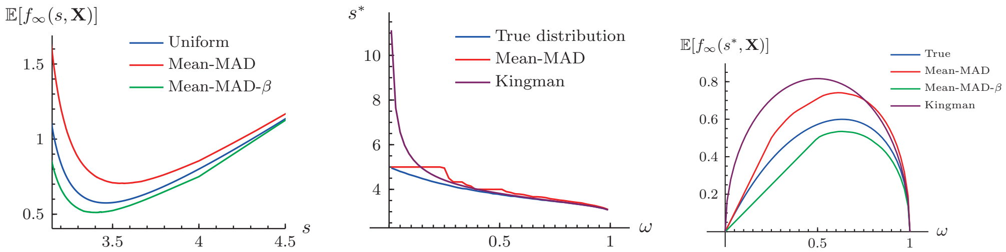
$$\min_{s \in (\mu, b]} \min_{\mathbb{P} \in \mathscr{P}_{(\mu,d,\beta)}} \mathbb{E}_{\mathbb{P}}[f_\infty(s, \mathbf{X})] \quad \text{and} \quad \min_{s \in (\mu, b]} \max_{\mathbb{P} \in \mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}[f_\infty(s, \mathbf{X})]. \tag{7.28}$$

Tight bounds for

$$\min_{\mathbb{P} \in \mathscr{P}_{(\mu,d,\beta)}} \mathbb{E}_{\mathbb{P}}[f_\infty(s, \mathbf{X})] \quad \text{and} \quad \max_{\mathbb{P} \in \mathscr{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}[f_\infty(s, \mathbf{X})] \tag{7.29}$$

do not follow directly from Lemma 7.4 and Lemma 7.2, because the function $f_\infty(s, \mathbf{X})$ can be viewed as a special case of $f_n(s, \mathbf{X})$ but then with $n \to \infty$. A limiting argument for $n \to \infty$ was provided in [210], where the distributions that attain the bounds in (7.29) were shown to correspond to the two-point and three-point distributions in Lemma 7.4 and Lemma 7.2, respectively. Figure 7.12a shows some examples of the tight lower and upper bounds (7.29), where we use uniformly $U(1, 5)$ distributed service times as a reference.

We next consider the optimal slot length, s^* say, where we also compare with a DRO approach in [138] based on mean-variance infinite-range ambiguity, which uses Kingman's bound



(a) $\omega = 3/4, U(1,5)$ **(b)** $U(1,5)$ **(c)** $U(1,5)$

Figure 7.12: Optimal stationary schedules and costs (reference values obtained by simulation)

$\mathbb{E}[W] \leqslant \sigma^2/(2(s-\mu))$ (tight for the D/G/1 queue, see [60]) and hence solves

$$\min_{s>\mu} \ \omega(s-\mu) + (1-\omega)\frac{\sigma^2}{2(s-\mu)}.$$

Figures 7.12b and 7.12c plot the optimal slot lengths and costs, respectively, for different values of ω. We use a bisection search procedure to find the optimal value s^* and solve the problems in (7.28). The objective function (7.26) is convex in s, and hence efficient numerical techniques can be applied. We also plot the objective function that follows from Kingman's bound for the matching mean-variance ambiguity set. Notice that the optimal mean-MAD slot length is close to the "true" optimum for all values of ω, while Kingman's bound suggests relatively large slot lengths when ω is small. This difference arises because the mean-MAD bound conditions on the range of service times, whereas Kingman's bound allows for unbounded support.

7.5. Conclusions and outlook

This chapter presents a novel distributionally robust approach for the ASP that, in contrast to existing robust methods, can incorporate explicitly independent service times in a min-max formulation. The key idea is to use MAD instead of variance as dispersion measure, so that the worst-case scenarios that maximize costs are formed by independent three-point distributions for the service times. These three-point distributions are independent of the chosen schedule and hence render the min-max problem amenable to standard stochastic programming techniques.

As explained in Section 7.1, existing DRO approaches for the ASP are confronted with formidable challenges when dealing with independent and identically distributed service times, due to the non-linearity of the stochastic program. Our novel approach beats these challenges by cleverly choosing the ambiguity set, resulting in exact yet easy-to-solve linear programs. Another way to circumvent the challenges is to relax the assumption of i.i.d. service times and allow for correlation and dependency structures, and using DRO approaches that optimize over joint probability measures instead of product measures. To the best of our knowledge, this is the first work that deals with independent random variables in a DRO setting for the ASP. Other DRO approaches allow either freely varying dependency structures or restricted linear dependence through conditioning on the covariance matrix, setting all covariance terms equal to zero; see [135, 171]. The latter approach, however, is only an approximation for the independent setting.

We have revealed several new structural properties of optimal robust schedules, including the "spiked patterns" in Figures 7.2-7.4, negligible early slots that resemble Bailey's rule, and optimal schedules that turn from dome-shaped to increasing when overtime costs become large. As shown, DRO approaches that allow for dependent service times prescribe entirely different schedules, such as decreasing patterns to counter several long service times that occur early during the day (due to positive correlation). A further comparison between correlated and independent settings is worth exploring in more detail. To measure the "price of correlations," the method in this chapter can serve as a benchmark for system performance under full independence of the driving random variables.

8

Some concluding thoughts

8.1. Discussion

Independence causes problems, an observation that will initially seem strange to most probabilists. After all, in probability theory, and certainly when it comes to sums of random variables or stochastic systems such as random walks and queues, the independence of variables is often a requirement for mathematical tractability. In contrast, if one does not exclude correlations, the analysis usually becomes much more challenging. This stands in stark contrast with the optimization point of view, at least when resolving extremal problems, to which the perspective is exactly the other way around: allowing for correlations leads to a relaxation that is mathematically simpler since modeling independence requires an infinite number of semi-infinite constraints.

Extremal stochastic models seek to establish upper and lower bounds for performance metrics using limited information available about the underlying random variables that drive the system, a renowned example being the eponymous bound derived by Kingman [132] for the expected delay in the GI/G/1 queue, which only uses mean-variance information about the service and interarrival times. However, a notable drawback of this bound is its inability to fully harness the knowledge that the underlying driving sequences consist of independent and identically distributed (i.i.d.) random variables. This limitation generally gives rise to a bound that is not as sharp as it could potentially be. In distributionally robust optimization (DRO), the issue regarding independence we address here is widely acknowledged. We adopt the terminology from DRO and in this chapter shall refer to the set of admissible probability distributions $\mathscr{P}$ as an ambiguity set. Many existing works that incorporate structural properties of distributions—

such as unimodality and symmetry—require the underlying ambiguity set to remain convex to employ the duality theory suited for solving such problems; see, for example, [177]. However, the presence of independence disrupts this convexity, resulting in mathematical intractability [100]. Moreover, we have not yet addressed the additional difficulties that arise from assuming identically distributed random variables. Fully capturing such a distributional characteristic requires an infinite number of moment constraints. Even within the stochastic programming literature, the arduous nature of addressing stochastic independence is widely acknowledged. The introduction of independence necessitates the computation of high-dimensional convolutions, which poses significant challenges. Indeed, even under the lenient assumption of independent uniform distributions governing the random variables, the simple class of linear two-stage stochastic programs with fixed recourse is deemed computationally intractable [72, 99]. Despite these additional complexities, it still seems of substantial value to add this structural information to the ambiguity set. Doing so reduces its size, resulting in tighter performance bounds. This parallels the scenario where the inclusion of the unimodality property in the ambiguity set leads to sharper bounds as it excludes the possibility of the worst-case distribution being a pathological discrete distribution supported on a limited set of points; see, e.g., [131]. As demonstrated in this chapter, the assumption of i.i.d. random variables does not eliminate the possibility of these discrete distributions being the extremal solutions. However, it does exclude overly conservative dependency structures.

As our motivating example, we again turn to the GI/G/1 queue. Let W_n be the waiting time of the nth customer. Further, let $X_n = V_n - U_n$ denote the difference between service and interarrival time. The waiting-time process then evolves according to the Lindley recursion

$$W_{n+1} = (W_n + X_n)^+, \ n \geqslant 0, \ W_0 = 0, \tag{8.1}$$

where $(x)^+ = \max\{x, 0\}$. By the i.i.d. assumption, (8.1) can equivalently be stated as

$$W_{n+1} \stackrel{d}{=} \max\{0, X_1, X_1 + X_2, \dots, X_1 + X_2 + \dots + X_n\} =: f_n(X_1, X_2, \dots, X_n),$$

where $\stackrel{d}{=}$ denotes equality in distribution. The sequence $\{W_n, n \geqslant 0\}$ can thus be expressed as a multivariate function $f_n(\cdot)$, which shall serve as the objective function for our extremal analysis. In the remainder, it is our aim to resolve

$$\max_{\mathbb{P}\in\mathcal{P}} \mathbb{E}[f_n(\mathbf{X})], \tag{8.2}$$

where $\mathcal{P}$ contains the available moment information and $\mathbf{X}$ is a random vector with elements $\{V_i, U_i\}_{i=1}^n$. Letting $n \to \infty$, this yields the classical extremal queue problem, as discussed in Chapter 2 and [50, 60, 222]. To obtain bounds for the stationary GI/G/1 waiting time, one considers the random variable $W := \lim_{n\to\infty} W_n$, which solves the stochastic fixed point equation

$$W \stackrel{d}{=} (W + V - U)^+. \tag{8.3}$$

Notice that for the expressions above to be valid, aside from stochastic independence, equality in distribution must hold throughout for all underlying random variables. This additional requirement, as aptly described by Kingman [133], introduces a challenging element of nonlinearity into the problem: "A still more difficult problem arises when all random variables considered are required to have the same distribution, since this requirement introduces an element of nonlinearity into the problem." To illustrate how this nonlinearity arises in the formulation of problem (8.2), let us consider the following example in two dimensions. Let $X_1,\ X_2$ denote two random variables both with support $\mathscr{X} = \{x_1, \dots, x_r\}$, and say we are trying to find a distribution $(p_1, \dots, p_r)$ on $\mathscr{X}$ that resolves $\max_{\mathbb{P}\in\mathscr{P}} \mathbb{E}[f_2(X_1, X_2)]$. This is tantamount to solving

$$\begin{aligned} \max \sum_{k=1}^{r}\sum_{l=1}^{r} & \max\{0, x_k, x_k + x_l\} p_k p_l \\ \text{s.t.} \sum_{k=1}^{r} & h_j(x_k) p_k = q_j, && j = 0, \dots, m, \\ & p_k \geqslant 0, && k = 1, \dots, r, \end{aligned} \tag{8.4}$$

where h_j and q_j, $\forall j$, capture the available moment information. The nonlinearity arising from the objective function in (8.4) becomes evident upon closer examination. It is apparent that the objective function now takes a quadratic form in the vector $(p_1, \dots, p_r)$. The computational complexity naturally increases as more random variables are introduced, exacerbating the computational intractability of the problem. Furthermore, if the support is not assumed to be finite, this adds another layer of complexity to the extremal analysis.

In this concluding chapter, our objective is to address extremal problems with i.i.d. random variables using simple yet effective solutions tailored to specific cases. We delve deeper into the recursive argument that was used to resolve the extremal queue problem in Chapter 2. Then, we show that the insensitivity property, emphasized throughout this thesis, can be extended to encompass various other types of distributional information. To facilitate its application, we present concise guidelines outlining the process of eliciting this property. Finally, we discuss two more extremal models featuring i.i.d. driving sequences, which can be effectively solved by leveraging the recursive argument and harnessing the insensitivity property.

8.2. A simple trick

An excellent exposition on extremal models for independent random variables is provided by Kingman [132], who discusses a two-dimensional example. Consider two independent random variables $X \in \mathscr{X},\ Y \in \mathscr{Y}$, where $\mathscr{X}, \mathscr{Y}$ denote their supports, with ambiguity sets $\mathscr{P}_X$ and $\mathscr{P}_Y$, respectively, and define

$$\phi_Y(x) := \mathbb{E}[f(x, Y)],\ x \in \mathscr{X}.$$

Then $\mathbb{E}[f(X, Y)] = \mathbb{E}[\phi_Y(X)]$ by the law of total expectation. Observe that ϕ_Y is $\mathscr{X}$-measurable by Fubini's theorem. Suppose that there exists a measure that satisfies the moment conditions. Then using the Richter-Rogosinski theorem with f replaced by the function ϕ_Y, we can establish the existence of an extremal random variable X^* with at most $m + 1$ support points. It follows that

$$\mathbb{E}[\phi_Y(X)] \leqslant \mathbb{E}[\phi_Y(X^*)] \implies \mathbb{E}[f(X, Y)] \leqslant \mathbb{E}[f(X^*, Y)].$$

Applying the same argument to $\phi_{X^*}(y) := \mathbb{E}[f(X^*, y)]$, it can be shown that there exists a Y^* such that

$$\mathbb{E}[f(X^*, Y)] \leqslant \mathbb{E}[f(X^*, Y^*)],$$

and so

$$\mathbb{E}[f(X, Y)] \leqslant \mathbb{E}[f(X^*, Y^*)].$$

So far, this result relies predominantly on the Richter-Rogosinski theorem [186]. It is natural to try and extend this result to the setting with an arbitrary number of random variables. In particular, consider the problem of finding bounds for

$$\mathbb{E}[f_n(X_1, X_2, \dots, X_n)],$$

where $X_1, X_2, \dots, X_n$ are independent random variables satisfying the moment conditions. By applying the above argument sequentially, it turns out that it is possible to carry over the two-dimensional result to the setting with an arbitrary number of independent random variables. Thus, we can restrict our attention, with respect to all the random variables being considered, to those that assume only a finite number of values. This results in the following theorem, which generalizes results discussed in [114, 133].

Theorem 8.1 **(Richter-Rogosinski tower rule).** *Suppose that there exist distributions for X_1, $X_2, \dots, X_n$ that all satisfy m moment conditions. Then the supremum of $\mathbb{E}[f_n(X_1, X_2, \dots, X_n)]$, where $X_1, X_2, \dots, X_n$ are independent, is unaltered if the respective ambiguity sets are restricted to distributions with at most $m+1$ support points.*

Unfortunately, Theorem 8.1 does not necessarily hold for identically distributed random variables (see, e.g., [133]). Despite this limitation, this theorem offers a framework for establishing sharp bounds for $\mathbb{E}[f_n(X_1, X_2, \dots, X_n)]$, even if $X_1, X_2, \dots, X_n$ are i.i.d. random variables. If we relax the assumption of identical distributions, it is sufficient to solve

$$\begin{aligned} \max_{\mathbb{P}_\otimes, \{\mathbb{P}_i\}_{i=1}^n} \ & \mathbb{E}_{\mathbb{P}_\otimes}\left[f_n(X_1, X_2, \dots, X_n)\right] \\ \text{s.t. } & \mathbb{P}_\otimes = \mathbb{P}_1 \otimes \mathbb{P}_2 \otimes \cdots \otimes \mathbb{P}_n, \\ & \mathbb{E}_{\mathbb{P}_i}[h_j(X_i)] = q_j, \quad i = 1, 2, \dots, n, \ j = 1, \dots, m, \end{aligned} \tag{8.5}$$

where the probability measure $\mathbb{P}_\otimes$ is written in this particular form to impose independence among the random variables $X_1, \dots, X_n$. To deal with problem (8.5), we adopt the approach outlined in Chapter 2 and sequentially optimize over the n random variables. This way, we can express the multivariate problem as

$$\max_{\mathbb{P}_n \in \mathcal{P}} \cdots \max_{\mathbb{P}_2 \in \mathcal{P}} \max_{\mathbb{P}_1 \in \mathcal{P}} \mathbb{E}_{\mathbb{P}_\otimes}[f_n(X_1, X_2, \dots, X_n)]. \tag{8.6}$$

It is important to note here that, due to the identicality of the random variables, the optimization is taken over the same ambiguity set $\mathcal{P}$ in each stage. Suppose we first consider the maximization with respect to X_1, then we need to consider as objective function

$$\phi(x_1) := \mathbb{E}[f_n(x_1, X_2, \dots, X_n)],$$

where the expectation $\mathbb{E}$ is taken with respect to the other random variables $X_2, \dots, X_n$. From Theorem 8.1, it readily follows that the extremal distributions have at most $m+1$ mass points. In order to find a candidate distribution for $\mathbb{E}[\phi(X_1)]$, we can try to exploit general properties such as curvature, as demonstrated in the proof of Theorem 2.1, in which convexity of $\phi(x_1)$ was leveraged. If an extremal distribution can be established which is fully characterized by the distributional information within the ambiguity set, such that it relies solely on the curvature properties of $\phi(x_1)$ and no longer depends on the specific characteristics of f_n—and, by extension, the other random variables—this extremal distribution constitutes the candidate to solve (8.5) and governs the extremal random variable, denoted as X^*. Then, by taking the expectation with respect to X^*, we obtain

$$\max_{\mathbb{P}_n \in \mathcal{P}^n_{(\mu,d)}} \cdots \max_{\mathbb{P}_2 \in \mathcal{P}^2_{(\mu,d)}} \mathbb{E}_{\mathbb{P}_\otimes}[f_n(X^*, X_2, \dots, X_n)]. \tag{8.7}$$

Since the expectation operator preserves properties such as monotonicity, convexity, etc., the worst-case expectation retains those properties as function of $x_2, \dots, x_n$. As a result, we can substitute X^* sequentially n times, and, since all the ambiguity sets are equivalent, the extremal $X_1, X_2, \dots, X_n$ are i.i.d. as X^*, hence resolving the original extremal problem (8.6) with identically distributed random variables. As a result, this recursive argument can solve both problems with independent and i.i.d. random variables.

Of course, the main challenge is finding such an extremal distribution that solves $\phi(X_1)$, which pertains to the notion of "insensitivity" that we have introduced earlier in this thesis. Distribution-free analysis, and distributionally robust optimization, for that matter, greatly benefit from this insensitivity property as it simplifies the mathematical analysis significantly. As a matter of fact, the recursive argument presented in this section relies heavily on this property. Without it, the argument would be considerably more complex, perhaps even impossible. While it is important to note that this simple trick based on the tower rule is certainly not a panacea for all extremal problems with i.i.d. driving sequences, it proves potentially valuable for particular models. Therefore, we complement this recursive argument with a set of guidelines that enables us to identify cases for which this simple trick can lead to solvable instances of the extremal problem with i.i.d. random variables. In the next section, we present several combinations of distributional information and objective functions that induce the desired insensitivity.

8.3. Guidelines for inducing insensitivity

We already observed in Chapter 3 that we can restrict our attention to two-point distributions in the search for the extremal distribution that attains the tight bound $\mathbb{E}_{\mathbb{P}}[\phi(X)]$ with $\mathbb{P} \in \mathcal{P}_{(\mu,\sigma)}$. We next prove a result that is similar to Lemma 3.2, but which requires only existence and convexity of the derivative ϕ'. The proof of this result elucidates why this setting with variance information attains similar favorable properties as those that we encountered for mean-MAD information in the previous chapters, albeit for a different class of functions. We prove the insensitivity result by following a series of steps: First, we determine a candidate primal solution by leveraging the complementary slackness property to make an educated guess. Then, we demonstrate that our guess remains primal feasible across all possible values of the distributional parameters. Subsequently, we establish the feasibility of the resulting dual solution. Finally, we demonstrate the optimality of these solutions using weak duality.

Proposition 8.2 **(Insensitive mean-variance solution).** *Consider a function $x \mapsto \phi(x)$ that is continuously differentiable on $[a,b]$, with its derivative denoted as ϕ', and a random variable $X \sim \mathbb{P} \in \mathcal{P}_{(\mu,\sigma)}$.*
(i) *If $\phi'(x)$ is convex on $[a,b]$, the extremal distribution that solves $\max_{\mathbb{P}\in\mathcal{P}_{(\mu,\sigma)}} \mathbb{E}_{\mathbb{P}}[\phi(X)]$ is a two-point distribution on the values $\left\{\mu - \frac{\sigma^2}{b-\mu}, b\right\}$ with respective probabilities $\left\{\frac{(b-\mu)^2}{(b-\mu)^2+\sigma^2}, \frac{\sigma^2}{(b-\mu)^2+\sigma^2}\right\}$.*
(ii) *If $\phi'(x)$ is concave on $[a,b]$, the extremal distribution that solves $\max_{\mathbb{P}\in\mathcal{P}_{(\mu,\sigma)}} \mathbb{E}_{\mathbb{P}}[\phi(X)]$ is a two-point distribution on the values $\left\{a, \mu + \frac{\sigma^2}{\mu-a}\right\}$ with respective probabilities $\left\{\frac{\sigma^2}{(\mu-a)^2+\sigma^2}, \frac{(\mu-a)^2}{(\mu-a)^2+\sigma^2}\right\}$.*

Proof. Consider a generic function $\phi(x)$ with a convex derivative $\phi'(x)$. For $\mathbb{P} \in \mathcal{P}_{(\mu,\sigma)}$, we must solve

$$\begin{aligned} \max_{\mathbb{P}(x)\geqslant 0} \quad & \int_a^b \phi(x)\mathrm{d}\mathbb{P}(x) \\ \text{s.t.} \quad & \int_a^b \mathrm{d}\mathbb{P}(x) = 1, \; \int_a^b x\,\mathrm{d}\mathbb{P}(x) = \mu, \; \int_a^b x^2\mathrm{d}\mathbb{P}(x) = \sigma^2 + \mu^2, \end{aligned} \tag{8.8}$$

which has the following dual:

$$\begin{aligned} \min_{\lambda_0,\lambda_1,\lambda_2} \quad & \lambda_0 + \lambda_1\mu + \lambda_2(\sigma^2 + \mu^2) \\ \text{s.t.} \quad & M(x) := \lambda_0 + \lambda_1 x + \lambda_2 x^2 \geqslant \phi(x), \; \forall x \in [a,b]. \end{aligned} \tag{8.9}$$

We seek a quadratic function that majorizes the objective function $\phi(x)$. Under the assumption that strong duality holds, we can provide a constructive argument to find the extremal distribution with help from the dual problem. Consequently, we can use the complementary slackness property to make an educated guess on the support of the extremal distribution. The candidate solution $M(x)$ that we will consider touches $\phi(x)$ at exactly two points, i.e., some $x_0 < \mu$ and the upper bound of the support b, see Figure 8.1. First we show that for each feasible pair of the mean μ and variance σ parameters, the two-point distribution with support $\{x_0, b\}$ is a member of the ambiguity set. Solving for x_0, p_{x_0} and p_b, using the moment constraints in (8.8), yields the distribution stated in the lemma. Notice that

$$p_{x_0} = \frac{(b-\mu)^2}{(b-\mu)^2 + \sigma^2},$$

from which it is apparent that $p_{x_0}, p_b \in [0,1]$. We also have $x_0 < \mu$ since $\sigma > 0$ and $b > \mu$. Finally, we check if $x_0 \geqslant a$. Obviously, this holds if $a = -\infty$. To show $x_0 \geqslant a$ also for $a > -\infty$, recall that

$$\sigma^2 \leqslant (b-\mu)(\mu - a).$$

Hence, using this bound, it follows that

$$x_0 = \mu - \frac{\sigma^2}{b-\mu} \geqslant \mu - \frac{(b-\mu)(\mu-a)}{b-\mu} = a.$$

As a consequence, the extremal distribution stated in Lemma 8.2 always constitutes a feasible primal solution for each μ, σ pair. Next the goal is to construct a dual solution $(\lambda_0, \lambda_1, \lambda_2)$ that achieves the same objective value as this primal solution, ensuring by weak duality that this primal-dual solution pair is optimal. To obtain the quadratic function illustrated in Figure 8.1, the dual variables should satisfy the following conditions:

$$\lambda_0 + \lambda_1 x_0 + \lambda_2 x_0^2 = \phi(x_0), \quad \lambda_0 + \lambda_1 b + \lambda_2 b^2 = \phi(b), \quad \lambda_1 + 2\lambda_2 x_0 = \phi'(x_0).$$

Solving for $\lambda_0, \lambda_1, \lambda_2$, substituting the expression for x_0 and plugging these values into the dual objective function yields, after some algebra,

$$\lambda_0 + \lambda_1 \mu + \lambda_2(\mu^2 + \sigma^2) = \frac{(b-\mu)^2}{(b-\mu)^2 + \sigma^2} \phi\left(\mu - \frac{\sigma^2}{b-\mu}\right) + \frac{\sigma^2}{(b-\mu)^2 + \sigma^2} \phi(b),$$

which agrees with the primal objective value. Hence, strong duality holds if this choice for λ_0, λ_1 and λ_2 is feasible. To complete the proof, it thus remains to establish feasibility of this dual solution. Denote by M_{x_0} the dual function that touches ϕ at x_0, and intersects at b.

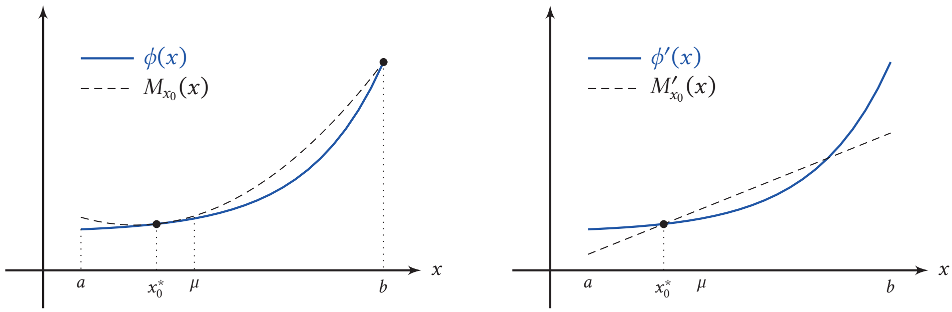
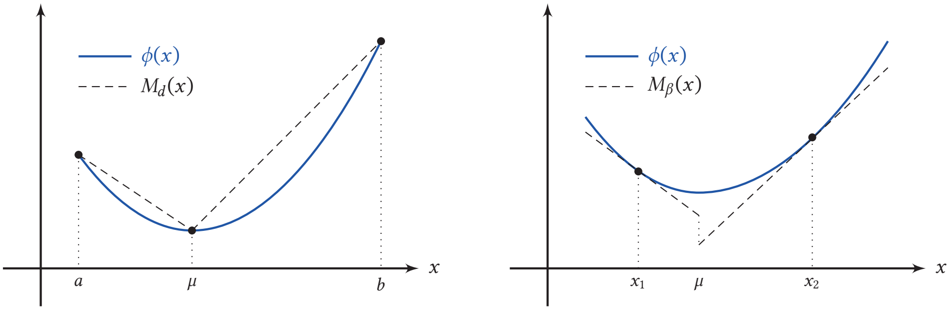


Figure 8.1: Objective function $\phi(x)$, dual function $M(x)$ and their derivatives

Now, by looking at the derivatives of the majorizing function $M(x)$ and the objective function $\phi(x)$, we can verify that dual feasibility indeed holds. To see this, notice that the derivatives can only intersect twice because ϕ' is convex and M' is linear. The first intersection point corresponds to a tangent point for the original functions. Due to dual feasibility, $M'(x)$ has to be greater than ϕ' directly after the intersection point and $\phi(x)$ can only coincide with $M(x)$ once, except for a possible second time at b. This then yields assertion (i). For assertion (ii), note that for concave $\phi'(x)$ the line of reasoning is almost identical, but instead considering the extremal distribution with support $\{a, x_0\}$ and the dual function that corresponds to this solution. □

The benefits of this result are readily apparent. If a function is a member of this class of functions with a convex derivative, then intuitive guesses and trial solutions for the primal solution and dual function are unnecessary, as we immediately obtain an extremal distribution that exhibits the insensitivity property.

It is quite instructive to compare the upper and lower bounds with mean-variance information to those with mean-MAD(-β) information. For this purpose, recall the dual functions corresponding to the mean-MAD(-β) ambiguity sets, as displayed in Figures 8.2a and 8.2b.



(a) The majorant $M_d(x)$ **(b)** The minorant $M_\beta(x)$

Figure 8.2: Some convex function $\phi(x)$ and the dual functions for mean-MAD(-β) information

Another natural guess for the primal solution in the mean-MAD setting is a two-point distribution with one support point fixed to either a or b and a second support point x_0 varying in the interval (μ, b) or (a, μ), respectively. This leads to three unknowns (two probabilities and the value of x_0), which can be determined using the three moment equations. However, it becomes clear from Figure 8.2a that several issues arise when turning to the dual problem. Because $\phi(x)$ is convex and M_d is piecewise affine, there does not exist a matching dual function for which $M_d(x)$ coincides with $\phi(x)$ on a point in the interior of $[a, b]$, except for $x = \mu$. Hence, the only alternative that yields a consistent solution is the three-point distribution with support $\{a, \mu, b\}$, which follows from solving the moment conditions

$$p_1 + p_2 + p_3 = 1, \quad p_1 a + p_2 \mu + p_3 b = \mu, \quad p_1|a - \mu| + p_3|b - \mu| = d,$$

yielding the extremal distribution that is ubiquitous in this thesis.

In contrast, a similar guess proves effective for the mean-MAD-β lower bound. Consider a two-point distribution characterized by two unknown support values x_1, x_2 and two unknown probabilities p_1, p_2. With four moment constraints in place, we can obtain a unique extremal distribution as the solution to the following system of equations:

$$p_1 + p_2 = 1, \quad p_1 x_1 + p_2 x_2 = \mu, \quad p_1|x_1 - \mu| + p_2|x_2 - \mu| = d, \quad p_2 = \beta.$$

Upon examining Figure 8.2b, it becomes apparent that this two-point distribution admits a feasible dual function, M_β. Therefore, the resulting solution serves as a viable candidate for the extremal distribution, meeting the requirements in both the primal and dual sense.

We proceed to establish a bound that closely resembles the Edmundson-Madanski (E-M) bound [146], but for which a weaker condition than convexity, called starshapeness, is sufficient. A function $\phi : \mathbb{R}_+ \mapsto \mathbb{R}$, with $\mathbb{R}_+$ denoting the nonnegative real numbers, is said to be starshaped if $\phi(x)/x$ is a nondecreasing function of $x \in \mathbb{R}_+$. However, it is usually easier to apply the following, more intuitive, definition: A function is starshaped if, and only if, it has the property that when one draws its graph and places a star at the origin, then this star illuminates from the origin every point of the upper part of the curve [155], as illustrated in Figure 8.3.

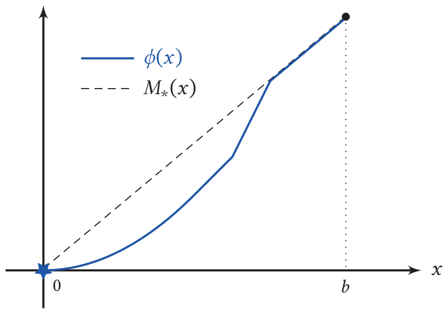


Figure 8.3: A starshaped function and its linear majorant

Let us consider, in conjunction with the class of starshaped functions, the ambiguity set

$$\mathscr{P}_{(\mu,b)} = \{\mathbb{P} \ : \ \text{supp}(X) \subseteq [0,b], \ \mathbb{E}_{\mathbb{P}}[X] = \mu\}.$$

From Figure 8.3, it follows that the next result holds.

Proposition 8.3 **(E-M bound for starshaped functions).** *Let X denote a random variable of which its distribution $\mathbb{P} \in \mathscr{P}_{(\mu,b)}$. Consider a function $x \mapsto \phi(x)$ for which $\phi(x)/x$ is nondecreasing on $[0,b]$. Then, $\max_{\mathbb{P}\in\mathscr{P}_{(\mu,b)}} \mathbb{E}_{\mathbb{P}}[\phi(X)]$ is achieved by a two-point distribution with support $\{0,b\}$ and respective probabilities $\{1-\mu/b, \mu/b\}$.*

Proof. Define $M_*(x) := \lambda_0 + \lambda_1 x$, which is the dual function corresponding to $\mathscr{P}_{(\mu,b)}$. Then, from the definition of starshapeness, it follows that the dual solution $\lambda_0 = \phi(0)$, $\lambda_1 = (\phi(b) - \phi(0))/b$ is feasible. Moreover, by complementary slackness, the corresponding primal solution places probability mass on the support $\{0,b\}$. It is then a straightforward exercise to show that the primal and dual objective values agree, and hence, by weak duality, this is the optimal primal-dual solution pair. □

We shall see later that starshapeness becomes useful for distribution-free analysis of the GI/G/c queue.

8.4. Tractable extremal models

8.4.1. Bounds for higher-order cumulants of GI/G/1 queue

We analyze the mean-variance extremal queue with a single server and unlimited waiting space in which customers are served based on the order of arrival. This system is driven by two independent sequences of i.i.d. distributed random variables, which represent the service times

$\{V_n\}$ and the interarrival times $\{U_n\}$ of customers, distributed as V and U, respectively. The service time of the nth customer is denoted by V_n, and the interarrival time between the arrivals of customers n and $n+1$ is represented by U_n. We assume that the first customer arrives precisely at time 0. It is further assumed that the mean service time $\mathbb{E}[V] = \mu_V$ and the mean interarrival time $\mathbb{E}[U] = \mu_U$ exist (and are finite), and that both V and U have a finite variance σ_V^2 and σ_U^2 respectively. To ensure stability, we assume that $\mu_V < \mu_U$. The waiting time of the nth customer, denoted by W_n, is defined as the time the customer spends in the queue before being served. The sequence $\{W_n, n \geqslant 0\}$ represents a Lindley process. We define the Lindley process through the recursion

$$W_{n+1} = (W_n + V_n - U_n)^+, \ n \geqslant 0,$$

in which $(x)^+ = \max\{x, 0\}$ and $W_0 = 0$. We study the steady-state waiting time, denoted by W, of which the distribution satisfies the stochastic fixed-point equation

$$W \stackrel{d}{=} (W + V - U)^+,$$

with $\stackrel{d}{=}$ denoting equality in distribution. Define $X_n := V_n - U_n$ and $S_n := X_1 + \cdots + X_n$, for $n \geqslant 1$, so that $W_n \stackrel{d}{=} \max\{S_k : 1 \leqslant k \leqslant n\}$—in other words, the waiting time process $\{W_n\}$ is generated by the random walk $\{S_n\}$. Under the finite moment conditions, W_n is finite almost surely with a finite mean given by

$$\mathbb{E}[W_n] = \sum_{k=1}^{n} \frac{1}{k}\mathbb{E}[S_k^+], \ \forall n \geqslant 0$$

and, assuming a stable system, W is a proper random variable with mean

$$\mathbb{E}[W] = \sum_{k=1}^{\infty} \frac{1}{k}\mathbb{E}[S_k^+] < \infty.$$

Let $c_m(W)$ denote the mth cumulant of the steady-state waiting time, which one can write as

$$c_m(W) = \sum_{k=1}^{\infty} \frac{1}{k}\mathbb{E}[(S_k^+)^m]. \tag{8.10}$$

Bounds for the higher cumulants of the steady-state waiting time were already obtained in [24] using stochastic comparison techniques. Let the random variables $V_{(2)}$ and $U_{(2)}$ follow the extremal two-point distributions stated in assertions (i) and (ii) of Lemma 8.2, respectively. In order to apply the bounds in Lemma 8.2, it is necessary to show that expression (8.10) can be written in terms of an objective function with a convex derivative. As a building block for expression (8.10), define an objective function as follows:

$$\phi(x) := \int_{a_Y}^{b_Y} h\left((y-x)^+\right) \mathrm{d}F(y) = h(0)F(x) + \int_{x}^{b_Y} h(y-x)\mathrm{d}F(y), \quad a_U \leqslant x \leqslant b_U, \tag{8.11}$$

where $h(x) = x^m$ with $m \geqslant 2$ being an integer, and F is the cumulative distribution function of some auxiliary real-valued random variable Y (which henceforth will correspond to V_1, $V_1 +$

$V_2 - U_2$, etc.). Since $h(0) = 0$, it is possible to reduce the Riemann-Stieltjes integral (8.11) further to

$$\phi(x) = \int_x^{b_Y} h(y-x)\mathrm{d}F(y), \quad a_U \leqslant x \leqslant b_U. \tag{8.12}$$

In Appendix B.5, we show that $\phi(x)$ is convex and continuously differentiable, with a concave derivative $\phi'(x)$, so that Lemma 8.2 can be employed.

LEMMA 8.4. *Consider a real-valued random variable Y with cumulative distribution function F on support $[a_Y, b_Y]$ and a finite moment-generating function. Then, the objective function $\phi(x)$, as defined in* (8.12)*, with $h(x) = x^m$, $m \geqslant 2$, is continuously differentiable on $[a_U, b_U]$. Moreover, its derivative, $\phi'(x)$, is concave on $[a_U, b_U]$.*

Note that we did not make any assumptions on the probability distribution F. Chen and Whitt [51], however, needed to derive sufficient conditions for the distribution function F so that their objective function yields a Chebyshev system on $[0, b]$, entailing that the extremal distributions are the two-point distributions stated in Lemma 8.2. However, in order to demonstrate this, they must make ancillary assumptions, such as the existence of a smooth probability density function and additional constraints on the support of Y (i.e., $a_Y \leqslant 0 < b \leqslant b_Y$), all of which turn out to be superfluous, as we show next. For the higher-order cumulants of the steady-state waiting time, we demonstrate the following result, which is a stronger assertion than Theorem 1 in [51].

THEOREM 8.5 **(Higher-order cumulants of steady-state waiting time).** *Consider the GI/G/1 queue with generic service time V whose distribution lies in the ambiguity set $\mathscr{P}_{(\mu_V,\sigma_V)}$ and generic interarrival time U whose distribution lies in the ambiguity set $\mathscr{P}_{(\mu_U,\sigma_U)}$. Consider the tight upper bounds for all cumulants $m \geqslant 2$ of the steady-state waiting time W.*

(i) *For given service time V, the tight upper bound follows from $U_{(2)}$.*

(ii) *For given interarrival time U, the tight upper bound follows from $V_{(2)}$.*

(iii) *The overall tight upper bound follows from $V_{(2)}$ and $U_{(2)}$.*

Proof. Let us provide a proof sketch for part (i) of the theorem, then (ii) and (iii) follow from a similar line of reasoning. According to Lemma 8.2, objective functions of the form (8.12) attain their maximal expected value under the extremal random variable $U_{(2)}$. To establish results for the steady-state waiting time, we first consider a truncated version of the infinite series expression (8.10):

$$f_{n,m}(u_1,\dots,u_n) = \sum_{k=1}^{n} \frac{1}{k}\mathbb{E}[(\max\{0, V_1 - u_1 + \dots + V_k - u_k\})^m]. \tag{8.13}$$

Observe that this expression has the required functional form. We investigate the case $m = 2$, i.e. the variance, as the argument is identical for higher cumulants. Consider the moment problem in U_1, thus fixing the distributions of $U_2, \dots, U_n, V_1, \dots, V_n$,

$$\max_{\mathbb{P}\in\mathscr{P}_{(\mu_U,\sigma_U)}} \int_{a_U}^{b_U} \left(\mathbb{E}[((V_1 - u_1)^+)^2] + \dots + \frac{1}{n}\mathbb{E}[((V_1 - u_1 + \dots + V_n - U_n)^+)^2] \right) \mathrm{d}\mathbb{P}(u_1), \tag{8.14}$$

where the expectations $\mathbb{E}$ are taken with respect to the distributions of $V_1, \dots, V_n, U_2, \dots, U_n$. Since the expectations in (8.14) are functions of the form $f(u_1)$, with $h(x) = x^2$ and Y representing $V_1, V_1 + V_2 - U_2$, etc., Lemma 8.2 holds for each of these terms. Since the two-point extremal distribution is independent of $u_2, \dots, u_n$, we can apply the univariate result recursively. The proof is then completed by a limit argument, where we approach the infinite series by letting $n \to \infty$. Consider the function

$$f_{n,m}(u_1, \dots, u_n) = \sum_{k=1}^{n} \frac{1}{k}\mathbb{E}[(\max\{0, V_1 - u_1 + \dots + V_k - u_k\})^m]$$

Then, for i.i.d. interarrival times distributed as U,

$$\max_{\mathbb{P}\in\mathscr{P}_{(\mu_U,\sigma_U)}} \mathbb{E}[f_{n,m}(U_1, \dots, U_n)]$$

is solved by the extremal distribution $U_{(2)}$. This yields the bounds

$$c_n := \sum_{k=1}^{n} \frac{1}{k}\mathbb{E}[(\max\{0, V_1 - U_1 + \dots + V_k - U_k\})^m] \leqslant \mathbb{E}_{\mathbb{P}}[f_{n,m}(U_1^*, \dots, U_n^*)] =: u_n$$

with $U_1^*, U_2^*, \dots$ i.i.d. as $U_{(2)}$. Since the sequences $\{c_n\}$ and $\{u_n\}$ both converge to well-defined limits, the result follows. For (ii), we prove that $\phi(v_1)$ has a convex derivative. For (iii), we combine (i) and (ii). □

8.4.2. Extremal GI/G/c queue with mean-support information

Consider a multi-server GI/G/c queueing system where the service time of the nth customer is denoted by V_n, and the interarrival time between the arrivals of customers n and $n + 1$ is represented by U_n. That is, the system is driven by the same two sequences as the GI/G/1 model in the previous subsection. Customers are served according to a first-come-first-served policy. The waiting time W_n of the nth customer is given by the remaining workload assigned to the first server at the time of the nth arrival, $W_n^{(1)}$. For a vector $\mathbf{x}$, let $x^{(i)}$ denote the ith smallest component, so that $x^{(1)} \leqslant x^{(2)} \leqslant \dots \leqslant x^{(c)}$. Let $(W_n^{(1)}, W_n^{(2)}, \dots, W_n^{(c)})$ denote the componentwise increasing vector of assigned workloads at the c servers as seen by the nth arrival. Let $\Phi(\mathbf{x}) = (x^{(1)}, x^{(2)}, \dots, x^{(c)})$ denote the operator that sorts the vector $\mathbf{x}$ in this increasing order. Then for $n = 1, 2, \dots$, the evolution of the workloads in the GI/G/c queueing system is described by the Kiefer-Wolfowitz recursion

$$(W_{n+1}^{(1)}, W_{n+1}^{(2)}, \dots, W_{n+1}^{(c)}) = \Phi\left((W_n^{(1)} + V_n - U_n)^+, (W_n^{(2)} - U_n)^+, \dots, (W_n^{(c)} - U_n)^+\right),$$

For $c = 1$, this expression reduces to Lindley's recursion

$$W_{n+1} = (W_n + V_n - U_n)^+, \quad n \geqslant 0.$$

By a simple induction argument, it can be shown that W_{n+1} is convex in the service-time parameter. However, Weber [220] provided counterexamples that show the GI/G/c queue no longer possesses this convexity property, so we need to find a weaker property. It is obvious that the waiting times are nondecreasing in the service times. Hence, it seems natural to try to show that a property stronger than monotonicity, but weaker than convexity, holds for the workloads in the GI/G/c queue as functions of the service times.

A property that satisfies this characterization is starshapeness of a nonnegative function, as discussed in the previous section. Since the pointwise maximum, the pointwise minimum and the sum of any two monotone functions is again monotone, the following result can be easily verified [194, Lemma 2.4].

Lemma 8.6. *Consider two starshaped functions* $\phi_i : \mathbb{R}_+ \to \mathbb{R}, i = 1, 2$. *Then* $\phi_{[1]}, \phi_{[2]}$ *and* ψ *defined by* $\phi_{[1]}(x) = \min\{\phi_1(x), \phi_2(x)\},\ \forall x \in \mathbb{R}_+,\ \phi_{[2]}(x) = \max\{\phi_1(x), \phi_2(x)\},\ \forall x \in \mathbb{R}_+$, *and* $\psi(x) = \phi_1(x) + \phi_2(x),\ \forall x \in \mathbb{R}_+$, *are all starshaped functions.*

Then it can be shown that the sorting operator Φ preserves starshapeness by repeated use of the previous result [194, Lemma 2.5].

Lemma 8.7. *Let* $\Phi : \mathbb{R}^c \to \mathbb{R}^c$ *denote a function that sorts the coordinates of the input vector into increasing order. If* $\phi_i,\ i = 1, 2, \dots, c$, *are all starshaped functions, then* $\psi_i, i = 1, 2, \cdots, c$, *defined by* $(\psi_1(x), \dots, \psi_c(x)) = \Phi(\phi_1(x), \dots, \phi_c(x)), x \in \mathbb{R}_+$, *are all starshaped functions.*

Close inspection of the Kiefer-Wolfowitz recursion reveals that all operations governing the workload dynamics of the GI/G/c queue maintain the starshapeness property. Then, since the expectation operator also preserves starshapeness, we can readily derive an extremal result for the GI/G/c queue with mean-support information by combining Proposition 8.3 with the recursive argument.

Theorem 8.8 **(Tight bounds for GI/G/c waiting time).** *Consider the GI/G/c queue with generic service time* V *whose distribution lies in the ambiguity set* $\mathcal{P}_{(\mu_V, b_V)}$. *For given interarrival time* U, *the tight upper bounds for the transient mean waiting time* $\mathbb{E}[W_n]$ *and expected steady-state waiting time* $\mathbb{E}[W]$ *are achieved by a two-point distribution for* V *with support* $\{0, b_V\}$ *and respective probabilities* $\{1 - \mu_V/b_V, \mu_V/b_V\}$.

8.5. Outlook

The insensitive extremal distributions discussed in this chapter can be broadly classified into two categories. The first category pertains to the mean-MAD approach, where the support is fixed to the maximum of $m+1$ point masses, in agreement with the Richter-Rogosinski theorem [186]. Consequently, the corresponding probabilities are obtained by solving the moment conditions. Alternatively, as with mean-variance information, properties of the objective function fix some of the mass points, enabling the determination of the remaining points and their corresponding probabilities through the moment conditions. This second category also includes, for example, the mean-MAD-β information set combined with a convex objective function.

It further includes the ambiguity set that incorporates support, mean, variance and skewness information, along with an objective function featuring a convex second derivative; see, e.g., [128].

Unfortunately, striving for a fully general solution to the extremal problem involving i.i.d. random variables necessitates referring back to the works of Hoeffding [114] and Kingman [133]. As these authors have adeptly articulated, the extremal problem induced by i.i.d. random variables is nonlinear in the probability distribution $\mathbb{P}$. In fact, due to the product structure resembling multilinear terms in a finite-dimensional mathematical programming problem, one could even characterize this problem as highly nonconvex in $\mathbb{P}$. As such, it is expected that future approaches for solving this problem must evolve through advancements in nonconvex semi-infinite optimization. Otherwise, one shall be left relying on either judiciously chosen combinations of objective functions and distributional information, or approximations that aim to approach these structural properties through infinite sequences of moments, as demonstrated in [27, 132], or by using infinitely-constrained ambiguity sets, as presented in [54].

A
Properties of MAD and DRO results

We recall some well-known properties of the MAD, see e.g. [21]. Denote by σ^2 the variance of the random variable X, whose distribution is known to belong to the set $\mathscr{P}_{(\mu,d)}$. Then

$$\frac{d^2}{4\beta(1-\beta)} \leqslant \sigma^2 \leqslant \frac{d(b-a)}{2}.$$

In particular, since

$$d^2 \leqslant 4\beta(1-\beta)\sigma^2 \leqslant \sigma^2,$$

it holds that $d \leqslant \sigma$. For a proof, we refer the reader to [21]. For some distributions, an explicit formula for d is available:

- Uniform distribution on $[a, b]$:
$$d = \frac{1}{4}(b-a)$$
- Normal distribution $N(\mu, \sigma^2)$:
$$d = \sqrt{\frac{2}{\pi}}\sigma$$
- Gamma distribution with parameters λ and k (for which $\mu = k/\lambda$):
$$d = \frac{2k^k}{\Gamma(k)\exp(k)}\frac{1}{\lambda}.$$

- Beta distribution with shape parameters γ, θ on support $[a, b]$:

$$d = \frac{2\gamma^{\gamma}\theta^{\theta}\Gamma(\gamma+\theta)}{(\gamma+\theta)^{\gamma+\theta+1}\Gamma(\gamma)\Gamma(\theta)}(b-a)$$

- Triangular distribution on $[a, b]$ with mode c:

$$d = \begin{cases} \frac{2(b+c-2a)^3}{81(a-b)(a-c)}, & \text{for } a+b<2c, \\ \frac{2(a+c-2b)^3}{81(a-b)(b-c)}, & \text{for } a+b>2c \end{cases}$$

- Binomial distribution with success probability p and N trials:

$$d = 2(1-p)^{N-\lfloor Np \rfloor} p^{\lfloor Np \rfloor + 1}(\lfloor Np \rfloor + 1)\binom{N}{\lfloor Np \rfloor + 1}$$

- Discrete uniform distribution on $[a, b]$, with a, b integers and $N = b - a + 1$:

$$d = \begin{cases} \frac{1}{4}N, & \text{for } N \text{ even}, \\ \frac{(N-1)(N+1)}{4N}, & \text{for } N \text{ odd}. \end{cases}$$

The MAD is known to satisfy the bound

$$0 \leqslant d \leqslant \frac{2(b-\mu)(\mu-a)}{b-a}. \tag{A.1}$$

Let $\beta = \mathbb{P}(X \geqslant \mu)$. For example, in the case of continuous symmetric distribution of X we know that $\beta = 0.5$. This quantity is known to satisfy the bounds:

$$\frac{d}{2(b-\mu)} \leqslant \beta \leqslant 1 - \frac{d}{2(\mu-a)}. \tag{A.2}$$

In Ben-Tal and Hochman [19], the following result was proved (for a much larger class of functions $f(\mathbf{y}, \mathbf{X})$ than in our case):

Proposition A.1. *If $f(\mathbf{y}, \cdot)$ is convex,*

$$\sup_{\mathbb{P}\in\mathcal{P}_{(\mu,\delta)}} \mathbb{E}_{\mathbb{P}}[f(\mathbf{y}, \mathbf{X})] = g_U(\mathbf{y}) = \sum_{\boldsymbol{\kappa}\in\{1,2,3\}^n} \prod_{i=1}^{n} p_{\kappa_i}^{(i)} f(\mathbf{y}, \xi_{\kappa_1}^{(1)}, \dots, \xi_{\kappa_n}^{(n)}), \tag{A.3}$$

with $p_{\kappa_i}^{(i)}, \xi_{\kappa_i}^{(i)}$ defined as in Lemma 6.3. *If $f(\mathbf{y}, \cdot)$ is concave,*

$$\sup_{\mathbb{P}\in\mathcal{P}_{(\mu,\delta,\beta)}} \mathbb{E}_{\mathbb{P}}[f(\mathbf{y}, \mathbf{X})] = g_L(\mathbf{y}) = \sum_{\boldsymbol{\kappa}\in\{1,2\}^n} \prod_{i=1}^{n} \hat{p}_{\kappa_i}^{(i)} f(\mathbf{y}, v_{\kappa_1}^{(1)}, \dots, v_{\kappa_n}^{(n)}), \tag{A.4}$$

with $v_1^{(i)} = \mu_i + \frac{\delta_i}{2\beta_i}$, $v_2^{(i)} = \mu_i - \frac{\delta_i}{2(1-\beta_i)}$ and $\hat{p}_1^{(i)} = \beta_i$, $\hat{p}_2^{(i)} = 1 - \beta_i$.

Hence, $g_U(\cdot)$ in (A.3) inherits the convexity in $\mathbf{y}$ from $f(\cdot,\mathbf{X})$ and its functional form depends only on the form of $f(\cdot,\mathbf{X})$, and similarly for $g_L(\cdot)$. The upper and lower bound give a closed interval for

$$\mathrm{Val}_{\mathbb{P}}(\mathbf{y}) = \mathbb{E}_{\mathbb{P}}[f(\mathbf{y},\mathbf{X})] \quad \forall \mathbb{P} \in \mathscr{P}_{(\mu,\delta,\beta)}. \tag{A.5}$$

Corollary A.2. *If* $f(\mathbf{y},\cdot)$ *is convex for all* $\mathbf{y}$ *then* $\mathrm{Val}_{\mathbb{P}}(\mathbf{y}) \in [g_L(\mathbf{y}), g_U(\mathbf{y})]$ $\forall \mathbb{P} \in \mathscr{P}_{(\mu,\delta,\beta)}$. *If* $f(\mathbf{y},\cdot)$ *is concave for all* $\mathbf{y}$ *then* $\mathrm{Val}_{\mathbb{P}}(\mathbf{y}) \in [g_U(\mathbf{y}), g_L(\mathbf{y})]$ $\forall \mathbb{P} \in \mathscr{P}_{(\mu,\delta,\beta)}$.

From Proposition A.1 we see that the extremal distribution is independent of $\mathbf{y}$. Hence, we can substitute the 3^n terms. This leads to a convex function in $\mathbf{y}$, and hence the minimization problem over $\mathbf{y}$ is tractable.

B
Proofs

B.1. Remaining proofs Chapter 3

Proof Lemma 3.3. From the derivations in [91], we know that the derivative of L can be written as

$$L'(\rho) = s + \left(\frac{s - L(\rho) + 1}{\rho} + \frac{2}{1-\rho}\right)(L(\rho) - s\rho). \tag{B.1}$$

Further differentiating both sides yields

$$\begin{aligned} L''(\rho) = (L(\rho) - s\rho)&\left(-\frac{s - L(\rho) + 1}{\rho^2} - \frac{L'(\rho)}{\rho} + \frac{2}{(1-\rho)^2}\right) \\ &+ \left(\frac{s - L(\rho) + 1}{\rho} + \frac{2}{1-\rho}\right)\left(L'(\rho) - s\right) \end{aligned} \tag{B.2}$$

and

$$\begin{aligned} L'''(\rho) = &\left(\frac{s - L(\rho) + 1}{\rho} - \frac{2}{\rho - 1}\right)L''(\rho) + 2\left(\frac{-s + L(\rho) - 1}{\rho^2} - \frac{L'(\rho)}{\rho} + \frac{2}{(\rho-1)^2}\right)\left(L'(\rho) - s\right) \\ &+ (L(\rho) - s\rho)\left(\frac{2(s - L(\rho) + 1)}{\rho^3} - \frac{L''(\rho)}{\rho} + \frac{2L'(\rho)}{\rho^2} - \frac{4}{(\rho-1)^3}\right). \end{aligned} \tag{B.3}$$

By Little's law, $W(\rho) = L(\rho)/(\rho s\mu)$. The third derivative of the expected wait is then given by

$$W'''(\rho) = \frac{\partial^3}{\partial\rho^3}\frac{L(\rho)}{\rho s\mu} = \frac{\rho\left(6L'(\rho) + \rho\left(\rho L'''(\rho) - 3L''(\rho)\right)\right) - 6L(\rho)}{s\mu\rho^4}. \tag{B.4}$$

To prove the claim, we shall show that $W'''(\rho) \geqslant 0$. Consecutively substituting (B.3), (B.2), (B.1) and finally (3.4) into (B.4), we obtain

$$W'''(\rho) = \frac{C}{s\mu(1-\rho)^4\rho^3}\Big(- s^2(1-\rho)^4((7C-6)\rho+3) + 2s(1-\rho)^2(\rho(2C((3C-7)\rho+1) + 9\rho - 4) + 1) - 6(C-2)((C-2)C+2)\rho^3 + s^3(1-\rho)^6\Big). \tag{B.5}$$

Since $\frac{C}{s\mu(1-\rho)^4\rho^3}$ is nonnegative, it suffices to show that

$$f(s,C) := \Big(- s^2(1-\rho)^4\left((7C-6)\rho+3\right) + 2s(1-\rho)^2(\rho(2C((3C-7)\rho+1) + 9\rho - 4) + 1) - 6(C-2)((C-2)C+2)\rho^3 + s^3(1-\rho)^6\Big) \geqslant 0.$$

To simplify some of the terms of $f(s,C)$, it is convenient to work with bounds for C in the remainder of the proof. We will use the simple bounds (i) $C \leqslant \rho$ and (ii) $C \leqslant 1 + \frac{s(1-\rho)^2}{2\rho} - \frac{(1-\rho)}{2\rho}\sqrt{4s\rho + s^2(1-\rho)^2}$ (see, e.g., [105]). It suffices to show for all $s \geqslant 2$ that $f(s,C(\rho)) \geqslant 0$, $\forall \rho \in (0,1)$, since the result is already demonstrated for the single-server queue. We proceed by showing that $f(n,C)$ is nondecreasing for $2 \leqslant n \leqslant s$. Then, to complete the proof, it remains to show that $f(2,C) \geqslant 0$ for all $\rho \in (0,1)$. Now notice that

$$g(x) := \frac{1}{(1-\rho)^2}\frac{\partial}{\partial n} f(n,x) = a_2x^2 - a_1x + a_0,$$

where

$$\begin{aligned} a_2 &= 12\rho^2, \\ a_1 &= 2\rho\left(7n(1-\rho)^2 + 14\rho - 2\right), \\ a_0 &= 18\rho^2 - 8\rho + 3(1-\rho)^4n^2 - 6(1-\rho)^2n + 12\rho(1-\rho)^2n + 2. \end{aligned}$$

After some algebra, one sees that, for $n \geqslant 2$, $a_0, a_1 \geqslant 0$, and clearly, $a_2 = 12\rho^2 > 0$. Since $a_0, a_1, a_2 \geqslant 0$, $g(x)$ has two positive roots. Denote the smaller root by x^-. Now, to show $g(C) \geqslant 0$, we demonstrate that $g'(C) \leqslant g'(x^-)$, which is sufficient as $g(x)$ is a convex quadratic function. We will instead prove this inequality for an upper bound on C:

$$C \leqslant 1 + \frac{n(1-\rho)^2}{2\rho} - \frac{(1-\rho)}{2\rho}\sqrt{4n\rho + n^2(1-\rho)^2} =: \bar{C},$$

where the inequality follows from (ii) and $n \leqslant s$. We next compute $g'(\bar{C})$ and $g'(x^-)$, and show that $g'(x^-) - g'(\bar{C})$ is nonnegative. Demonstrating $g'(\bar{C}) \leqslant g'(x^-)$ is sufficient since g is a quadratic decreasing function on $[0, \bar{C}]$. Notice that

$$\begin{aligned} g'(x^-) &= -2\rho(1-\rho)\sqrt{(n\left(52\rho + 13(1-\rho)^2n + 44\right) - 20)} \\ g'(\bar{C}) &= -2\rho(1-\rho)\left((1-\rho)n + 6\sqrt{n\left(4\rho + (1-\rho)^2n\right)} - 2\right). \end{aligned}$$

Since $g'(x^-), g'(\bar{C}) < 0$, it is sufficient to show $\frac{|g'(x^-)|}{2\rho(1-\rho)} \leqslant \frac{|g(\bar{C})|}{2\rho(1-\rho)}$ by demonstrating nonnegativity of

$$\begin{aligned}\bar{g}(n,\rho) &:= \left(\frac{g(\bar{C})}{2\rho(1-\rho)}\right)^2 - \left(\frac{g'(x^-)}{2\rho(1-\rho)}\right)^2 \\ &= \left(6\sqrt{n\left(4\rho+(1-\rho)^2 n\right)} + (1-\rho)n - 2\right)^2 - n\left(52\rho + 13(1-\rho)^2 n + 44\right) + 20.\end{aligned}$$

After some tedious calculations, it follows from standard calculus that $\frac{\partial \bar{g}}{\partial n} > 0$, for $n \geqslant 2$. So, for fixed ρ, the auxiliary function $\bar{g}(\cdot,\rho)$ is minimized at $n = 2$. Hence,

$$\bar{g}(n,\rho) \geqslant \bar{g}(2,\rho) = 96\rho^2 - 48\rho\sqrt{1+\rho^2} + 24 \geqslant 17.5692 > 0,$$

for all $\rho \in (0,1)$. Here the final inequality follows from minimizing $96\rho^2 - 48\rho\sqrt{1+\rho^2} + 24$. Therefore, $g(C) \geqslant 0$ or, equivalently, $\frac{\partial}{\partial n} f(n,C) \geqslant 0$ for $2 \leqslant n \leqslant s$. In the remainder of the proof, it thus suffices to concentrate on $f(2,C)$. Define

$$\begin{aligned}h(x) := f(2,x) = \Big(&-4(1-\rho)^4\left((7x-6)\rho+3\right) + 4(1-\rho)^2(\rho(2x((3x-7)\rho+1) \\ &+ 9\rho - 4) + 1) - 6(x-2)((x-2)x+2)\rho^3 + 8(1-\rho)^6\Big).\end{aligned}$$

Observe that

$$h''(x) = 12\rho^2(\rho(4\rho - 3x - 4) + 4).$$

From the well-known bound (i), it follows that $h''(x) \geqslant 0$ for $x \leqslant \rho$ since $h''(\rho) = 12(\rho-2)^2\rho^2 \geqslant 0$. Further, it can be shown that

$$h'(\rho) = 2\rho(\rho^4 + 4\rho^3 - 18\rho^2 + 20\rho - 10) \leqslant 0, \ \forall \rho \in (0,1).$$

Since $h''(x) \geqslant 0$, $h'(x) \leqslant 0$ for all $x \leqslant \rho$. Hence, for $\rho \in (0,1)$,

$$f(2,C) = h(C) \geqslant h(\rho) = -2\rho^2((\rho-2)\rho((\rho-2)\rho+2)-2) \geqslant 0,$$

in which the final inequality follows from some straightforward calculus. This completes the proof. □

B.2. Remaining proofs Chapter 4

Proof of Theorem 4.4. We solve problems (4.12) and (4.13) by considering the four scenarios depicted in Figure 4.3. Scenario 1a implies $F_m(0) = 0$, $F_m(t) = F_m(b) = 1$, which gives

$$\lambda_0 = \frac{m}{t}, \quad \lambda_1^- = 0, \quad \lambda_1^+ = \frac{t-2m+b}{t}, \quad \lambda_2 = -\frac{1}{t},$$

and objective value

$$\lambda_0 + \lambda_1^+\frac{1}{2} + \lambda_2 d_m = \frac{b-2d_m}{t} + \frac{1}{2}.$$

Solving the primal problem (4.12) with probability masses on the points $\{0, t, b\}$ gives

$$\int_x \mathbb{1}\{x \geqslant t\} \mathrm{d}\mathbb{P}(x) = p_t + p_b = \frac{b - 2d_m}{2t} + \frac{1}{2}.$$

Scenario 1b implies that $F(0) = F(t) = F(m) = F(b) = 1$, and hence $\lambda_0 = 1$, $\lambda_1^+ = \lambda_1^- = \lambda_2 = 0$ with objective value 1.

Scenario 2a implies $F(m) = 0$, $F(t) = 1$ which gives

$$\lambda_0 = \lambda_1^- = \lambda_1^+ = 0, \quad \lambda_2 = \frac{1}{t - m},$$

with objective value

$$\lambda_2 d_m = \frac{d_m}{t - m}.$$

Solving the optimal probabilities for the primal problem (4.12) indeed shows that $p_t = \frac{d_m}{t-m}$. Scenario 2b implies that $F(0) = 0$, $F(m) = F(t) = F(b) = 1$, and

$$\lambda_0 = 1, \quad \lambda_1^- = -1, \quad \lambda_1^+ = \lambda_2 = 0,$$

with objective value

$$\lambda_0 + (\lambda_1^- + \lambda_1^+)\frac{1}{2} = \frac{1}{2}.$$

Solving (4.12), with support $\{m, t, b\}$, gives $p_m = 1/2$. □

Proof of Theorem 4.5. For a random variable X with distribution $\mathbb{P} \in \mathscr{P}_{(\mu,b,d,\beta)}$, we now solve

$$\begin{aligned} \sup_{\mathbb{P} \in \mathcal{M}^+} \quad & \int_x \mathbb{1}_{\{x \geqslant t\}} \, \mathrm{d}\mathbb{P}(x) \\ \text{s.t.} \quad & \int_x \mathrm{d}\mathbb{P}(x) = 1, \ \int_x x \, \mathrm{d}\mathbb{P}(x) = \mu, \\ & \int_x |x - \mu| \, \mathrm{d}\mathbb{P}(x) = d, \ \int_x \mathbb{1}_{\{x \geqslant \mu\}} \, \mathrm{d}\mathbb{P}(x) = \beta, \end{aligned} \tag{B.6}$$

which is a semi-infinite linear program with four equality constraints.

Consider the dual of (B.6),

$$\begin{aligned} \inf_{\lambda_0, \lambda_1, \lambda_2, \lambda_3} \quad & \lambda_0 + \lambda_1 \mu + \lambda_2 d + \lambda_3 \beta \\ \text{s.t.} \quad & \mathbb{1}_{\{x \geqslant t\}} \leqslant \lambda_0 + \lambda_1 x + \lambda_2 |x - \mu| + \lambda_3 \mathbb{1}_{\{x \geqslant \mu\}}, \ \forall x \in [0, b]. \end{aligned} \tag{B.7}$$

Define $F(x) = \lambda_0 + \lambda_1 x + \lambda_2 |x - \mu| + \lambda_3 \mathbb{1}_{\{x \geqslant \mu\}}$. Then the inequality in (B.7) can be written as $\mathbb{1}_{\{x \geqslant t\}} \leqslant F(x)$, $\forall x$, i.e. $F(x)$ majorizes $\mathbb{1}_{\{x \geqslant t\}}$. Note that $F(x)$ has both a "kink" and a jump discontinuity at μ. There are four candidate scenarios, which are described in Figure B.1. When $t \in [0, \mu)$, $F(x)$ touches $\mathbb{1}_{\{x \geqslant t\}}$ in $\{0, t\} \cup [\mu, b]$ (scenario 1a), or $F(x) = 1$ and touches in $[t, b]$ (scenario 1b). When $t \in [\mu, b]$, $F(x)$ touches in $[0, \mu] \cup \{t\}$ (scenario 2a), or in $[0, \mu) \cup [t, b]$ (scenario 2b).

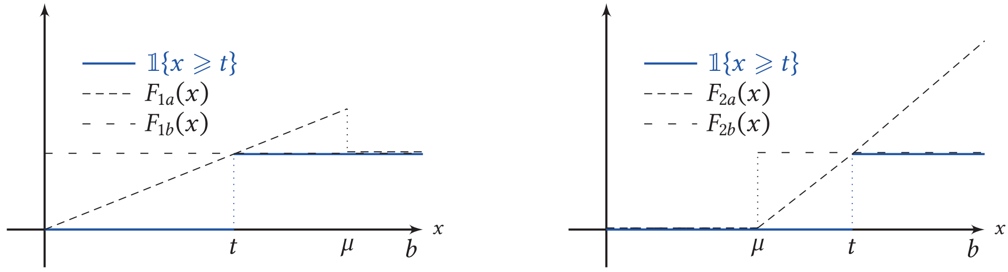


Figure B.1: Scenario 1 and the majorizing functions $F_{1a}(x)$ and $F_{1b}(x)$ under scenarios 1a and 1b, respectively. Scenario 2 and the majorizing functions $F_{2a}(x)$ and $F_{2b}(x)$ under scenarios 2a and 2b, respectively.

Scenario 1a implies $F(0) = 0,\ F(t) = F(\mu) = F(b) = 1$, which gives

$$\lambda_0 = \frac{\mu}{2t}, \quad \lambda_1 = \frac{1}{2t}, \quad \lambda_2 = -\frac{1}{2t}, \quad \lambda_3 = \frac{\mu - t}{t},$$

and objective value

$$\lambda_0 + \lambda_1\mu + \lambda_2 d + \lambda_3\beta = \frac{(1-\beta)\mu + \beta t}{t} - \frac{d}{2t}.$$

Solving the primal problem (B.6) with probability masses on the points $\{0, t, \mu, b\}$ gives

$$\int_x \mathbb{1}\{x \geqslant t\} \mathrm{d}\mathbb{P}(x) = p_t + p_\mu + p_b = \frac{(1-\beta)\mu + \beta t}{t} - \frac{d}{2t}.$$

Since primal and dual feasible solutions have the same objective value we have strong duality and hence found the optimal solutions.

Scenario 1b implies that $F(0) = F(t) = F(\mu) = F(b) = 1$, and hence $\lambda_0 = 1,\ \lambda_1 = \lambda_2 = \lambda_3 = 0$ with objective value 1. It is clear that the optimal primal objective value is also equal to 1.

Scenario 2a implies $F(0) = F(\mu) = 0,\ F(t) = 1$ which gives

$$\lambda_0 = -\frac{\mu}{2(t-\mu)}, \quad \lambda_1 = \lambda_2 = \frac{1}{2(t-\mu)}, \quad \lambda_3 = 0,$$

with objective value

$$\lambda_0 + \lambda_1\mu + \lambda_2 d + \lambda_3\beta = \frac{d}{2(t-\mu)}.$$

Solving (B.6) with probabilities masses on $\{0, \nu, \mu, b\}$, with $\nu \in (0, \mu)$, indeed shows that $p_t = \frac{d}{2(t-\mu)}$.

Scenario 2b implies that $F(0) = 0,\ F(\mu) = F(t) = F(b) = 1$, which gives as the dual feasible solution

$$\lambda_0 = \lambda_1 = \lambda_2 = 0, \quad \lambda_3 = 1,$$

and objective value

$$\lambda_0 + \lambda_1\mu + \lambda_2 d + \lambda_3\beta = \beta.$$

Finding the optimal probabilities of (B.6) confirms that $p_0 = (1-\beta)$. □

Proof Corollary 4.6. We solve

$$\begin{aligned} \inf_{\mathbb{P}\in\mathcal{M}^+} \quad & \int_x \mathbb{1}_{\{x>t\}}\,\mathrm{d}\mathbb{P}(x) \\ \text{s.t.} \quad & \int_x \mathrm{d}\mathbb{P}(x) = 1,\ \int_x x\,\mathrm{d}\mathbb{P}(x) = \mu, \\ & \int_x |x-\mu|\,\mathrm{d}\mathbb{P}(x) = d,\ \int_x \mathbb{1}_{\{x\geqslant\mu\}}\,\mathrm{d}\mathbb{P}(x) = \beta, \end{aligned} \tag{B.8}$$

which is a semi-infinite linear program with four equality constraints. The dual problem is given by

$$\begin{aligned} \sup_{\lambda_0,\lambda_1,\lambda_2,\lambda_3} \quad & \lambda_0 + \lambda_1\mu + \lambda_2 d + \lambda_3\beta \\ \text{s.t.} \quad & \mathbb{1}_{\{x>t\}} \geqslant \lambda_0 + \lambda_1 x + \lambda_2|x-\mu| + \lambda_3\mathbb{1}_{\{x\geqslant\mu\}} =: F(x),\ \forall x\in[0,b]. \end{aligned} \tag{B.9}$$

The proof is similar to that of Theorem 4.5 but, since we are minimizing, $F(x)$ is a minorizing function. Note that $F(x)$ has both a “kink” and a jump discontinuity at μ. There are four candidate solutions, which are depicted in Figure B.2. When $t \in [0,\mu)$, $F(x)$ touches $\mathbb{1}_{\{x>t\}}$ in $\{t\}\cup[\mu,b]$ (scenario 1a) or in $[0,t]\cup[\mu,b]$ (scenario 1b). When $t\in[\mu,b]$, $F(x)$ touches in $[0,\mu)\cup\{t,b\}$ (scenario 2a), or $F(x)=0$ and touches in $[0,t]$ (scenario 2b).

Scenario 1a implies $F(t)=0$, $F(\mu)=F(b)=1$, which gives the dual solution

$$\lambda_0 = \frac{2t-\mu}{2(t-\mu)},\quad \lambda_1 = -\frac{1}{2(t-\mu)},\quad \lambda_2 = \frac{1}{2(t-\mu)},\quad \lambda_3 = 0,$$

and objective value

$$\lambda_0+\lambda_1\mu+\lambda_2 d+\lambda_3\beta = 1-\frac{d}{2(\mu-t)}.$$

Solving the primal problem (B.8) with probability masses on the points $\{t,\mu,\nu,b\}$, with $\nu\in(\mu,b)$, gives

$$\int_x \mathbb{1}\{x>t\}\,\mathrm{d}\mathbb{P}(x) = 1-p_t = 1-\frac{d}{2(\mu-t)}.$$

Scenario 1b implies that $F(0)=F(t)=0$, $F(\mu)=F(b)=1$, and hence $\lambda_0=\lambda_1=\lambda_2=0$, $\lambda_3=1$ with objective value β. Now solving the primal problem with probability masses on $\{0,t,\mu,b\}$ gives us

$$\int_x \mathbb{1}\{x>t\}\,\mathrm{d}\mathbb{P}(x) = p_\mu + p_b = \beta.$$

Scenario 2a implies that $F(0)=F(t)=0$ and $F(b)=1$, which results in

$$\lambda_0 = \frac{\mu}{2(t-b)},\quad \lambda_1=\lambda_2=\frac{1}{2(b-t)},\quad \lambda_3 = -\frac{(\mu-t)}{(b-t)},$$

with objective value

$$\lambda_0+\lambda_1\mu+\lambda_2 d+\lambda_3\beta = \frac{\beta(\mu-t)}{(b-t)}+\frac{d}{2(b-t)}.$$

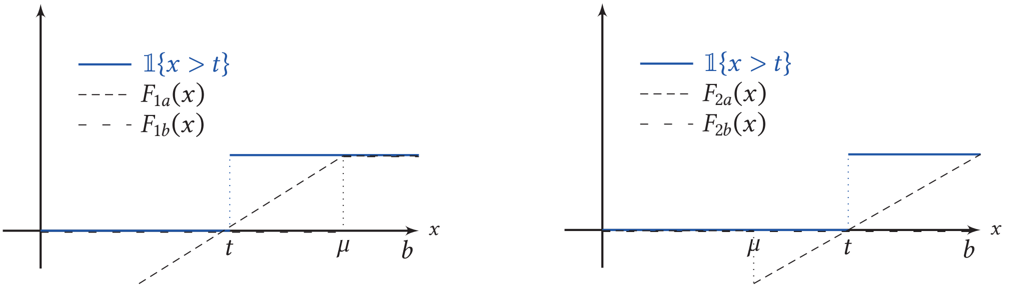


Figure B.2: Scenario 1 and the minorizing functions $F_{1a}(x)$ and $F_{1b}(x)$ under scenarios 1a and 1b, respectively. Scenario 2 and the minorizing functions $F_{2a}(x)$ and $F_{2b}(x)$ under scenarios 2a and 2b, respectively.

Indeed, solving the primal problem with probability masses on $\{0, \nu, t, b\}$, with $\nu \in (0, \mu)$, gives $p_b = \frac{\beta(\mu-t)}{(b-t)} + \frac{d}{2(b-t)}$.

Scenario 2b implies that $F(0) = F(\mu) = F(t) = F(b) = 0$, which gives the dual feasible solution

$$\lambda_0 = \lambda_1 = \lambda_2 = \lambda_3 = 0,$$

with objective value 0. All probability mass is placed on points that are less than or equal to t. Hence, the optimal primal objective value is also equal to 0. □

B.3. Remaining proofs Chapter 5

Proof of Lemma 5.1. Suppose the moment constraints are consistent so that $\mathcal{P}$ is nonempty. Let $\tilde{\tau}$ and τ^* be the optimal values, and optimal solution, of (5.4) and (5.8), respectively. To prove the result, it suffices to show that

$$\tau^* = \sup_{\mathbb{P}\in\mathcal{P}} \frac{\mathbb{E}_{\mathbb{P}}[g(X)\mathbb{1}_{\Xi}(X)]}{\mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\Xi}(X)]} =: \tilde{\tau}.$$

Notice that

$$\frac{\mathbb{E}_{\mathbb{P}}[g(X)\mathbb{1}_{\Xi}(X)]}{\mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\Xi}(X)]} \leqslant \tilde{\tau},$$

for all $\mathbb{P} \in \mathcal{P}$. Rewriting and taking the supremum over $\mathcal{P}$, the inequality above is equivalent to

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[g(X)\mathbb{1}_{\Xi}(X) - \tilde{\tau}\mathbb{1}_{\Xi}(X)] \leqslant 0,$$

which implies $\tilde{\tau}$ is a feasible solution to problem (5.8). Since $\tilde{\tau}$ is optimal to (5.4),

$$\tau^* \leqslant \sup_{\mathcal{P}} \frac{\mathbb{E}_{\mathbb{P}}[g(X)\mathbb{1}_{\Xi}(X)]}{\mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\Xi}(X)]} = \tilde{\tau}. \tag{B.10}$$

As τ^* is a feasible solution to (5.8), it holds that

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[g(X)\mathbb{1}_{\Xi}(X) - \tau^*\mathbb{1}_{\Xi}(X)\right] \leqslant 0. \tag{B.11}$$

We next show that this implies

$$\sup_{\mathbb{P}\in\mathcal{P}} \frac{\mathbb{E}_{\mathbb{P}}[g(X)\mathbb{1}_{\Xi}(X)]}{\mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\Xi}(X)]} \leqslant \tau^*. \tag{B.12}$$

For the sake of contradiction, assume that for an arbitrary $\epsilon > 0$, there exists a sequence of probability measures $\{\mathbb{P}_k\}$ such that

$$\tau^* + \epsilon \leqslant \frac{\mathbb{E}_{\mathbb{P}_k}[g(X)\mathbb{1}_{\Xi}(X)]}{\mathbb{E}_{\mathbb{P}_k}[\mathbb{1}_{\Xi}(X)]}$$

as k grows large. Fixing $\mathbb{P}_k$ for a sufficiently large k, we obtain

$$\epsilon\mathbb{E}_{\mathbb{P}_k}[\mathbb{1}_{\Xi}(X)] \leqslant \mathbb{E}_{\mathbb{P}_k}[g(X)\mathbb{1}_{\Xi}(X) - \tau^*\mathbb{1}_{\Xi}(X)].$$

Since $\mathbb{E}_{\mathbb{P}_k}[\mathbb{1}_{\Xi}(X)] = \mathbb{P}_k(X \in \Xi) > 0$ and $\epsilon > 0$, this inequality contradicts (B.11). Moreover, if $\mathbb{P}_k(X \in \Xi) = 0$, the fractional objective function diverges. However, this also yields a contradiction by boundedness of the optimal values. Hence, it follows from combining (B.10) and (B.12) that $\tau^* = \tilde{\tau}$. This completes the proof. □

Proof of Proposition 5.5. Replacing the second-moment function x^2 by $|x-\mu|$ and substituting d for $(\sigma^2 + \mu^2)$ in (5.12) yields the dual problem for $\sup_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}[X|X \geqslant t]$,

$$\begin{aligned} \inf_{\lambda_0,\lambda_1,\lambda_2,\lambda_3} \quad & \lambda_3 \\ \text{subject to} \quad & \lambda_0 + \lambda_1\mu + \lambda_2 d \leqslant 0, \\ & \lambda_0 + \lambda_1 x + \lambda_2|x-\mu| \geqslant 0, \qquad \forall a \leqslant x < t, \\ & \lambda_0 + \lambda_1 x + \lambda_2|x-\mu| \geqslant x - \lambda_3, \quad \forall t \leqslant x \leqslant b. \end{aligned} \tag{B.13}$$

Denote the left-hand sides of the second and third constraints by $D(x) := \lambda_0+\lambda_1 x+\lambda_2|x-\mu|$. The function $D(x)$ is dual feasible when it is greater than or equal to 0 for $x < t$ and greater than or equal to $x-\lambda_3$ for $t \leqslant x \leqslant b$. First, we consider the case $t < \mu$. To solve the dual problem, we shall consider three cases for the shape of the dual function $D(x)$, as illustrated in Figure B.3.

First, we discuss the case where $D(x)$ is a straight line (i.e., $\lambda_2 = 0$). If $\lambda_1 = 0$, then $D_1(x)$ is a horizontal line, which is dual feasible because, for a suitable choice of λ_3, this function lies above $\mathbb{1}_{\{x\geqslant t\}}(x-\lambda_3)$. Note that for this solution to satisfy $\lambda_0+\lambda_1\mu+\lambda_2 d \leqslant 0$, the constant $\lambda_0 = 0$. Since we are minimizing λ_3 (or maximizing $-\lambda_3$), we push the function $x-\lambda_3$ upward until it hits the horizontal line, so this choice for the dual function yields b as the optimal objective value. The dual function $D_1(x)$ cannot have a positive slope because this would imply $\lambda_0 + \lambda_1\mu > 0$. Note that a line with negative slope $\lambda_1 < 0$ will only increase λ_3. Hence, the horizontal line with $\lambda_0 = 0$ is the best feasible option. A primal solution that attains this value is the distribution with support $\left\{\frac{b(d-2\mu)+2\mu^2}{-2b+d+2\mu}, b\right\}$ and probabilities

$$p_1 = 1 - \frac{d}{2(b-\mu)}, \quad p_2 = \frac{d}{2(b-\mu)}.$$

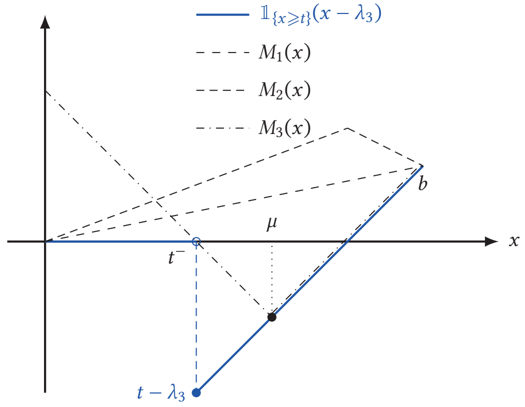


Figure B.3: $D_1(x)$, $D_2(x)$, and $D_3(x)$

Using complementary slackness, we argue that the second case, $D_2(x)$, can be omitted. From Figure B.3, observe that the corresponding primal solution is supported on the values a, b. However, we cannot construct a two-point distribution that, in general, satisfies the moment constraints. Therefore, this second case does not provide a useful solution from which we can obtain a tight bound.

For the third case (wedge), let $D_3(x)$ coincide with $\mathbb{1}\{x \geqslant t\}(x - \lambda_3)$ at $x = t$, μ and b. Choosing $D_3(x)$ in this particular way, the dual variables satisfy

$$\lambda_0 = \frac{(t + \lambda_3)\mu - 2t\lambda_3}{2(t - \mu)}, \; \lambda_1 = \frac{\lambda_3 + t - 2\mu}{2(t - \mu)}, \; \lambda_2 = \frac{t - \lambda_3}{2(t - \mu)}.$$

Substituting these values into $\lambda_0 + \lambda_1\mu + \lambda_2 d \leqslant 0$, we obtain

$$-\lambda_3 + \mu + \frac{d(\lambda_3 - t)}{2(\mu - t)} \leqslant 0,$$

in which the left-hand side is a decreasing (linear) function of λ_3 for $t < \mu - \frac{d(b-\mu)}{2(b-\mu)-d}$. Since we are minimizing with respect to λ_3, we choose λ_3 such that equality is attained. Hence, $\lambda_3^* = \mu + \frac{d(\mu-t)}{2(\mu-t)-d}$. The distribution with support $\left\{t^-, \lambda_3^*\right\}$, and respective probabilities

$$p_1 = \frac{d}{2(\mu - t)}, \quad p_2 = 1 - \frac{d}{2(\mu - t)},$$

achieves λ_3^* asymptotically. Combining the two feasible cases, and ensuring these bounds are tight by constructing primal feasible solutions that (asymptotically) achieve these bounds, we arrive at the desired result. □

Proof of Proposition 5.6. In this setting, the dual is given by

$$
\begin{aligned}
\inf_{\lambda_0,\lambda_1,\lambda_2,\lambda_3} \quad & \lambda_3 \\
\text{subject to} \quad & \lambda_0 + \lambda_1\mu + \lambda_2 d \leqslant 0, \\
& \lambda_0 + \lambda_1 x + \lambda_2 d(x) \geqslant 0, \qquad \forall x < t, \\
& \lambda_0 + \lambda_1 x + \lambda_2 d(x) \geqslant x - \lambda_3, \quad \forall x \geqslant t.
\end{aligned}
\tag{B.14}
$$

Define $V(x) := \lambda_0 + \lambda_1 x + \lambda_2 d(x)$. Analogous to the variance and MAD settings, a candidate solution is the majorant $V(x)$ that touches at $x = t^-$, and is tangent to $x - \lambda_3$ at a point $x_0 > t$ that will be determined a posteriori. Using these insights, we solve the following system of equations to determine p_t, p_{x_0} and the dual variables $\lambda_0, \lambda_1, \lambda_2, \lambda_3$ (with x_0 fixed):

$$
\begin{aligned}
& p_t t + p_{x_0} x_0 = \mu, \ p_t d(t) + p_{x_0} d(x_0) = \bar{\sigma}, \\
& \lambda_0 + \lambda_1 t + \lambda_2 d(t) = 0, \ \lambda_0 + \lambda_1 x_0 + \lambda_2 d(x_0) = x_0 - \lambda_3 \\
& \lambda_1 + \lambda_2 d'(x_0) = 1, \ \lambda_0 + \lambda_1\mu + \lambda_2(\mu^2 + \sigma^2) = 0,
\end{aligned}
$$

where the first line contains the moment constraint, the second and third line fix the shape of $V(x)$, and finally, we assume that the dual constraint $\lambda_0 + \lambda_1\mu + \lambda_2(\mu^2 + \sigma^2) \leqslant 0$ is tight. The derivative $d'(x)$ is assumed to be the right derivative, in order to allow for nondifferentiable dispersion functions. Notice that the resulting dual solution is always feasible, since $V(x)$ is convex ($\lambda_2 \geqslant 0$ is a necessary condition for dual feasibility), and therefore, $V(x) \geqslant (x - \lambda_3)\mathbb{1}_{\Xi}(x), \ \forall x,$ by the constraints $V'(x_0) = 1$ and $V(t) = 0$. Solving the equations, one obtains

$$
\begin{aligned}
p_t &= \frac{\bar{\sigma} t - \mu d(t)}{t d(x_0) - x_0 d(t)}, \ p_2 = \frac{\bar{\sigma} x_0 - \mu d(x_0)}{x_0 d(t) - t d(x_0)}, \\
\lambda_3 &= \frac{-(t-\mu)\,(d(x_0) - x_0 d'(x_0)) - \mu d(t) + \bar{\sigma} t}{(t-\mu) d'(x_0) - d(t) + \bar{\sigma}},
\end{aligned}
$$

and

$$
\begin{aligned}
\lambda_0 = \frac{\mu d(t) - \bar{\sigma} t}{(t-\mu) d'(x_0) - d(t) + \bar{\sigma}}, \ \lambda_1 &= \frac{\bar{\sigma} - d(t)}{(t-\mu) d'(x_0) - d(t) + \sigma}, \\
\lambda_2 &= \frac{t-\mu}{(t-\mu) d'(x_0) - d(t) + \bar{\sigma}}.
\end{aligned}
$$

To guarantee strong duality, we choose x_0 such that

$$
\begin{aligned}
x_0 &= \frac{-(t-\mu)\,(d(x_0) - x_0 d'(x_0)) - \mu d(t) + \bar{\sigma} t}{(t-\mu) d'(x_0) - d(t) + \bar{\sigma}} = \lambda_3 \\
\iff & \frac{(t - x_0)\bar{\sigma} + (x_0 - \mu) d(t) + (\mu - t) d(x_0)}{(t-\mu) d'(x_0) - d(t) + \bar{\sigma}} = 0,
\end{aligned}
$$

and the normalization constraint $p_t + p_{x_0} = 1$ hold. Both conditions are equivalent to

$$
(t - x_0)\bar{\sigma} + (x_0 - \mu) d(t) + (\mu - t) d(x_0) = 0.
$$

Consequently, the x_0^* that follows from solving

$$\frac{\bar{\sigma}t - \mu d(t)}{td(x_0) - x_0 d(t)} + \frac{\bar{\sigma}x_0 - \mu d(x_0)}{x_0 d(t) - td(x_0)} = 1,$$

is optimal. Hence, the claim follows. □

Proof of Proposition 5.7. The class of symmetric pairs of Dirac measures (i.e., $\delta_{\mu-x},\ \delta_{\mu+x},\ x \geqslant 0$) generates the set of symmetric distributions about μ. From Theorem 5.2, it follows that the dual problem is given by

$$\begin{aligned}
\inf_{\lambda_0,\lambda_1,\lambda_2,\lambda_3} \quad & \lambda_3 \\
\text{subject to} \quad & \lambda_0 + \lambda_1\mu + \lambda_2(\sigma^2 + \mu^2) \leqslant 0, \\
& 2\lambda_0 + 2\lambda_1\mu + 2\lambda_2(x^2 + \mu^2) + \lambda_3 \mathbb{1}_\Xi(\mu - x) + \lambda_3 \mathbb{1}_\Xi(\mu + x) \\
& \qquad \geqslant (\mu - x)\mathbb{1}_\Xi(\mu - x) + (\mu + x)\mathbb{1}_\Xi(\mu + x),\ \forall x \geqslant 0.
\end{aligned} \tag{B.15}$$

The last constraint can be reduced to

$$\begin{aligned}
2\lambda_0 + 2\lambda_1\mu + 2\lambda_2(x^2 + \mu^2) &\geqslant -2\lambda_3 + 2\mu, \quad \forall 0 \leqslant x < \mu - t, \\
2\lambda_0 + 2\lambda_1\mu + 2\lambda_2(x^2 + \mu^2) &\geqslant x + \mu - \lambda_3, \quad \forall x \geqslant \mu - t.
\end{aligned}$$

Notice that $\mathbb{1}_\Xi(\mu + x) = 1,\ \forall x \geqslant 0$, since it is assumed that $t < \mu$. Define the quadratic function $M^{\text{sym}}(x) := 2\lambda_0 + 2\lambda_1\mu + 2\lambda_2(x^2 + \mu^2)$. We suggest two possible solutions for the dual problem. The first solution, denoted as $M_1^{\text{sym}}(x)$, touches $-2\lambda_3 + 2\mu$ at $x = 0$ and $x + \mu - \lambda_3$ at $\mu - t$. The second solution, $M_2^{\text{sym}}(x)$, is a quadratic function that touches $x + \mu - \lambda_3$ at an optimal point x_0 that is unknown a priori. We further postulate that in both dual solutions, the constraint $\lambda_0 + \lambda_1\mu + \lambda_2(\mu^2 + \sigma^2) \leqslant 0$ is tight. In the interest of space, we omit the figure, but it is easily verified that the suggested solutions are dual feasible. The corresponding primal solutions follow from complementary slackness and are the pairs of Dirac measure $\delta_{\mu-x},\ \delta_{\mu+x}$ in which for x we substitute the points at which the dual function coincides with the right-hand sides of the constraints in (B.15).

The dual variables which correspond to $M_1^{\text{sym}}(x)$ are

$$\lambda_1 = \frac{-\lambda_0 + \frac{(t-\mu)(\mu^2+\sigma^2)}{2(t-\mu)^2-\sigma^2}}{\mu},\ \lambda_2 = \frac{\mu - t}{2(t-\mu)^2 - \sigma^2},$$

yielding

$$\lambda_3^* = \frac{2(t-\mu)^2\mu - t\sigma^2}{2(t-\mu)^2 - \sigma^2}$$

as our guess for the optimal value of the dual problem. The proposed solution is feasible for the dual problem since $M_1^{\text{sym}}(x)$ is convex (as $\lambda_2 \geqslant 0$) and tangent to $-2\lambda_3 + 2\mu$ at $x = 0$, and further some straightforward calculations show that the derivative of $M_1^{\text{sym}}(x)$ at $x = \mu - t$ is greater than 1, so that $M_1^{\text{sym}}(x) \geqslant x + \mu - \lambda_3^*,\ \forall x \geqslant \mu - t$. For the primal probabilities, it follows from the variance constraint

$$(1-p)\mu^2 + \frac{1}{2}pt^2 + \frac{1}{2}p(2\mu - t)^2 = \mu^2 + \sigma^2$$

that

$$p = \frac{\sigma^2}{(t-\mu)^2}.$$

Hence,

$$\mathbb{E}[X \,|\, X \geqslant t] = \frac{\frac{\sigma^2}{2(t-\mu)^2}(2\mu - t) + \left(1 - \frac{\sigma^2}{(t-\mu)^2}\right)\mu}{\frac{\sigma^2}{2(t-\mu)^2} + \left(1 - \frac{\sigma^2}{(t-\mu)^2}\right)} = \frac{2\mu(t-\mu)^2 - \sigma^2 t}{2(t-\mu)^2 - \sigma^2} = \lambda_3^*,$$

so that these are the optimal primal-dual solutions by weak duality. For the second case, we determine the candidate support point x_0 first. From the moment constraints and the fact that the solution should be symmetric about μ, we find that $x_0^* = \sigma$ and therefore, the primal candidate is given by the distribution $\frac{1}{2}\delta_{\mu-\sigma} + \frac{1}{2}\delta_{\mu+\sigma}$. The second dual solution then yields

$$\lambda_1 = -\frac{4\lambda_0\sigma + \mu^2 + \sigma^2}{4\mu\sigma},\ \lambda_2 = \frac{1}{4\sigma},\ \lambda_3^* = \mu + \sigma.$$

Thus,

$$\mathbb{E}[X \,|\, X \geqslant t] = \mu + x_0^* = \mu + \sigma = \lambda_3^*.$$

For $t \geqslant \mu$, there exists a sequence of measures supported on $\{\mu, \mu - k, \mu + k\}$ that is feasible in the primal, and for which the conditional expectation diverges, as $k \to \infty$. This feasible sequence is given by

$$\mathbb{P}_k = \left(1 - \frac{\sigma^2}{k^2}\right)\delta_\mu + \frac{1}{2}\frac{\sigma^2}{k^2}\delta_{\mu-k} + \frac{1}{2}\frac{\sigma^2}{k^2}\delta_{\mu+k}.$$

It is then easily verified that $\lim_{k\to\infty} \mathbb{E}_{\mathbb{P}_k}[X \,|\, X \geqslant t]$ diverges. Combining the cases above, while checking for feasibility of the primal solutions, the claim follows. □

Proof of Proposition 5.8. Symmetric, unimodal distributions with mode μ can be generated by rectangular/uniform distributions, possibly including a Dirac measure at μ (i.e., $\delta_{[\mu-z,\mu+z]}$, $z \geqslant 0$). From Theorem 5.2, it follows that the dual problem is given by

$$\begin{aligned}
\inf_{\lambda_0,\lambda_1,\lambda_2,\lambda_3} \quad & \lambda_3 \\
\text{subject to} \quad & \lambda_0 + \lambda_1\mu + \lambda_2(\sigma^2 + \mu^2) \leqslant 0, \\
& \int_{\mu-x}^{\mu+x} \lambda_0 + \lambda_1 z + \lambda_2 z^2 \mathrm{d}z \geqslant \int_{\mu-x}^{\mu+x} (z - \lambda_3)\mathbb{1}_\Xi(z)\,\mathrm{d}z,\ \forall x > 0, \\
& \lambda_0 + \lambda_1\mu + \lambda_2\mu^2 \geqslant (\mu - \lambda_3).
\end{aligned} \tag{B.16}$$

After computing the integral on the left-hand side of the penultimate constraint, one obtains

$$\frac{2}{3}x\left(3\lambda_0 + 3\mu(\lambda_1 + \lambda_2\mu) + \lambda_2 x^2\right) \geqslant \int_{\mu-x}^{\mu+x} (z - \lambda_3)\mathbb{1}_\Xi(z)\,\mathrm{d}z,\ \forall x > 0.$$

For the right-hand side, we distinguish two cases so that we can split the semi-infinite constraint into two sets, resulting in the system of inequalities

$$\begin{aligned} \frac{2}{3}x\left(3\lambda_0 + 3\mu(\lambda_1 + \lambda_2\mu) + \lambda_2 x^2\right) &\geqslant 2x(\mu - \lambda_3), \quad \forall 0 < x < \mu - t, \\ \frac{2}{3}x\left(3\lambda_0 + 3\mu(\lambda_1 + \lambda_2\mu) + \lambda_2 x^2\right) &\geqslant -\frac{1}{2}(t - x - \mu)(t + x - 2\lambda_3 + \mu), \quad \forall x \geqslant \mu - t. \end{aligned} \tag{B.17}$$

Again, we can make an educated guess for an optimal dual solution, and using weak duality, prove optimality by constructing a matching primal solution using the complementary slackness property. The dual solution is now characterized by a third-order polynomial function, $M^{\text{uni}}(x) := \frac{2}{3}x\left(3\lambda_0 + 3\mu(\lambda_1 + \lambda_2\mu) + \lambda_2 x^2\right)$. We show through primal-dual reasoning that there are merely two feasible options for the extremal distribution. Using this insight, we optimize the primal problem directly by plugging in the candidate form of the extremal distribution. Notice that as $M^{\text{uni}}(x)$ needs to be convex in order to be dual feasible, there cannot exist a tangent point on the interval $[0, \mu - t)$ because otherwise $M^{\text{uni}}(x)$ would need to intersect $2x(\mu - \lambda_3)$. Furthermore, there can exist only one tangent point $x = x_0^*$ at which $M^{\text{uni}}(x)$ coincides with the quadratic function $-\frac{1}{2}(t - x - \mu)(t + x - 2\lambda_3 + \mu)$, as $M^{\text{uni}}(x)$ is a cubic function. By complementary slackness, the corresponding extremal distribution is then given by the mixture of a Dirac measure at μ and a uniform distribution on the interval $[\mu - x_0^*, \mu + x_0^*]$, for the first case, or a uniform distribution on $[\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma]$ for the second. Indeed, the latter uniform distribution is the only one that is feasible for the primal problem. From these observations, it follows that the primal problem can be reduced to a finite-dimensional (nonconvex) optimization problem. The objective function of the primal problem can be rewritten as

$$\mathbb{E}[X \mid X \geqslant t] = \frac{(1-p)\mu + p\int_t^{\mu + x_0} \frac{z}{2x_0}\mathrm{d}z}{(1-p) + p\int_t^{\mu + x_0} \frac{1}{2x_0}\mathrm{d}z} = \frac{4x_0\mu - p(t + x_0 - \mu)(t - x_0 + \mu)}{4x_0 - 2p(t + x_0 - \mu)}. \tag{B.18}$$

From the variance constraint

$$(1-p)\mu^2 + p\frac{(\mu - x_0)^2 + (\mu - x_0)(\mu + x_0) + (\mu + x_0)^2}{3} = \sigma^2 + \mu^2,$$

it follows that $p = \frac{3\sigma^2}{x_0^2}$. In order to be a probability, it should hold that $p \leqslant 1$, and hence $x_0 \geqslant \sqrt{3}\sigma$. The variable p can be eliminated from (B.18), yielding the optimization problem

$$\max_{x_0} \left\{ \frac{4\mu(x_0)^3 - 3\sigma^2(-\mu + t + x_0)(\mu + t - x_0)}{4(x_0)^3 - 6\sigma^2(-\mu + t + x_0)} \; : \; x_0 \geqslant \sqrt{3}\sigma \right\}. \tag{B.19}$$

From standard arguments, it follows that the maximum of (B.19) must be attained at a critical point of the objective function or at the boundary of the feasible region. The first case follows from determining the critical point by solving the first-order condition

$$6\sigma^2 x_0^2\left(3(t - \mu)^2 - x_0^2\right) + 9\sigma^4(\mu - t - x_0)^2 = 0.$$

The second case corresponds to the boundary of the feasible region for which $p = 1$. As a consequence, the optimal tangent point $x_0^* = \sqrt{3}\sigma$. It is easy to verify that this solution yields the second case.

Finally, for the third case, $t \geqslant \mu$, we construct a maximizing sequence $\{\mathbb{P}_k\}$ for which $\mathbb{E}[X \mid X \geqslant t]$ diverges. To this end, consider

$$\mathbb{P}_k = \left(1 - \frac{3\sigma^2}{k^2}\right)\delta_\mu + \frac{3\sigma^2}{k^2}\delta_{[\mu-k,\mu+k]},$$

which is feasible for the primal problem. Then

$$\mathbb{E}_{\mathbb{P}_k}[X \mid X \geqslant t] = \frac{\int_t^{\mu+k} \frac{z}{2k}\mathrm{d}z}{\int_t^{\mu+k} \frac{1}{2k}\mathrm{d}z} = \frac{1}{2}(t+\mu+k) \overset{k\to\infty}{\longrightarrow} \infty,$$

hence resulting in the third case. Combining the cases above completes the proof. □

Proof of Proposition 5.9. The primal can be equivalently stated as

$$\sup_{\mathbb{P}\in\mathcal{P}_{(\mu,\sigma)}} \mathbb{E}[\mathbb{1}_{\{X\geqslant z\}} \mid X \geqslant p],$$

which is (weakly) dual to

$$\begin{aligned} \inf_{\lambda_0,\lambda_1,\lambda_2,\lambda_3} \quad & \lambda_3 \\ \text{subject to} \quad & \lambda_0 + \lambda_1\mu + \lambda_2(\sigma^2+\mu^2) \leqslant 0, \\ & \lambda_0 + \lambda_1 x + \lambda_2 x^2 \geqslant (\mathbb{1}_{\{x\geqslant z\}}(x) - \lambda_3)\mathbb{1}_{\Xi}(x),\ \forall x \geqslant 0. \end{aligned} \tag{B.20}$$

The right-hand side of the constraint is equal to 0, for $x < t$, $-\lambda_3$ for $t \leqslant x < z$, and $1-\lambda_3$ for $x \geqslant z$. We discuss three cases, in the order of their appearance in the claim. The first dual solution, $M_1(x)$, corresponds to a convex quadratic function that touches $(\mathbb{1}_{\{x\geqslant z\}}(x) - \lambda_3)\mathbb{1}_{\Xi}(x)$ at z and x_0, where the latter point lies between p and z. The primal probabilities, dual variables, and the support point x_0 follow from solving

$$\begin{gathered} p_{x_0} + p_z = 1,\ p_{x_0}x_0 + p_z z = \mu,\ p_{x_0}x_0^2 + p_z z^2 = \mu^2+\sigma^2, \\ \lambda_0 + \lambda_1 x_0 + \lambda_2 x_0^2 = -\lambda_3,\ \lambda_0 + \lambda_1 z + \lambda_2 z^2 = 1-\lambda_3, \\ \lambda_1 + 2\lambda_2 x_0 = 0,\ \lambda_0 + \lambda_1\mu + \lambda_2(\mu^2+\sigma^2) = 0, \end{gathered}$$

yielding as solution

$$\begin{aligned} p_{x_0} &= \frac{(z-\mu)^2}{\sigma^2+(z-\mu)^2},\ p_z = \frac{\sigma^2}{\sigma^2+(z-\mu)^2},\ x_0 = \mu + \frac{\sigma^2}{\mu - z}, \\ \lambda_0 &= \frac{(z-\mu)\left(\mu^2(z-\mu) - \sigma^2(\mu+z)\right)}{(\sigma^2+(z-\mu)^2)^2},\ \lambda_1 = \frac{2(z-\mu)\left(\mu^2+\sigma^2-\mu z\right)}{(\sigma^2+(z-\mu)^2)^2}, \\ \lambda_2 &= \frac{(z-\mu)^2}{(\sigma^2+(z-\mu)^2)^2},\ \lambda_3 = \frac{\sigma^2}{\sigma^2+(z-\mu)^2}. \end{aligned}$$

Indeed, by weak duality, this gives the best possible bound since

$$\mathbb{E}[\mathbb{1}_{\{x\geqslant z\}}(X)] = p_z = \frac{\sigma^2}{\sigma^2+(z-\mu)^2} = \lambda_3.$$

For the second case, let $M_2(x)$ denote a quadratic function that touches at $x = z$ and some point x_0, like $M_1(x)$, but additionally agrees with $(\mathbb{1}_{\{x \geqslant z\}}(x) - \lambda_3)\mathbb{1}_{\Xi}(x)$ at $x = p$. Again, we can use a similar set of conditions, as described above, to find the optimal primal and dual solutions, but now with a three-point distribution with $x_0 = \frac{\mu^2+\sigma^2-\mu z}{\mu - z}$ (which follows from the conditions). This leads to the second bound. For brevity, we omit the detailed calculations. Finally, for the third case, $M_3(x)$ is a quadratic function that touches at z and a point $0 \leqslant x \leqslant p$. The same set of calculations leads to the third upper bound, which is equal to the constant 1. We combine the cases above in such a way that the primal distributions are feasible. This completes the proof. □

B.3.1. Conic reformulations

Proof of second claim Theorem 5.10. We will focus on the second semi-infinite constraint (the first constraint can be dealt with analogously), which can equivalently be written as the collection of robust counterparts

$$\lambda_0 + \boldsymbol{\lambda}_1^\top \mathbf{x} + \boldsymbol{\lambda}_2^\top \mathbf{u} + \lambda_3 \geqslant \mathbf{s}_l(\boldsymbol{\nu})^\top \mathbf{x} + t_l(\boldsymbol{\nu}), \quad \forall (\mathbf{x}, \mathbf{u}) \in \mathscr{C}, \ \forall l \in \mathscr{L}. \tag{B.21}$$

Let "cl" denote the closure of a set. We next generate a proper cone from the uncertainty set as follows: Define $\mathscr{K} := \mathrm{cl}\left(\{(\mathbf{x}, \mathbf{u}, w) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} : (\mathbf{z}/w, \mathbf{u}/w) \in \mathscr{C}, w > 0\}\right)$, of which $\mathscr{K}^*$ is the dual cone. The semi-infinite constraint (B.21) is satisfied if, and only if,

$$\begin{aligned} \inf \quad & (\boldsymbol{\lambda}_1 - \mathbf{s}_l(\boldsymbol{\nu}))^\top \mathbf{x} + \boldsymbol{\lambda}_2^\top \mathbf{u} \\ \text{s.t.} \quad & (\mathbf{x}, \mathbf{u}, 1) \in \mathscr{K}, \end{aligned} \tag{B.22}$$

is greater than, or equal to, $t_l(\boldsymbol{\nu}) - \lambda_0 - \lambda_3$. By conic duality, the strong dual of (B.22) is given by

$$\begin{aligned} \sup \quad & -w_l \\ \text{s.t.} \quad & \boldsymbol{\lambda}_1 - \mathbf{s}_l(\boldsymbol{\nu}) - \mathbf{a}_l = \mathbf{0}, \\ & \boldsymbol{\lambda}_2 - \mathbf{b}_l = \mathbf{0}, \\ & (\mathbf{a}_l, \mathbf{b}_l, w_l) \in \mathscr{K}^*. \end{aligned} \tag{B.23}$$

Then the semi-infinite constraint (B.21) is satisfied if, and only if, there exist solutions $(\mathbf{a}_l, \mathbf{b}_l, w_l) \in \mathscr{K}^*$, for all $l \in \mathscr{L}$, such that the constraints in (B.23) are satisfied and $-w_l$ is not less than $t_l(\boldsymbol{\nu}) - \lambda_0 - \lambda_3$. Using this dual characterization, we can rewrite (5.28) in the following way:

$$\begin{aligned} \inf \quad & \lambda_3 \\ \text{s.t.} \quad & \lambda_0 + \boldsymbol{\lambda}_1^\top \boldsymbol{\mu} + \boldsymbol{\lambda}_2^\top \boldsymbol{\sigma} \leqslant 0, \\ & \lambda_0 + \boldsymbol{\lambda}_1^\top \mathbf{x} + \boldsymbol{\lambda}_2^\top \mathbf{u} \geqslant 0, && \forall (\mathbf{x}, \mathbf{u}) \in \overline{\mathscr{C}}, \\ & \lambda_0 + \lambda_3 - t_l(\mathbf{x}) - w_l \geqslant 0, && \forall l \in \mathscr{L}, \\ & \boldsymbol{\lambda}_1 - \mathbf{s}_l(\boldsymbol{\nu}) - \mathbf{a}_l = \mathbf{0}, && \forall l \in \mathscr{L}, \\ & \boldsymbol{\lambda}_2 - \mathbf{b}_l = \mathbf{0}, && \forall l \in \mathscr{L}, \\ & (\mathbf{a}_l, \mathbf{b}_l, w_l) \in \mathscr{K}^*, && \forall l \in \mathscr{L}, \\ & \boldsymbol{\nu} \in \mathscr{V}, \ \lambda_0 \in \mathbb{R}, \ \boldsymbol{\lambda}_1 \in \mathbb{R}^n, \ \boldsymbol{\lambda}_2 \in \mathbb{R}^m_+, \ \lambda_3 \in \mathbb{R}. \end{aligned} \tag{B.24}$$

Since $\overline{\mathscr{C}}$ is also conic representable, an analogous argument enables us to reformulate the first semi-infinite constraint, reducing (B.24) to a finite-dimensional conic optimization problem. Then, the second claim follows. □

We use the following result to derive the LMI reformulations for the Chebyshev ambiguity set $\mathscr{P}_{(\mu,\Sigma)}$.

LEMMA B.1 **(S-Lemma, 176).** *Consider two quadratic functions of* $\mathbf{x} \in \mathbb{R}^n$, $q_i(\mathbf{x}) = \mathbf{x}^\top \mathrm{C}_i \mathbf{x} + 2\mathbf{c}_i^\top \mathbf{x} + \bar{c}_i$, $i = 0, 1$, *with* $q_1(\overline{\mathbf{x}}) > 0$ *for some* $\overline{\mathbf{x}}$. *Then*

$$q_0(\mathbf{x}) \geqslant 0 \quad \forall \mathbf{x} : q_1(\mathbf{x}) \geqslant 0$$

if, and only if, there exists $\tau \geqslant 0$ *such that*

$$\begin{pmatrix} \bar{c}_0 & \mathbf{c}_0^\top \\ \mathbf{c}_0 & \mathrm{C}_0 \end{pmatrix} - \tau \begin{pmatrix} \bar{c}_1 & \mathbf{c}_1^\top \\ \mathbf{c}_1 & \mathrm{C}_1 \end{pmatrix} \in \mathbb{S}_+^{n+1}.$$

Proof of Corollary 5.11. By the S-Lemma, we have that

$$\lambda_0 + \boldsymbol{\lambda}_1^\top \mathbf{x} + \mathbf{x}^\top \boldsymbol{\Lambda} \mathbf{x} \geqslant 0,\ \forall \mathbf{x} : \mathbf{c}^\top \mathbf{x} \geqslant \bar{c} \iff \exists \tau \geqslant 0 : \begin{bmatrix} \lambda_0 + \tau \bar{c} & \frac{1}{2}(\boldsymbol{\lambda}_1 - \tau \mathbf{c})^\top \\ \frac{1}{2}(\boldsymbol{\lambda}_1 - \tau \mathbf{c}) & \boldsymbol{\Lambda} \end{bmatrix} \succcurlyeq \mathbf{0}.$$

Analogously, for the second semi-infinite constraint,

$$\begin{aligned} &\lambda_0 + (\boldsymbol{\lambda}_1 - \mathbf{s}(\boldsymbol{\nu}))^\top \mathbf{x} + \mathbf{x}^\top \boldsymbol{\Lambda} \mathbf{x} + \lambda_3 - t_l(\boldsymbol{\nu}) \geqslant 0,\ \forall \mathbf{x} : \mathbf{c}^\top \mathbf{x} \leqslant \bar{c} \\ \iff \exists \chi_l \geqslant 0 : &\begin{bmatrix} \lambda_0 - \chi_l \bar{c} & \frac{1}{2}(\boldsymbol{\lambda}_1 - \mathbf{s}_l(\boldsymbol{\nu}) + \chi_l \mathbf{c})^\top \\ \frac{1}{2}(\boldsymbol{\lambda}_1 - \mathbf{s}_l(\boldsymbol{\nu}) + \chi_l \mathbf{c}) & \boldsymbol{\Lambda} \end{bmatrix} \succcurlyeq \mathbf{0}. \end{aligned}$$

These LMIs yield the following semidefinite programming problem:

$$\begin{aligned} \inf \quad & \lambda_3 \\ \text{s.t.} \quad & \lambda_0 + \boldsymbol{\lambda}_1^\top \boldsymbol{\mu} + \langle \boldsymbol{\Lambda}, \Sigma \rangle \leqslant 0, \\ & \begin{bmatrix} \lambda_0 + \tau \bar{c} & \frac{1}{2}(\boldsymbol{\lambda}_1 - \tau \mathbf{c})^\top \\ \frac{1}{2}(\boldsymbol{\lambda}_1 - \tau \mathbf{c}) & \boldsymbol{\Lambda} \end{bmatrix} \succcurlyeq \mathbf{0}, \\ & \begin{bmatrix} \lambda_0 + \lambda_3 - t_l(\boldsymbol{\nu}) - \chi_l \bar{c} & \frac{1}{2}(\boldsymbol{\lambda}_1 - \mathbf{s}_l(\boldsymbol{\nu}) + \chi_l \mathbf{c})^\top \\ \frac{1}{2}(\boldsymbol{\lambda}_1 - \mathbf{s}_l(\boldsymbol{\nu}) + \chi_l \mathbf{c}) & \boldsymbol{\Lambda} \end{bmatrix} \succcurlyeq \mathbf{0}, \qquad \forall l \in \mathscr{L}, \\ & \boldsymbol{\nu} \in \mathscr{V},\ \lambda_0 \in \mathbb{R},\ \boldsymbol{\lambda}_1 \in \mathbb{R}^n,\ \boldsymbol{\Lambda} \in \mathbb{S}_+^n,\ \lambda_3 \in \mathbb{R},\ \tau \in \mathbb{R}_+,\ \boldsymbol{\chi} \in \mathbb{R}_+^{|\mathscr{L}|}, \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ denotes the trace inner product. □

Proof of Corollary 5.12. The model with MAD information follows from defining separate constraints for the positive and negative parts of the absolute value terms. This yields

$$\begin{aligned}
\inf_{\boldsymbol{\nu},\lambda_0,\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\lambda_3} \quad & \lambda_3 \\
\text{s.t.} \quad & \lambda_0 + \boldsymbol{\lambda}_1^\top \mathbf{m} + \boldsymbol{\lambda}_2^\top \mathbf{f} \leqslant 0, \\
& \lambda_0 + \boldsymbol{\lambda}_1^\top \mathbf{x} + \boldsymbol{\lambda}_2^\top \mathbf{d}(\mathbf{x}) \geqslant 0, && \forall \mathbf{x} : \mathbf{c}^\top \mathbf{x} \geqslant \bar{c}, \\
& \lambda_0 + \boldsymbol{\lambda}_1^\top \mathbf{x} - \boldsymbol{\lambda}_2^\top \mathbf{d}(\mathbf{x}) \geqslant 0, && \forall \mathbf{x} : \mathbf{c}^\top \mathbf{x} \geqslant \bar{c}, \\
& \lambda_0 + \boldsymbol{\lambda}_1^\top \mathbf{x} + \boldsymbol{\lambda}_2^\top \mathbf{d}(\mathbf{x}) + \lambda_3 \geqslant \mathbf{s}_l(\boldsymbol{\nu})^\top \mathbf{x} + t_l(\boldsymbol{\nu}), && \forall \mathbf{x} : \mathbf{c}^\top \mathbf{x} \leqslant \bar{c},\ \forall l \in \mathscr{L}, \\
& \lambda_0 + \boldsymbol{\lambda}_1^\top \mathbf{x} - \boldsymbol{\lambda}_2^\top \mathbf{d}(\mathbf{x}) + \lambda_3 \geqslant \mathbf{s}_l(\boldsymbol{\nu})^\top \mathbf{x} + t_l(\boldsymbol{\nu}), && \forall \mathbf{x} : \mathbf{c}^\top \mathbf{x} \leqslant \bar{c},\ \forall l \in \mathscr{L}, \\
& \boldsymbol{\nu} \in \mathscr{V},\ \lambda_0 \in \mathbb{R},\ \boldsymbol{\lambda}_1 \in \mathbb{R}^n,\ \boldsymbol{\lambda}_2 \in \mathbb{R}_+^{n^2},\ \lambda_3 \in \mathbb{R},
\end{aligned}$$

where $\mathbf{d}(\mathbf{x}) = \mathbf{m}_0 + \mathbf{D}\mathbf{x}$ describes the affine functions $X_i - m_i$ and $(X_i \pm X_j) - (m_i \pm m_j)$. Let us focus on the first semi-infinite constraint. Using the fact that $\mathbf{d}$ is an affine, multi-valued function of $\mathbf{x}$, the first semi-infinite constraint can be rewritten as

$$\lambda_0 + \boldsymbol{\lambda}_2^\top \mathbf{m}_0 + (\boldsymbol{\lambda}_1 + \mathbf{D}^\top \boldsymbol{\lambda}_2)^\top \mathbf{x} \geqslant 0.$$

Then, standard LP duality yields a finite-dimensional linear reformulation since

$$\begin{aligned}
&\lambda_0 + \boldsymbol{\lambda}_2^\top \mathbf{m}_0 + (\boldsymbol{\lambda}_1 + \mathbf{D}^\top \boldsymbol{\lambda}_2)^\top \mathbf{x} \geqslant 0, \forall \mathbf{x} : \mathbf{c}^\top \mathbf{x} \geqslant \bar{c} \\
&\iff \lambda_0 + \boldsymbol{\lambda}_2^\top \mathbf{m}_0 + \min_{\mathbf{x} : \mathbf{c}^\top \mathbf{x} \geqslant \bar{c}} (\boldsymbol{\lambda}_1 + \mathbf{D}^\top \boldsymbol{\lambda}_2)^\top \mathbf{x} \geqslant 0 \\
&\iff \lambda_0 + \boldsymbol{\lambda}_2^\top \mathbf{m}_0 + \max_{\tau \leqslant 0} \left\{ \bar{c}\tau \ : \ \tau \mathbf{c} = \boldsymbol{\lambda}_1 + \mathbf{D}^\top \boldsymbol{\lambda}_2 \right\} \geqslant 0 \\
&\iff \exists \tau \leqslant 0 : \lambda_0 + \boldsymbol{\lambda}_2^\top \mathbf{m}_0 + \bar{c}\tau \geqslant 0,\ \tau \mathbf{c} = \boldsymbol{\lambda}_1 + \mathbf{D}^\top \boldsymbol{\lambda}_2.
\end{aligned}$$

A similar argument applies to the other semi-infinite constraints. Therefore, all robust counterparts can be rewritten in terms of linear inequalities, yielding the result. □

B.4. Remaining proofs Chapter 7

Proof of Proposition 7.8. We first consider the marginal moment model for the uncertain service times discussed in Mak et al. [149]. Assume knowledge of only the marginal moments of the joint distribution and consider the min-max problem

$$\min_{\mathbf{s} \in \mathscr{S}} \max_{\mathbb{P} \in \mathscr{F}_{(\mathscr{D},\mathscr{Q})}} \mathbb{E}_{\mathbb{P}}[f_n(\mathbf{s}, \mathbf{X})], \tag{B.25}$$

where the probability measure $\mathbb{P}$ is chosen from the marginal moment ambiguity set $\mathscr{F}_{(\mathscr{D},\mathscr{Q})}$ with $\mathscr{D} = [a_1, b_1] \times \ldots \times [a_n, b_n]$ the Cartesian product of the supports of the individual service times and $\mathscr{Q}$ an information set that comprises vectors $(\mathbb{E}[h_1(X_k)], \ldots, \mathbb{E}[h_m(X_k)]) = (q_k^{(1)}, \ldots, q_k^{(m)})$, $\forall k$, containing information on m generalized moments for all service time distributions. For the setting with traditional power moments (i.e., $h_l(x) = x^l$), Mak et al. [149] show that problem

(B.25) is equivalent to

$$\begin{aligned}
\min_{\mathbf{s},\boldsymbol{\zeta},\boldsymbol{\lambda}} & \sum_{k=1}^{n} \zeta_k + \sum_{k=1}^{n}\sum_{l=1}^{m} q_k^{(l)} \lambda_k^{(l)} \\
\text{s.t.} & \sum_{i=k}^{\min\{j,n\}} \max_{x_i \in [a_i,b_i]} \left((x_i - s_i)\,\pi_{ij} - \sum_{l=1}^{m} \lambda_k^{(l)} h_l(x_i) - \zeta_i \right) \leqslant 0, \quad \text{for } 1 \leqslant k \leqslant n,\ k \leqslant j \leqslant n+1, \\
& s \in \mathcal{S},
\end{aligned} \tag{B.26}$$

with π_{ij} as stated in Proposition 7.8. These parameters characterize the extreme points of the polyhedron Λ that describes the feasible set corresponding to the dual of LP (7.3), see Mak et al. [149, Section 2]. Recall that (7.3) determines the total costs $f_n(\mathbf{s}, \mathbf{x})$ for a given realization of service times $\mathbf{x}$. Close inspection of the proof of Proposition 2 in Mak et al. [149] shows that this result also holds for generalized moment functions such as the MAD dispersion function $h(x) = |x - \mu|$. To see this, note that strong duality for generalized moment problems holds for the inner maximization in (B.25) when the generalized moment vector lies in the interior of the set of feasible moment vectors; see, e.g., [177]. For mean-MAD ambiguity, this specializes to the conditions $\mu_k \in (a_k, b_k)$, $d_k \in (0, \frac{2(b_k-\mu_k)(\mu_k-a_k)}{(b_k-a_k)})$, $\forall k$. By analyzing the dual problem of the semi-infinite LP $\max_{\mathbb{P}\in\mathcal{F}_{(\mathcal{D},\mathcal{Q})}} \mathbb{E}_{\mathbb{P}}[f_n(\mathbf{s}, \mathbf{X})]$, an analog of Lemma 1 in Mak et al. [149] for generalized moments can be derived. Then, to obtain problem (B.26), the proof of Proposition 2 in Mak et al. [149] exploits the structure of the polyhedron Λ but does not require the functionals $h_l(x)$ to be of a specific form. We thus conclude that equivalence between (B.25) and (B.26) also holds for generalized moments.

Specializing (B.26) to the case where the mean and MAD are known, (B.25) gives rise to

$$\begin{aligned}
\min_{\mathbf{s},\boldsymbol{\zeta},\boldsymbol{\lambda}^{(1)},\boldsymbol{\lambda}^{(2)}} & \sum_{i=1}^{n} \left(\zeta_i + \mu_i \lambda_i^{(1)} + d_i \lambda_i^{(2)}\right) \\
\text{s.t.} & \sum_{i=k}^{\min\{n,j\}} \max_{x_i \in [a_i,b_i]} \left\{ (x_i - s_i)\,\pi_{ij} - \lambda_i^{(1)} x_i - \lambda_i^{(2)} |x_i - \mu_i| - \zeta_i \right\} \leqslant 0, \\
& \qquad \forall 1 \leqslant k \leqslant n, 1 \leqslant k \leqslant j \leqslant n+1, \\
& \mathbf{s} \in \mathcal{S}.
\end{aligned}$$

Since the function to be maximized is piecewise linear, the optimal solution lies either on the boundary points a_i, b_i or at the kink point μ_i. Hence,

$$\begin{aligned}
& \max_{x_i \in [a_i,b_i]} \{(x_i - s_i)\pi_{ij} - \lambda_i^{(1)} x_i - \lambda_i^{(2)} |x_i - \mu_i| - \zeta_i\} \\
&= \max_{x_i \in \{a_i,\mu_i,b_i\}} \{(x_i - s_i)\pi_{ij} - \lambda_i^{(1)} x_i - \lambda_i^{(2)} |x_i - \mu_i| - \zeta_i\}.
\end{aligned}$$

Using the auxiliary variable ξ_{ij} to replace the maximum operator by linear inequalities, we obtain

$$\begin{aligned}
&\sum_{i=k}^{\min\{n,j\}} \xi_{ij} \leqslant \sum_{i=k}^{\min\{n,j\}} \zeta_i + s_i \pi_{ij}, && \text{for } 1 \leqslant k \leqslant n, 1 \leqslant k \leqslant j \leqslant n+1,\\
&\xi_{ij} \geqslant \left(\pi_{ij} - \lambda_i^{(1)}\right) b_i - \lambda_i^{(2)}(b_i - \mu_i), && \text{for } 1 \leqslant i \leqslant n,\ i \leqslant j \leqslant n+1,\\
&\xi_{ij} \geqslant \left(\pi_{ij} - \lambda_i^{(1)}\right) \mu_i, && \text{for } 1 \leqslant i \leqslant n,\ i \leqslant j \leqslant n+1,\\
&\xi_{ij} \geqslant \left(\pi_{ij} - \lambda_i^{(1)}\right) a_i - \lambda_i^{(2)}(\mu_i - a_i), && \text{for } 1 \leqslant i \leqslant n,\ i \leqslant j \leqslant n+1.
\end{aligned}$$

This yields (7.16). □

B.5. Remaining proofs Chapter 8

Proof of Lemma 8.4. First, note that (8.12) is well defined because the moment-generating function of Y is finite valued. The objective function $\phi(x)$ is convex since it reflects the expectation of a convex function, i.e., $\mathbb{E}[((Y-x)^+)^m]$. From integration by parts, it follows that

$$\begin{aligned}
\int_x^{b_Y} h(y-x)\mathrm{d}F(y) &= -\int_x^{b_Y} h(y-x)\mathrm{d}(1-F(y))\\
&= 0 + \int_x^{b_Y} h'(y-x)(1-F(y))\mathrm{d}y
\end{aligned}$$

in which the second equality follows from $h(0)=0,\ F(b_Y)=1$. Taking the derivative,

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_x^{b_Y} h(y-x)\mathrm{d}F(y) = \frac{\mathrm{d}}{\mathrm{d}x}\int_x^{b_Y} h'(y-x)\,(1-F(y))\,\mathrm{d}y = -\int_x^{b_Y} h''(y-x)\,(1-F(y))\,\mathrm{d}y,$$

where the final line follows from Leibniz's integral rule and the observation that $h'(0)=0$. After some rewriting, this yields

$$\begin{aligned}
-\int_x^{b_Y} h''(s-x)\,(1-F(s))\,\mathrm{d}s &= -\int_x^{b_Y} h''(s-x)\int_s^{b_Y} \mathrm{d}F(y)\,\mathrm{d}s\\
&= -\int_x^{b_Y}\int_x^{y} h''(s-x)\mathrm{d}s\,\mathrm{d}F(y)\\
&= -\int_x^{b_Y} h'(y-x)\mathrm{d}F(y).
\end{aligned}$$

Hence,

$$\phi'(x) = -\int_x^{b_Y} h'(y-x)\mathrm{d}F(y),$$

which is a continuous function of x for $h(x)=x^m,\ m \geqslant 2$, supporting the claim that $\phi(x)$ is continuously differentiable. It then remains to show that $\phi'(x)$ is concave. By differentiating a second time, one obtains

$$\phi''(x) = \int_x^{b_Y} h''(y-x)\mathrm{d}F(y),$$

which is clearly nonincreasing in x for $h(x)=x^m,\ m \geqslant 2.$ As a consequence, the derivative $\phi'(x)$ is concave. □

Bibliography

[1] Abate, J., Choudhury, G., and Whitt, W. (1993). Calculation of the GI/G/1 waiting-time distribution and its cumulants from Pollaczek's formulas. *Archiv fur Elektronik und Ubertragungstechnik (International Journal of Electronics and Communication)*, 47(5/6):311–321.

[2] Abate, J. and Whitt, W. (1992). The Fourier-series method for inverting transforms of probability distributions. *Queueing Systems*, 10:5–88.

[3] Abate, J. and Whitt, W. (1995). Numerical inversion of Laplace transforms of probability distributions. *INFORMS Journal on Computing*, 7(1):36–43.

[4] Abdel-Malek, L., Montanari, R., and Morales, L. C. (2004). Exact, approximate, and generic iterative models for the multi-product newsboy problem with budget constraint. *International Journal of Production Economics*, 91(2):189–198.

[5] Ardestani-Jaafari, A. and Delage, E. (2016). Robust optimization of sums of piecewise linear functions with application to inventory problems. *Operations Research*, 64(2):474–494.

[6] Arrow, K. J., Harris, T., and Marschak, J. (1951). Optimal inventory policy. *Econometrica: Journal of the Econometric Society*, 19(3):250–272.

[7] Asmussen, S. (2003). *Applied Probability and Queues*. Springer-Verlag, New York, 2nd edition.

[8] Asmussen, S. and Albrecher, H. (2010). *Ruin Probabilities*. World Scientific, Singapore, 2nd edition.

[9] Avramidis, A. N., Deslauriers, A., and L'Ecuyer, P. (2004). Modeling daily arrivals to a telephone call center. *Management Science*, 50(7):896–908.

[10] Bailey, N. T. (1952). A study of queues and appointment systems in hospital out-patient departments, with special reference to waiting-times. *Journal of the Royal Statistical Society: Series B (Methodological)*, 14(2):185–199.

[11] Balas, E. and Zemel, E. (1980). An algorithm for large zero-one knapsack problems. *Operations Research*, 28(5):1130–1154.

[12] Ban, G.-Y. and Rudin, C. (2019). The big data newsvendor: Practical insights from machine learning. *Operations Research*, 67(1):90–108.

[13] Begen, M. A., Levi, R., and Queyranne, M. (2012). A sampling-based approach to appointment scheduling. *Operations Research*, 60(3):675–681.

[14] Begen, M. A. and Queyranne, M. (2011). Appointment scheduling with discrete random durations. *Mathematics of Operations Research*, 36(2):240–257.

[15] Ben-Tal, A., Den Hertog, D., De Waegenaere, A., Melenberg, B., and Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357.

[16] Ben-Tal, A., Den Hertog, D., and Vial, J.-P. (2015). Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, 149:265–299.

[17] Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). *Robust Optimization*. Princeton University Press, Princeton, NJ.

[18] Ben-Tal, A., Goryashko, A., Guslitzer, E., and Nemirovski, A. (2004). Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376.

[19] Ben-Tal, A. and Hochman, E. (1972). More bounds on the expectation of a convex function of a random variable. *Journal of Applied Probability*, 9(4):803–812.

[20] Ben-Tal, A. and Hochman, E. (1976). Stochastic programs with incomplete information. *Operations Research*, 24(2):336–347.

[21] Ben-Tal, A. and Hochman, E. (1985). Approximation of expected returns and optimal decisions under uncertainty using mean and mean absolute deviation. *Zeitschrift für Operations Research*, 29(7):285–300.

[22] Ben-Tal, A. and Nemirovski, A. (1998). Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805.

[23] Ben-Tal, A. and Nemirovski, A. (2001). *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. SIAM, Philadelphia.

[24] Bergman, R., Daley, D. J., Rolski, T., and Stoyan, D. (1979). Bounds for cumulants of waiting-times in GI/GI/1 queues. *Mathematische Operationsforschung und Statistik. Series Optimization*, 10(2):257–263.

[25] Bertsimas, D., Doan, X. V., Natarajan, K., and Teo, C.-P. (2010). Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research*, 35(3):580–602.

[26] Bertsimas, D. and Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science*, 66(3):1025–1044.

[27] Bertsimas, D. and Natarajan, K. (2007). A semidefinite optimization approach to the steady-state analysis of queueing systems. *Queueing Systems*, 56(1):27–39.

[28] Bertsimas, D. and Popescu, I. (2002). On the relation between option and stock prices: A convex optimization approach. *Operations Research*, 50(2):358–374.

[29] Bertsimas, D. and Popescu, I. (2005). Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization*, 15(3):780–804.

[30] Bertsimas, D., Sim, M., and Zhang, M. (2018). Adaptive distributionally robust optimization. *Management Science*, 65(2):604–618.

[31] Bienaymé, I.-J. (1853). Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés. *Comptes Rendus de l'Académie des Sciences*, 37:309–324.

[32] Birge, J. R. and Dulá, J. H. (1991). Bounding separable recourse functions with limited distribution information. *Annals of Operations Research*, 30(1):277–298.

[33] Birge, J. R. and Louveaux, F. (2011). *Introduction to Stochastic Programming*. Springer, New York.

[34] Birge, J. R. and Wets, R. J.-B. (1987). Computing bounds for stochastic programming problems by means of a generalized moment problem. *Mathematics of Operations Research*, 12(1):149–162.

[35] Boonstra, G., van Eekelen, W. J. E. C., and van Leeuwaarden, J. S. H. (2023). Robust knapsack ordering for a partially-informed newsvendor with budget constraint. Preprint available at https://arxiv.org/abs/2301.02662.

[36] Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization.* Cambridge University Press, Cambridge, UK.

[37] Bradley, J. R. and Glynn, P. W. (2002). Managing capacity and inventory jointly in manufacturing systems. *Management Science*, 48(2):273–288.

[38] Brockett, P. L. and Cox Jr, S. H. (1985). Insurance calculations using incomplete information. *Scandinavian Actuarial Journal*, 1985(2):94–108.

[39] Casella, G. and Berger, R. L. (2002). *Statistical Inference.* Duxbury Press, Belmont, CA, 2nd edition.

[40] Cayirli, T. and Veral, E. (2003). Outpatient scheduling in health care: A review of literature. *Production and Operations Management*, 12(4):519–549.

[41] Charnes, A. and Cooper, W. W. (1962). Programming with linear fractional functionals. *Naval Research Logistics Quarterly*, 9(3-4):181–186.

[42] Chebyshev, P. L. (1867). Des valeurs moyennes. *Journal de Mathématiques Pures et Appliquées*, 12:177–184.

[43] Chebyshev, P. L. (1874). Sur les valeurs limites des intégrales. *Journal de Mathématiques Pures et Appliquées*, 19:157–160.

[44] Chen, B. P. and Henderson, S. G. (2001). Two issues in setting call centre staffing levels. *Annals of Operations Research*, 108:175–192.

[45] Chen, H., Hu, M., and Perakis, G. (2022). Distribution-free pricing. *Manufacturing & Service Operations Management*, 24(4):1939–1958.

[46] Chen, R. R. and Robinson, L. W. (2014). Sequencing and scheduling appointments with potential call-in patients. *Production and Operations Management*, 23(9):1522–1538.

[47] Chen, Y. and Hasenbein, J. J. (2017). Staffing large-scale service systems with distributional uncertainty. *Queueing Systems*, 87(1):55–79.

[48] Chen, Y. and Hasenbein, J. J. (2020). Knowledge, congestion, and economics: Parameter uncertainty in Naor's model. *Queueing Systems*, 96(1):83–99.

[49] Chen, Y. and Whitt, W. (2019). Extremal GI/GI/1 queues given two moments. Working paper, Columbia University, New York.

[50] Chen, Y. and Whitt, W. (2020). Algorithms for the upper bound mean waiting time in the GI/GI/1 queue. *Queueing Systems*, 94(3–4):327–356.

[51] Chen, Y. and Whitt, W. (2021). Extremal cumulants in GI/GI/1 queues given two moments: Exploiting Tchebycheff systems. *Queueing Systems*, 97(1-2):101–124.

[52] Chen, Y., Xu, M., and Zhang, Z. G. (2009). A risk-averse newsvendor model under the CVaR criterion. *Operations Research*, 57(4):1040–1044.

[53] Chen, Z., Sim, M., and Xiong, P. (2020). Robust stochastic optimization made easy with RSOME. *Management Science*, 66(8):3329–3339.

[54] Chen, Z., Sim, M., and Xu, H. (2019). Distributionally robust optimization with infinitely constrained ambiguity sets. *Operations Research*, 67(5):1328–1344.

[55] Choi, S., Ruszczyński, A., and Zhao, Y. (2011). A multiproduct risk-averse newsvendor with law-invariant coherent measures of risk. *Operations Research*, 59(2):346–364.

[56] Chung, K. (2001). *A Course in Probability Theory.* Academic Press, London.

[57] Cohen, J. W. (1982). *The Single Server Queue*. North-Holland Publishing Co., Amsterdam, 2nd edition.

[58] Cox, S. H. (1991). Bounds on expected values of insurance payments and option prices. *Transactions of the Society of Actuaries*, 43:231–260.

[59] Cozman, F. G. (1999). Calculation of posterior bounds given convex sets of prior probability measures and likelihood functions. *Journal of Computational and Graphical Statistics*, 8(4):824–838.

[60] Daley, D. J., Kreinin, A. Y., and Trengove, C. D. (1992). Inequalities concerning the waiting-time in single-server queues: A survey. In Bhat, U. N. and Basawa, I. V., editors, *Queueing and Related Models*, pages 177–223. Clarendon Press, Oxford.

[61] Das, B., Dhara, A., and Natarajan, K. (2021). On the heavy-tail behavior of the distributionally robust newsvendor. *Operations Research*, 69(4):1077–1099.

[62] de Kemp, M. A., Mandjes, M., and Olver, N. (2021). Performance of the smallest-variance-first rule in appointment sequencing. *Operations Research*, 69(6):1909–1935.

[63] de Klerk, E., Kuhn, D., and Postek, K. (2020). Distributionally robust optimization with polynomial densities: Theory, models and algorithms. *Mathematical Programming*, 181:265–296.

[64] De Schepper, A. and Heijnen, B. (1995). General restrictions on tail probabilities. *Journal of Computational and Applied Mathematics*, 64(1-2):177–188.

[65] De Vylder, F. and Goovaerts, M. J. (1982). Analytical best upper bounds on stop-loss premiums. *Insurance: Mathematics and Economics*, 1(3):163–175.

[66] Delage, E. and Ye, Y. (2010a). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612.

[67] Delage, E. and Ye, Y. (2010b). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612.

[68] Denton, B. and Gupta, D. (2003). A sequential bounding approach for optimal appointment scheduling. *IIE Transactions*, 35(11):1003–1016.

[69] Denton, B., Viapiano, J., and Vogl, A. (2007). Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science*, 10(1):13–24.

[70] Dokov, S. P. and Morton, D. P. (2005). Second-order lower bounds on the expectation of a convex function. *Mathematics of Operations Research*, 30(3):662–677.

[71] Dupačová, J. (1987). The minimax approach to stochastic programming and an illustrative application. *Stochastics: An International Journal of Probability and Stochastic Processes*, 20(1):73–88.

[72] Dyer, M. and Stougie, L. (2006). Computational complexity of stochastic programming problems. *Mathematical Programming*, 106:423–432.

[73] Eckberg Jr, A. (1977). Sharp bounds on Laplace-Stieltjes transforms, with applications to various queueing problems. *Mathematics of Operations Research*, 2(2):135–142.

[74] Edelson, N. and Hilderbrand, D. (1975). Congestion tolls for Poisson queuing processes. *Econometrica: Journal of the Econometric Society*, 43(1):81–92.

[75] Elmachtoub, A. N., Gupta, V., and Hamilton, M. L. (2021). The value of personalized pricing. *Management Science*, 67(10):6055–6070.

[76] Erdogan, S. A. and Denton, B. (2013). Dynamic appointment scheduling of a stochastic server with uncertain demand. *INFORMS Journal on Computing*, 25(1):116–132.

[77] Erlebacher, S. J. (2000). Optimal and heuristic solutions for the multi-item newsvendor problem with a single capacity constraint. *Production and Operations Management*, 9(3):303–318.

[78] Esteban-Pérez, A. and Morales, J. M. (2022). Distributionally robust stochastic programs with side information based on trimmings. *Mathematical Programming*, 195(1-2):1069–1105.

[79] Feller, W. (1971). *An Introduction to Probability Theory and its Applications. Vol. II.* John Wiley & Sons Inc., New York, 2nd edition.

[80] Fisher, M. and Raman, A. (1996). Reducing the cost of demand uncertainty through accurate response to early sales. *Operations Research*, 44(1):87–99.

[81] Foss, S., Konstantopoulos, T., and Zachary, S. (2007). Discrete and continuous time modulated random walks with heavy-tailed increments. *Journal of Theoretical Probability*, 20(3):581–612.

[82] Fridgeirsdottir, K. and Chiu, S. (2005). A note on convexity of the expected delay cost in single-server queues. *Operations Research*, 53(3):568–570.

[83] Gallego, G. (1992). A minmax distribution free procedure for the (Q, R) inventory model. *Operations Research Letters*, 11(1):55–60.

[84] Gallego, G. and Moon, I. (1993). The distribution free newsboy problem: Review and extensions. *Journal of the Operational Research Society*, 44(8):825–834.

[85] Gao, R. and Kleywegt, A. (2022). Distributionally robust stochastic optimization with Wasserstein distance. *Mathematics of Operations Research*.

[86] Gauss, C. F. (1821). Theoria combinationis observationum erroribus minimis obnoxiae, pars prior. *Commentationes Societatis Regiae Scientiarum Gottingensis Recentiores*, 33:523–533.

[87] Ghosal, S. and Wiesemann, W. (2020). The distributionally robust chance-constrained vehicle routing problem. *Operations Research*, 68(3):716–732.

[88] Giannakopoulos, Y., Poças, D., and Tsigonias-Dimitriadis, A. (2020). Robust revenue maximization under minimal statistical information. In *International Conference on Web and Internet Economics*, pages 177–190, Cham, Switzerland. Springer.

[89] Glasserman, P. (1997). Bounds and asymptotics for planning critical safety stocks. *Operations Research*, 45(2):244–257.

[90] Gorissen, B. L. (2015). Robust fractional programming. *Journal of Optimization Theory and Applications*, 166(2):508–528.

[91] Grassmann, W. (1983). The convexity of the mean queue size of the M/M/c queue with respect to the traffic intensity. *Journal of Applied Probability*, 20(4):916–919.

[92] Grechuk, B., Molyboha, A., and Zabarankin, M. (2010). Chebyshev inequalities with law invariant deviation measures. *Probability in the Engineering and Informational Sciences*, 24(1):145–170.

[93] Green, L. V. and Savin, S. (2008). Reducing delays for medical appointments: A queueing approach. *Operations Research*, 56(6):1526–1538.

[94] Gupta, V. and Osogami, T. (2011). On Markov–Krein characterization of the mean waiting time in M/G/K and other queueing systems. *Queueing Systems*, 68(3):339–352.

[95] Hadley, G. and Whitin, T. M. (1963). *Analysis of Inventory Systems.* Prentice-Hall, Englewood Cliffs, NJ.

[96] Hajek, B. (1983). The proof of a folk theorem on queuing delay with applications to routing in networks. *Journal of the ACM (JACM)*, 30(4):834–851.

[97] Han, Q., Du, D., and Zuluaga, L. F. (2014). A risk-and ambiguity-averse extension of the max-min newsvendor order formula. *Operations Research*, 62(3):535–542.

[98] Han, S., Tao, M., Topcu, U., Owhadi, H., and Murray, R. M. (2015). Convex optimal uncertainty quantification. *SIAM Journal on Optimization*, 25(3):1368–1387.

[99] Hanasusanto, G. A., Kuhn, D., and Wiesemann, W. (2016). A comment on "computational complexity of stochastic programming problems". *Mathematical Programming*, 159:557–569.

[100] Hanasusanto, G. A., Roitch, V., Kuhn, D., and Wiesemann, W. (2015). A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming*, 151(1):35–62.

[101] Hanasusanto, G. A., Roitch, V., Kuhn, D., and Wiesemann, W. (2017). Ambiguous joint chance constraints under mean and dispersion information. *Operations Research*, 65(3):751–767.

[102] Haneveld, W. K. K. (2013). *Duality in Stochastic Linear and Dynamic Programming*. Springer Science & Business Media, Berlin.

[103] Harel, A. (1990a). Convexity properties of the Erlang loss formula. *Operations Research*, 38(3):499–505.

[104] Harel, A. (1990b). Convexity results for single-server queues and for multiserver queues with constant service times. *Journal of Applied Probability*, 27(2):465–468.

[105] Harel, A. (2010). Sharp and simple bounds for the Erlang delay and loss formulae. *Queueing Systems*, 64(2):119–143.

[106] Harel, A. and Zipkin, P. (1987). The convexity of a general performance measure for multiserver queues. *Journal of Applied Probability*, 24(3):725–736.

[107] Harrison, J. M. and Zeevi, A. (2005). A method for staffing large call centers based on stochastic fluid models. *Manufacturing & Service Operations Management*, 7(1):20–36.

[108] Hassin, R. (2016). *Rational Queueing*. CRC press, Boca Raton, FL.

[109] Hassin, R., Haviv, M., and Oz, B. (2023). Strategic behavior in queues with arrival rate uncertainty. *European Journal of Operational Research*, 309(1):217–224.

[110] Hassin, R. and Mendel, S. (2008). Scheduling arrivals to queues: A single-server model with no-shows. *Management Science*, 54(3):565–572.

[111] Heemskerk, M., van Leeuwaarden, J., and Mandjes, M. (2017). Scaling limits for infinite-server systems in a random environment. *Stochastic Systems*, 7(1):1–31.

[112] Heijnen, B. (1990). Best upper and lower bounds on modified stop loss premiums in case of known range, mode, mean and variance of the original risk. *Insurance: Mathematics and Economics*, 9(2-3):207–220.

[113] Hettich, R. and Kortanek, K. O. (1993). Semi-infinite programming: Theory, methods, and applications. *SIAM Review*, 35(3):380–429.

[114] Hoeffding, W. (1955). The extrema of the expected value of a function of independent random variables. *The Annals of Mathematical Statistics*, 26(2):268–275.

[115] Hoeffding, W. and Shrikhande, S. (1955). Bounds for the distribution function of a sum of independent, identically distributed random variables. *The Annals of Mathematical Statistics*, 26(3):439–449.

[116] Illés, T. and Terlaky, T. (2002). Pivot versus interior point methods: Pros and cons. *European Journal of Operational Research*, 140(2):170–190.

[117] Isii, K. (1962). On sharpness of Tchebycheff-type inequalities. *Annals of the Institute of Statistical Mathematics*, 14(1):185–197.

[118] Isii, K. et al. (1960). The extrema of probability determined by generalized moments (i) bounded random variables. *Annals of the Institute of Statistical Mathematics*, 12(2):119–134.

[119] Jagers, A. A. and van Doorn, E. A. (1986). On the continued Erlang loss function. *Operations Research Letters*, 5(1):43–46.

[120] Jansen, K., Haezendonck, J., and Goovaerts, M. (1986). Upper bounds on stop-loss premiums in case of known moments up to the fourth order. *Insurance: Mathematics and Economics*, 5(4):315–334.

[121] Janssen, A. J. E. M., van Leeuwaarden, J. S. H., and Mathijsen, B. W. J. (2015). Novel heavy-traffic regimes for large-scale service systems. *SIAM Journal on Applied Mathematics*, 75(2):787–812.

[122] Ji, R. and Lejeune, M. A. (2021). Data-driven optimization of reward-risk ratio measures. *INFORMS Journal on Computing*, 33(3):1120–1137.

[123] Jiang, R., Shen, S., and Zhang, Y. (2017). Integer programming approaches for appointment scheduling with random no-shows and service durations. *Operations Research*, 65(6):1638–1656.

[124] Joiner, M. C. and van der Kogel, A. (2019). *Basic Clinical Radiobiology*. CRC Press, Boca Raton, FL, 5th edition.

[125] Jongbloed, G. and Koole, G. (2001). Managing uncertainty in call centres using poisson mixtures. *Applied Stochastic Models in Business and Industry*, 17(4):307–318.

[126] Kaandorp, G. C. and Koole, G. (2007). Optimal outpatient appointment scheduling. *Health Care Management Science*, 10(3):217–229.

[127] Kaas, R., Goovaerts, M., Dhaene, J., and Denuit, M. (2008). *Modern Actuarial Risk Theory: Using R*. Springer-Verlag, Berlin, 2nd edition.

[128] Karlin, S. and Studden, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics*, volume 15. John Wiley & Sons, New York.

[129] Kellerer, H., Pferschy, U., and Pisinger, D. (2004). *Knapsack Problems*. Springer-Verlag, Berlin.

[130] Kemperman, J. H. B. (1968). The general moment problem, a geometric approach. *The Annals of Mathematical Statistics*, 39(1):93–122.

[131] Kemperman, J. H. B. (1971). Moment problems with convexity conditions I. In *Optimizing Methods in Statistics*, pages 115–178. Elsevier.

[132] Kingman, J. F. C. (1962). Some inequalities for the queue GI/G/1. *Biometrika*, 49(3/4):315–324.

[133] Kingman, J. F. C. (1963). On inequalities of the Tchebychev type. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 59, pages 135–146. Cambridge University Press.

[134] Kleer, P. and van Leeuwaarden, J. S. H. (2022). Optimal stopping theory for a distributionally robust seller. Preprint available at https://arxiv.org/abs/2206.02477.

[135] Kong, Q., Lee, C.-Y., Teo, C.-P., and Zheng, Z. (2013). Scheduling arrivals to a stochastic service delivery system using copositive cones. *Operations Research*, 61(3):711–726.

[136] Kong, Q., Lee, C.-Y., Teo, C.-P., and Zheng, Z. (2016). Appointment sequencing: Why the smallest-variance-first rule may not be optimal. *European Journal of Operational Research*, 255(3):809–821.

[137] Kong, Q., Li, S., Liu, N., Teo, C.-P., and Yan, Z. (2020). Appointment scheduling under time-dependent patient no-show behavior. *Management Science*, 66(8):3480–3500.

[138] Kuiper, A., Mandjes, M., and de Mast, J. (2017). Optimal stationary appointment schedules. *Operations Research Letters*, 45(6):549–555.

[139] Lasserre, J. B. (2001). Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817.

[140] Lau, H.-S. and Lau, A. H.-L. (1996). The newsstand problem: A capacitated multiple-product single-period inventory problem. *European Journal of Operational Research*, 94(1):29–42.

[141] Lavine, M. (1991). Sensitivity in Bayesian statistics: The prior and the likelihood. *Journal of the American Statistical Association*, 86(414):396–399.

[142] Lee, H. L. and Cohen, M. A. (1983). A note on the convexity of performance measures of M/M/c queueing systems. *Journal of Applied Probability*, 20(4):920–923.

[143] Liu, C. and Hasenbein, J. J. (2019). Naor's model with heterogeneous customers and arrival rate uncertainty. *Operations Research Letters*, 47(6):594–600.

[144] Liu, N. (2016). Optimal choice for appointment scheduling window under patient no-show behavior. *Production and Operations Management*, 25(1):128–142.

[145] Liu, Y., Meskarian, R., and Xu, H. (2017). Distributionally robust reward-risk ratio optimization with moment constraints. *SIAM Journal on Optimization*, 27(2):957–985.

[146] Madansky, A. (1959). Bounds on the expectation of a convex function of a multivariate random variable. *The Annals of Mathematical Statistics*, 30(3):743–746.

[147] Mahes, A., Mandjes, M., Boon, M., and Taylor, P. (2021). Dynamic appointment scheduling. Preprint available at https://arxiv.org/abs/2103.09596.

[148] Mak, H.-Y., Rong, Y., and Zhang, J. (2014). Sequencing appointments for service systems using inventory approximations. *Manufacturing & Service Operations Management*, 16(2):251–262.

[149] Mak, H.-Y., Rong, Y., and Zhang, J. (2015). Appointment scheduling with limited distributional information. *Management Science*, 61(2):316–334.

[150] Mallows, C. (1956). Generalizations of Tchebycheff's inequalities. *Journal of the Royal Statistical Society: Series B (Methodological)*, 18(2):139–168.

[151] Mallows, C. L. and Richter, D. (1969). Inequalities of Chebyshev type involving conditional expectations. *The Annals of Mathematical Statistics*, 40(6):1922–1932.

[152] Mancilla, C. and Storer, R. (2012). A sample average approximation approach to stochastic appointment sequencing and scheduling. *IIE Transactions*, 44(8):655–670.

[153] Markov, A. (1884). On certain applications of algebraic continued fractions. Ph. D. thesis (in Russian), St Petersburg.

[154] Marshall, A. W. and Olkin, I. (1960). Multivariate Chebyshev inequalities. *The Annals of Mathematical Statistics*, pages 1001–1014.

[155] Marshall, A. W., Olkin, I., and Arnold, B. C. (1979). *Inequalities: Theory of Majorization and its Applications*. Springer, New York.

[156] Mercer, A. (1960). A queueing problem in which the arrival times of the customers are scheduled. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 108–113.

[157] Mohajerin Esfahani, P. and Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1-2):115–166.

[158] MOSEK (2019). MOSEK optimization software. Technical report, version 9.1.9. http://docs.mosek.com/9.1/toolbox/index.html.

[159] Nahmias, S. (1982). Perishable inventory theory: A review. *Operations Research*, 30(4):680–708.

[160] Nahmias, S. (2009). *Production and Operations Analysis*. McGraw-hill Education, New York, 6th edition.

[161] Nahmias, S. and Schmidt, C. P. (1984). An efficient heuristic for the multi-item newsboy problem with a single constraint. *Naval Research Logistics Quarterly*, 31(3):463–474.

[162] Naor, P. (1969). The regulation of queue size by levying tolls. *Econometrica: Journal of the Econometric Society*, 37(1):15–24.

[163] Natarajan, K., Sim, M., and Uichanco, J. (2018). Asymmetry and ambiguity in newsvendor models. *Management Science*, 64(7):3146–3167.

[164] Natarajan, K. and Teo, C.-P. (2017). On reduced semidefinite programs for second order moment bounds with applications. *Mathematical Programming*, 161(1):487–518.

[165] Natarajan, K., Teo, C. P., and Zheng, Z. (2011). Mixed 0-1 linear programs under objective uncertainty: A completely positive representation. *Operations Research*, 59(3):713–728.

[166] Natarajan, K. and Zhou, L. (2007). A mean-variance bound for a three-piece linear function. *Probability in the Engineering and Informational Sciences*, 21(4):611–621.

[167] Nesterov, Y. et al. (1997). Structure of non-negative polynomials and optimization problems. Technical report, Université Catholique de Louvain, Center for Operations Research.

[168] Nguyen, V. A., Zhang, F., Blanchet, J., Delage, E., and Ye, Y. (2020). Distributionally robust local non-parametric conditional estimation. *Advances in Neural Information Processing Systems*, 33:15232–15242.

[169] Nguyen, V. A., Zhang, F., Blanchet, J., Delage, E., and Ye, Y. (2021). Robustifying conditional portfolio decisions via optimal transport. Preprint available at https://arxiv.org/abs/2103.16451.

[170] Owhadi, H., Scovel, C., Sullivan, T. J., McKerns, M., and Ortiz, M. (2013). Optimal uncertainty quantification. *SIAM Review*, 55(2):271–345.

[171] Padmanabhan, D., Natarajan, K., and Murthy, K. (2021). Exploiting partial correlations in distributionally robust optimization. *Mathematical Programming*, 186:209–255.

[172] Parrilo, P. A. (2000). Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. Ph.D. thesis, California Institute of Technology.

[173] Parrilo, P. A. (2003). Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming*, 96:293–320.

[174] Perakis, G. and Roels, G. (2008). Regret in the newsvendor model with partial information. *Operations Research*, 56(1):188–203.

[175] Perkó, Z., Bortfeld, T., Hong, T., Wolfgang, J., and Unkelbach, J. (2018). Derivation of mean dose tolerances for new fractionation schemes and treatment modalities. *Physics in Medicine & Biology*, 63(3):035038.

[176] Pólik, I. and Terlaky, T. (2007). A survey of the S-lemma. *SIAM Review*, 49(3):371–418.

[177] Popescu, I. (2005). A semidefinite programming approach to optimal-moment bounds for convex classes of distributions. *Mathematics of Operations Research*, 30(3):632–657.

[178] Popescu, I. (2007). Robust mean-covariance solutions for stochastic optimization. *Operations Research*, 55(1):98–112.

[179] Postek, K., Ben-Tal, A., Den Hertog, D., and Melenberg, B. (2018). Robust optimization with ambiguous stochastic constraints under mean and dispersion information. *Operations Research*, 66(3):814–833.

[180] Qi, J. (2016). Mitigating delays and unfairness in appointment systems. *Management Science*, 63(2):566–583.

[181] Rahimian, H. and Mehrotra, S. (2019). Distributionally robust optimization: A review. Preprint available at https://arxiv.org/abs/1908.05659.

[182] Randhawa, R. S. (2016). Optimality gap of asymptotically derived prescriptions in queueing systems. *Queueing Systems*, 83(1):131–155.

[183] Robinson, L. W. and Chen, R. R. (2003). Scheduling doctors' appointments: Optimal and empirically-based heuristic policies. *IIE Transactions*, 35(3):295–307.

[184] Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42.

[185] Rockafellar, R. T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471.

[186] Rogosinski, W. W. (1958). Moments of non-negative mass. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 245(1240):1–27.

[187] Rolski, T. (1972). Some inequalities for GI/M/n queues. *Applicationes Mathematicae*, 1(13):42–47.

[188] Roos, E., Brekelmans, R., van Eekelen, W. J. E. C., den Hertog, D., and van Leeuwaarden, J. S. H. (2022). Tight tail probability bounds for distribution-free decision making. *European Journal of Operational Research*, 299(3):931–944.

[189] Ross, S. M. (1996). *Stochastic Processes*. John Wiley & Sons, New York, 2nd edition.

[190] Rujeerapaiboon, N., Kuhn, D., and Wiesemann, W. (2018). Chebyshev inequalities for products of random variables. *Mathematics of Operations Research*, 43(3):887–918.

[191] Scarf, H. E. (1958). A min-max solution of an inventory problem. In Arrow, K. J., Karlin, S., and Scarf, H. E., editors, *Studies in the Mathematical Theory of Inventory and Production*, pages 201–209. Stanford University Press, Stanford, CA.

[192] Shaked, M. and Shanthikumar, J. G. (1988). Stochastic convexity and its applications. *Advances in Applied Probability*, 20(2):427–446.

[193] Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. Springer, New York.

[194] Shanthikumar, J. G. and Wu, C. (1991). On the starshapeness of G/G/c queueing systems. *Advances in Applied Probability*, 23(2):431–435.

[195] Shanthikumar, J. G. and Yao, D. D. (1988). Strong stochastic convexity and its applications in parametric optimization of queueing systems. In *Proceedings of the 27th IEEE Conference on Decision and Control*, pages 657–662. IEEE.

[196] Shapiro, A. (2001). On duality theory of conic linear problems. In Goberna, M. A. and Lopez, M. A., editors, *Semi-Infinite Programming: Recent Advances*, pages 135–165. Kluwer Academic Publishers, Dordrecht, the Netherlands.

[197] Shapiro, A. (2006). Worst-case distribution analysis of stochastic programs. *Mathematical Programming*, 107(1):91–96.

[198] Shapiro, A., Dentcheva, D., and Ruszczyński, A. (2009). *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia, 3th edition.

[199] Shapiro, A. and Kleywegt, A. (2002). Minimax analysis of stochastic problems. *Optimization Methods and Software*, 17(3):523–542.

[200] Silver, E. A., Pyke, D. F., and Peterson, R. (1998). *Inventory Management and Production Planning and Scheduling*. John Wiley & Sons, New York, 3th edition.

[201] Smith, J. E. (1995). Generalized Chebychev inequalities: Theory and applications in decision analysis. *Operations Research*, 43(5):807–825.

[202] Spitzer, F. (1956). A combinatorial lemma and its application to probability theory. *Transactions of the American Mathematical Society*, 82(2):323–339.

[203] Srivastava, P. R., Wang, Y., Hanasusanto, G. A., and Ho, C. P. (2021). On data-driven prescriptive analytics with side information: A regularized Nadaraya-Watson approach. Preprint available at https://arxiv.org/abs/2110.04855.

[204] Steckley, S. G., Henderson, S. G., and Mehrotra, V. (2009). Forecast errors in service systems. *Probability in the Engineering and Informational Sciences*, 23(2):305–332.

[205] Stieltjes, T.-J. (1894). Recherches sur les fractions continues. *Annales de la Faculté des Sciences de Toulouse: Mathématiques*, 8(4):1–122.

[206] Stieltjes, T.-J. (1895). Recherches sur les fractions continues. *Annales de la Faculté des Sciences de Toulouse: Mathématiques*, 9(1):5–47.

[207] Ten Eikelder, S. C. M., Ajdari, A., Bortfeld, T., and den Hertog, D. (2022). Adjustable robust treatment-length optimization in radiation therapy. *Optimization and Engineering*, 23(4):1949–1986.

[208] Vairaktarakis, G. L. (2000). Robust multi-item newsboy models with a budget constraint. *International Journal of Production Economics*, 66(3):213–226.

[209] van Eekelen, W. J. E. C. (2023). A generalized moment approach to sharp bounds for conditional expectations.

[210] van Eekelen, W. J. E. C., den Hertog, D., and van Leeuwaarden, J. S. H. (2022). MAD dispersion measure makes extremal queue analysis simple. *INFORMS Journal on Computing*, 34(3):1681–1692.

[211] van Eekelen, W. J. E. C., den Hertog, D., and van Leeuwaarden, J. S. H. (2023a). Distributionally robust appointment scheduling that can deal with independent service times.

[212] van Eekelen, W. J. E. C., Hanasusanto, G. A., Hasenbein, J. J., and van Leeuwaarden, J. S. H. (2023b). Second-order bounds for the M/M/s queue with random arrival rate.

[213] van Leeuwaarden, J. S. H. and Stegehuis, C. (2021). Robust subgraph counting with distribution-free random graph analysis. *Physical Review E*, 104(4):044313.

[214] van Parys, B. P., Goulart, P. J., and Kuhn, D. (2016). Generalized Gauss inequalities via semidefinite programming. *Mathematical Programming*, 156(1-2):271–302.

[215] Vandenberghe, L., Boyd, S., and Comanor, K. (2007). Generalized Chebyshev bounds via semidefinite programming. *SIAM Review*, 49(1):52–64.

[216] Vasicek, O. A. (1977). An inequality for the variance of waiting time under a general queuing discipline. *Operations Research*, 25(5):879–884.

[217] Wang, P. P. (1993). Static and dynamic scheduling of customer arrivals to a single-server system. *Naval Research Logistics*, 40(3):345–360.

[218] Wang, Y., Prasad, M. N., Hanasusanto, G. A., and Hasenbein, J. J. (2022). Distributionally robust observable strategic queues. Preprint available at https://arxiv.org/abs/2204.03221.

[219] Weber, R. R. (1980). Note—on the marginal benefit of adding servers to G/GI/*m* queues. *Management Science*, 26(9):946–951.

[220] Weber, R. R. (1983). A note on waiting times in single server queues. *Operations Research*, 31(5):950–951.

[221] Weisser, T., Legat, B., Coey, C., Kapelevich, L., and Vielma, J. P. (2019). Polynomial and moment optimization in Julia and JuMP. https://pretalx.com/juliacon2019/talk/QZBKAU/.

[222] Whitt, W. (1984). On approximations for queues, I: Extremal distributions. *AT&T Bell Laboratories Technical Journal*, 63(1):115–138.

[223] Whitt, W. (2006). Staffing a call center with uncertain arrival rate and absenteeism. *Production and Operations Management*, 15(1):88–102.

[224] Wiesemann, W., Kuhn, D., and Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376.

[225] Xie, W. and Ahmed, S. (2018). On deterministic reformulations of distributionally robust joint chance constrained optimization problems. *SIAM Journal on Optimization*, 28(2):1151–1182.

[226] Xin, L. and Goldberg, D. A. (2021). Time (in)consistency of multistage distributionally robust inventory models with moment constraints. *European Journal of Operational Research*, 289(3):1127–1141.

[227] Xu, H., Liu, Y., and Sun, H. (2018). Distributionally robust optimization with matrix moment constraints: Lagrange duality and cutting plane methods. *Mathematical Programming*, 169(2):489–529.

[228] Yanıkoğlu, İ., Gorissen, B. L., and den Hertog, D. (2019). A survey of adjustable robust optimization. *European Journal of Operational Research*, 277(3):799–813.

[229] Yue, J., Chen, B., and Wang, M.-C. (2006). Expected value of distribution information for the newsvendor problem. *Operations Research*, 54(6):1128–1136.

[230] Zacharias, C. and Pinedo, M. (2014). Appointment scheduling with no-shows and overbooking. *Production and Operations Management*, 23(5):788–801.

[231] Zacharias, C. and Yunes, T. (2020). Multimodularity in the stochastic appointment scheduling problem with discrete arrival epochs. *Management Science*, 66(2):744–763.

[232] Žáčková, J. (1966). On minimax solutions of stochastic linear programming problems. *Časopis pro Pěstování Matematiky*, 91(4):423–430.

[233] Zan, J., Hasenbein, J. J., Morton, D. P., and Mehrotra, V. (2018). Staffing call centers under arrival-rate uncertainty with Bayesian updates. *Operations Research Letters*, 46(4):379–384.

[234] Zhou, S., Ding, Y., Huh, W. T., and Wan, G. (2021). Constant job-allowance policies for appointment scheduling: Performance bounds and numerical analysis. *Production and Operations Management*, 30(7):2211–2231.

[235] Zhu, S. and Fukushima, M. (2009). Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, 57(5):1155–1168.

Summary

In this thesis, we discuss distribution-free methods for stochastic models. The traditional approach to these models often assumes full information with respect to the probability distributions of the random variables under consideration. However, in this thesis, a distribution-free perspective is adopted by assuming only partial knowledge of these distributions, often limited to moment information. Distributionally robust analysis seeks to determine the worst-case model performance by optimizing over a set of probability distributions that comply with this partial information. This necessitates us to solve semi-infinite optimization problems through the application of duality theory.

The thesis tries to bridge the research on generalized moment problems, distributionally robust stochastic programming and extremal queueing models, three prevalent themes in the applied probability and optimization literature. It uses methods from generalized moment problem literature to derive new distribution-free bounds which can then be used for solving stochastic optimization problems and performing extremal analysis on queueing models. One of the key contributions emphasized in this thesis is the use of semi-infinite linear optimization problems and primal-dual techniques to establish tight (i.e., "best possible") bounds for the distribution-free analysis and optimization of stochastic models, facilitating distribution-free decision making. This work further highlights specific combinations of objective functions and ambiguity sets that yield worst-case probability distributions that are, in some sense, insensitive to the precise stochastic model dynamics under consideration, thereby significantly simplifying distribution-free analysis.

The different chapters detail various applications of the distribution-free bounds. In Chapter 2, we investigate the extremal queue problem, focusing on the worst-case performance of the GI/G/1 queue under mean-dispersion constraints for the interarrival- and service-time distributions. We use the mean absolute deviation from the mean (MAD) as the dispersion measure instead of the more common variance. Our key observation is that the expected waiting time can be expressed as a componentwise convex function of random variables, which allows us to use known tight bounds for the distribution-free analysis of the GI/G/1 queue. This approach leverages the extremal distribution's insensitivity property, providing sharp upper and lower bounds for the moments of the waiting time while incorporating the i.i.d. assumption. In Chapter 3, we examine an M/M/s queue with a random arrival rate characterized by its mean, variance and support. We establish tight bounds for the expected waiting time by determining a worst-case distribution supported on two points. The proofs rely on the convex derivative of the expected

waiting time with respect to the arrival rate. These bounds have applications in rational queueing where individuals decide to join or balk based on expected utility and limited market size knowledge. Chapter 4 introduces new bounds for the tail probability of random variables with known support, mean and mean absolute deviation. These bounds result from solving semi-infinite linear programs using the weak duality framework. We apply these bounds to analyze the newsvendor model, stop-loss reinsurance and a chance-constrained optimization problem in radiotherapy. In Chapter 5, we address conditional expectation bounds based on moment information and observed random events. We reformulate this problem as a semi-infinite linear program, enabling us to derive tight bounds for conditional expectations using the duality theory for generalized moment problems. Chapter 6 focuses on the multi-item newsvendor problem with budget constraint, in which we leverage the mean-MAD bounds from earlier chapters to simplify the optimization problem. After reducing this problem to a stochastic program with a simple structure, it follows that a greedy approach can be used to find the optimal order quantities for the newsvendor. Chapter 7 tackles the appointment scheduling problem, in which distribution-free analysis techniques are applied to minimize costs under worst-case scenarios. We address challenges posed by the independence assumption and explore distribution-free methods similar to those used in Chapter 2 for the GI/G/1 queue. Chapter 8 concludes the thesis by exploring classes of functions and distributional information that induce the insensitivity property and suggesting future research directions for a broader framework of distribution-free analysis for i.i.d. stochastic models.

Samenvatting

In dit proefschrift behandelen we verdelingsvrije methoden voor stochastische modellen. De traditionele benadering voor deze modellen gaat vaak uit van volledige informatie met betrekking tot de kansverdelingen van de betreffende stochasten. In dit proefschrift wordt echter een verdelingsvrije benadering gehanteerd door slechts gedeeltelijke kennis over deze verdelingen te veronderstellen, vaak beperkt tot informatie betreffende de momenten van de toevalsvariabelen. Verdelingsvrije analyse zoekt dan naar de slechtst mogelijke modelprestaties door te optimaliseren over een verzameling van kansverdelingen die voldoen aan deze gedeeltelijke informatie. Dit vereist vervolgens het oplossen van semi-oneindige optimaliseringsproblemen met behulp van dualiteitstheorie.

We streven ernaar in dit proefschrift een verband te leggen tussen het onderzoek naar gegeneraliseerde momentproblemen, kansverdelingsvrije stochastische optimaliseringsproblemen en worst-case wachtrijmodellen, drie veel voorkomende onderzoeksrichtingen in de literatuur over toegepaste kansrekening en optimalisatie. We gebruiken methoden uit de literatuur over gegeneraliseerde momentproblemen om nieuwe verdelingsvrije begrenzingen af te leiden, die vervolgens gebruikt kunnen worden voor het oplossen van stochastische optimaliseringsproblemen en het uitvoeren van worst-case analyse op wachtrijmodellen. Een van de belangrijkste bijdragen die dit proefschrift benadrukt, is het gebruik van semi-oneindige lineaire optimaliseringsproblemen en dualiteitstechnieken om scherpe (oftewel, optimale) grenzen vast te stellen voor de verdelingsvrije analyse en optimalisatie van stochastische modellen. Dit proefschrift belicht verder specifieke combinaties van doelfuncties en gedeeltelijke informatie die worst-case kansverdelingen opleveren welke, in zekere zin, ongevoelig zijn voor de precieze stochastische modeldynamiek van het probleem, wat de analyse aanzienlijk vereenvoudigt.

Dit proefschrift behandelt bovenstaande bijdragen aan de hand van diverse toepassingen, die zijn verdeeld over de verschillende hoofdstukken. In hoofdstuk 2 onderzoeken we het worst-case wachtrijprobleem en richten we ons op de slechtst mogelijke prestaties van de GI/G/1 wachtrij onder gemiddelde- en dispersiebeperkingen voor de tussenaankomst- en servicetijdverdelingen. We gebruiken de gemiddelde absolute afwijking van het gemiddelde als de dispersiemaat in plaats van de, meer gangbare, variantie. Onze belangrijkste observatie is dat de verwachte wachttijd kan worden uitgedrukt als een componentgewijze convexe functie van toevalsvariabelen, wat ons in staat stelt om bekende begrenzingen te gebruiken voor de analyse van de GI/G/1 wachtrij. Deze benadering maakt gebruik van de ongevoeligheidseigenschap van de worst-case kansverdeling, en levert scherpe boven- en ondergrenzen op voor de momenten van de wachttijd, die, zelfs onder de onafhankelijkheidsaanname (welke verdelingsvrije analyse normaal gesproken aanzienlijk lastiger maakt), valide zijn. In hoofdstuk 3 beschouwen we een M/M/s wachtrij met een stochastische aankomstintensiteit, gekenmerkt door zijn gemiddelde, variantie en bereik. We stellen scherpe grenzen vast voor de verwachte wachttijd door een worst-case verdeling vast te stellen die ondersteund wordt op twee punten. De wiskundige bewijzen in dit hoofdstuk maken gebruik van de convexe afgeleide van de verwachte wachttijd als functie van de aankomstintensiteit. Deze grenzen kunnen worden toegepast op de rationele wachtrijtheorie, waarbij klanten beslissen of ze wel of niet in de wachtrij gaan staan gebaseerd op het verwachte nut hiervan en de beperkte kennis over de totale marktomvang. In hoofdstuk 4 introduceren we nieuwe grenzen voor de staartkans van toevalsvariabelen met een bekend bereik, gemiddelde en gemiddelde absolute afwijking. Deze grenzen zijn opnieuw het resultaat van het oplossen van semi-oneindige lineaire optimaliseringsproblemen met behulp van zwakke dualiteit. In hoofdstuk 5 behandelen we begrenzingen voor conditionele verwachtingen die gebaseerd zijn op momentinformatie en waargenomen stochastische gebeurtenissen. We herformuleren dit probleem als een semi-oneindig lineair optimaliseringsprobleem, wat ons in staat stelt om scherpe grenzen af te leiden voor conditionele verwachtingen met behulp van de dualiteitstheorie voor gegeneraliseerde momentproblemen. In hoofdstuk 6 kijken we naar het multi-item newsvendor probleem met een beperkt budget, waarbij we de eerder vervaardigde grenzen benutten om het optimaliseringsprobleem te vereenvoudigen. Door dit probleem te reduceren tot een stochastisch optimaliseringsprobleem met een zeer eenvoudige structuur, kunnen we een efficiënt algoritme formuleren om optimale bestelhoeveelheden te vinden. In hoofdstuk 7 bespreken we het optimaal plannen van afspraken met behulp van verdelingsvrije analysetechnieken. Het doel is om de kosten onder de slechtst mogelijke omstandigheden te minimaliseren. We pakken de uitdagingen aan die voortkomen uit de aanname van onafhankelijkheid met behulp van verdelingsvrije methoden die vergelijkbaar zijn met die uit hoofdstuk 2 voor de GI/G/1 wachtrij. In hoofdstuk 8 sluiten we het proefschrift af door klassen van functies en informatie te verkennen die de ongevoeligheidseigenschap opleveren. Tevens beschrijven we mogelijke toekomstige onderzoeksrichtingen voor de ontwikkeling van een breder theoretisch raamwerk voor de verdelingsvrije analyse van stochastische modellen met onafhankelijke toevalsvariabelen.

CENTER DISSERTATION SERIES

CentER for Economic Research, Tilburg University, the Netherlands

No.	Author	Title	ISBN	Published
672	Joobin Ordoobody	The Interplay of Structural and Individual Characteristics	978 90 5668 674 1	February 2022
673	Lucas Avezum	Essays on Bank Regulation and Supervision	978 90 5668 675 8	March 2022
674	Oliver Wichert	Unit-Root Tests in High-Dimensional Panels	978 90 5668 676 5	April 2022
675	Martijn de Vries	Theoretical Asset Pricing under Behavioral Decision Making	978 90 5668 677 2	June 2022
676	Hanan Ahmed	Extreme Value Statistics using Related Variables	978 90 5668 678 9	June 2022
677	Jan Paulick	Financial Market Information Infrastructures: Essays on Liquidity, Participant Behavior, and Information Extraction	978 90 5668 679 6	June 2022
678	Freek van Gils	Essays on Social Media and Democracy	978 90 5668 680 2	June 2022
679	Suzanne Bies	Examining the Effectiveness of Activation Techniques on Consumer Behavior in Temporary Loyalty Programs	978 90 5668 681 9	July 2022
680	Qinnan Ruan	Management Control Systems and Ethical Decision Making	978 90 5668 682 6	June 2022
681	Lingbo Shen	Essays on Behavioral Finance and Corporate Finance	978 90 5668 683 3	August 2022
682	Joshua Eckblad	Mind the Gales: An Attention-Based View of Startup Investment Arms	978 90 5668 684 0	August 2022
683	Rafael Greminger	Essays on Consumer Search	978 90 5668 685 7	August 2022
684	Suraj Upadhyay	Essay on policies to curb rising healthcare expenditures	978 90 5668 686 4	September 2022

No.	Author	Title	ISBN	Published
685	Bert-Jan Butijn	From Legal Contracts to Smart Contracts and Back Again: An Automated Approach	978 90 5668 687 1	September 2022
686	Sytse Duiverman	Four essays on the quality of auditing: Causes and consequences	978 90 5668 688 8	October 2022
687	Lucas Slot	Asymptotic Analysis of Semidefinite Bounds for Polynomial Optimization and Independent Sets in Geometric Hypergraphs	978 90 5668 689 5	September 2022
688	Daniel Brosch	Symmetry reduction in convex optimization with applications in combinatorics	978 90 5668 690 1	October 2022
689	Emil Uduwalage	Essays on Corporate Governance in Sri Lanka	978 90 5668 691 8	October 2022
690	Mingjia Xie	Essays on Education and Health Economics	978 90 5668 692 5	October 2022
691	Peerawat Samranchit	Competition in Digital Markets	978 90 5668 693 2	October 2022
692	Jop Schouten	Cooperation, allocation and strategy in interactive decision-making	978 90 5668 694 9	December 2022
693	Pepijn Wissing	Spectral Characterizations of Complex Unit Gain Graphs	978 90 5668 695 6	November 2022
694	Joris Berns	CEO attention, emotion, and communication in corporate financial distress	978 90 5668 696 3	November 2022
695	Tom Aben	The (long) road towards smart management and maintenance: Organising the digital transformation of critical infrastructures	978 90 5668 697 0	December 2022
696	Gülbike Mirzaoğlu	Essays in Economics of Crime Prevention and Behavior Under Uncertainty	978 90 5668 698 7	February 2023
697	Suwei An	Essays on incentive contracts, M&As, and firm risk	978 90 5668 699 4	February 2023
698	Jorgo Goossens	Non-standard Preferences in Asset Pricing and Household Finance	978 90 5668 700 7	February 2023

No.	Author	Title	ISBN	Published
699	Santiago Bohorquez Correa	Risk and rewards of residential energy efficiency	978 90 5668 701 4	April 2023
700	Gleb Gertsman	Behavioral Preferences and Beliefs in Asset Pricing	978 90 5668 702 1	May 2023
701	Gabriella Massenz	On the Behavioral Effects of Tax Policy	978 90 5668 703 8	May 2023
702	Yeqiu Zheng	The Effect of Language and Temporal Focus on Cognition, Economic Behaviour, and Well-Being	978 90 5668 704 5	May 2023
703	Michela Bonani	Essays on Innovation, Cooperation, and Competition Under Standardization	978 90 5668 705 2	June 2023
704	Fabien Ize	The Role of Transparency in Fairness and Reciprocity Issues in Manager-Employee Relationships	978 90 5668 706 9	June 2023
705	Kristel de Nobrega	Cyber Defensive Capacity and Capability: A Perspective from the Financial Sector of a Small State	978 90 5668 707 6	July 2023
706	Christian Peters	The Microfoundations of Audit Quality	978 90 5668 708 3	June 2023
707	Felix Kirschner	Conic Optimization with Applications in Finance and Approximation Theory	978 90 5668 709 0	July 2023
708	Zili Su	Essays on Equity Incentive and Share Pledging in China	978 90 5668 710 6	September 2023
709	Rafael Escamilla	Managing the Nanostore Supply Chain: Base-of-the-Pyramid Retail in Emerging Markets	978 90 5668 711 3	September 2023
710	Tomas Jankauskas	Essays in Empirical Finance	978 90 5668 712 0	August 2023
711	Tung Nguyen Huy	Fostering Sustainable Land Management in Sub-Saharan Africa: Evidence from Ghana and Burkina Faso	978 90 5668 713 7	September 2023
712	Daniel Karpati	Essays in Finance & Health	978 90 5668 714 4	September 2023
713	Mylène Struijk	IT Governance in the Digital Era: Insights from Meta-Organizations	978 90 5668 715 1	September 2023

No.	Author	Title	ISBN	Published
714	Albert Rutten	Essays on Work and Retirement	978 90 5668 716 8	November 2023
715	Yan Liu	Essays on Credit Rating Agencies in China	978 90 5668 717 5	October 2023
716	Xiaoyue Zhang	Distortions and industrial upgrading in China	978 90 5668 718 2	September 2023
717	Andries van Beek	Solutions in multi-actor projects with collaboration and strategic incentives	978 90 5668 719 9	October 2023
718	Andries Steenkamp	Polynomial Optimization: Matrix Factorization Ranks, Portfolio Selection, and Queueing Theory	978 90 5668 720 5	October 2023
719	Luis Vargas	Sum-of-Squares Representations for Copositive Matrices and Independent Sets in Graphs	978 90 5668 721 2	November 2023
720	Zhenshu Wu	Essays in Corporate Finance and ESG	978 90 5668 722 9	November 2023
721	Frank de Meijer	Integrality and Cutting Planes in Semidefinite Programming Approaches for Combinatorial Optimization	978 90 5668 723 6	November 2023
722	Chayanin Wipusanawan	Standard-Essential Patents, Innovation, and Competition	978 90 5668 724 3	November 2023
723	Ke Wang	Essays in Corporate Risk	978 90 5668 725 0	November 2023
724	Mustafa Ahci	Essays on corporate disclosures, innovation, and investments	978 90 5668 726 7	November 2023
725	Frederique van Leeuwen	Motifs: From Theory to Practice	978 90 5668 727 4	December 2023
726	Lkhagvaa Erdenesuren	Expectation, Anticipation, and Identification: Essays on Subjective Expectations and Economic Decision-Making	978 90 5668 728 1	December 2023
727	Wouter van Eekelen	Distributionally Robust Views on Queues and Related Stochastic Models	978 90 5668 729 8	December 2023

This dissertation explores distribution-free methods for stochastic models. Traditional approaches operate on the premise of complete knowledge about the probability distributions of the underlying random variables that govern these models. In contrast, this work adopts a distribution-free perspective, assuming only partial knowledge of these distributions, often limited to generalized moment information. Distributionally robust analysis seeks to determine the worst-case model performance. It involves optimization over a set of probability distributions that comply with this partial information, a task tantamount to solving a semi-infinite linear program. To address such an optimization problem, a solution approach based on the concept of weak duality is used. Through the proposed weak-duality argument, distribution-free bounds are derived for a wide range of stochastic models. Further, these bounds are applied to various distributionally robust stochastic programs and used to analyze extremal queueing models—central themes in applied probability and mathematical optimization.

Wouter van Eekelen (Roosendaal en Nispen, the Netherlands, 1995) earned his Bachelor's degree in Industrial Engineering from Eindhoven University of Technology in 2016. He then completed a dual Master's degree in Applied Mathematics and Operations Management from the same university in 2019. In September of that year, he joined the Department of Econometrics and Operations Research at Tilburg University, as a Ph.D. candidate in the field of Stochastic Operations Research. As of September 2023, he has assumed his new role as a postdoctoral principal researcher at the University of Chicago Booth School of Business.

ISBN: 978 90 5668 729 8
DOI: 10.26116/8j9b-c575

TILBURG SCHOOL OF ECONOMICS
AND MANAGEMENT

CentER Dissertation Series