Data-driven optimization and analytics for operations management applications

by

Joline Ann Villaranda Uichanco

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Operations Research

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2013

Author . . . .
Sloan School of Management
August 1, 2013

Certified by . . . .
Georgia Perakis
William F. Pounds Professor of Management
Professor of Operations Research and Operations Management
Thesis Supervisor

Certified by . . . .
Retsef Levi
J. Spencer Standish (1945) Professor of Management
Associate Professor of Operations Managementat
Thesis Supervisor

Accepted by . . . .
Dimitris Bertsimas
Boeing Professor of Operations Research
Co-director, Operations Research Center

Data-driven optimization and analytics for operations management applications

by

Joline Ann Villaranda Uichanco

Submitted to the Sloan School of Management
on August 1, 2013, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Operations Research

Abstract

In this thesis, we study data-driven decision making in operation management contexts, with a focus on both theoretical and practical aspects.

The first part of the thesis analyzes the well-known newsvendor model but under the assumption that, even though demand is stochastic, its probability distribution is not part of the input. Instead, the only information available is a set of independent samples drawn from the demand distribution. We analyze the well-known sample average approximation (SAA) approach, and obtain new tight analytical bounds on the accuracy of the SAA solution. Unlike previous work, these bounds match the empirical performance of SAA observed in extensive computational experiments. Our analysis reveals that a distribution's weighted mean spread (WMS) impacts SAA accuracy. Furthermore, we are able to derive distribution parametric free bound on SAA accuracy for log-concave distributions through an innovative optimization-based analysis which minimizes WMS over the distribution family.

In the second part of the thesis, we use spread information to introduce new families of demand distributions under the minimax regret framework. We propose order policies that require only a distribution's mean and spread information. These policies have several attractive properties. First, they take the form of simple closed-form expressions. Second, we can quantify an upper bound on the resulting regret. Third, under an environment of high profit margins, they are provably near-optimal under mild technical assumptions on the failure rate of the demand distribution. And finally, the information that they require is easy to estimate with data. We show in extensive numerical simulations that when profit margins are high, even if the information in our policy is estimated from (sometimes few) samples, they often manage to capture at least 99% of the optimal expected profit.

The third part of the thesis describes both applied and analytical work in collaboration with a large multi-state gas utility. We address a major operational resource allocation problem in which some of the jobs are scheduled and known in advance, and some are unpredictable and have to be addressed as they appear. We employ a novel decomposition approach that solves the problem in two phases. The first is a job scheduling phase, where regular jobs are scheduled over a time horizon. The second is a crew assignment phase, which assigns jobs to maintenance crews under a stochastic number of future emergencies. We propose heuristics for both phases using linear programming relaxation and list scheduling. Using our models, we develop a decision support tool for the utility which is currently being piloted in one of the company's sites. Based on the utility's data, we project that the tool will result in 55% reduction in overtime hours.

Thesis Supervisor: Georgia Perakis
Title: William F. Pounds Professor of Management

Professor of Operations Research and Operations Management

Thesis Supervisor: Retsef Levi
Title: J. Spencer Standish (1945) Professor of Management
Associate Professor of Operations Managementat

Acknowledgments

There are many people I would like to thank who have helped and supported me in my Ph.D. journey, which culminates with this thesis.

I dedicate this thesis to my mother, Jocelyn, who has sacrificed so much throughout the years and who has raised two children despite personal difficulties. Without her, I would not be where I am. Through her example, she has taught me tenacity, resilience, and strong values. I also thank my brother, Joseph, who I know I can count on in times of trouble.

I send my heartfelt thanks to my two fantastic research advisors, Retsef and Georgia, who I have known since my Masters program at MIT. I could not have had better advisors. My Ph.D. journey has had many emotional highs and lows which I could not have weathered without them. They have given me mentorship, moral support and encouragement. I also thank my Masters advisors, Karthik Natarajan and Melvyn Sim, who still remain my mentors.

I would like to send special thanks to Steve Graves and Melvyn Sim for participating in my thesis committee.

Thanks to employees at National Grid, whose help was valuable in writing my thesis. Thanks to Mallik Angalakudati, Sid Balwani, Jorge Calzada, Bikram Chatterjee, and Nick Raad.

I thank all the friends that I made throughout the past five years. I am grateful for the friendship I have found in Ilke Kalcioglu, who has been my roommate for two years and who remains like a sister to me. I thank Ruben Lobel for also being supportive and being such a good friend. Special thanks also goes to Sun Wei, who is also like a sister and who never fails to give advice when I need it.

I thank my good friends, Chaitanya Bandi, Ozan Candogan, Mihalis Markakis, and Yehua Wei, who were in the academic job search with me. I'm glad we went through this process together. I wish them all the best for the future.

I'm grateful for my friends from the ORC. I'm grateful to Ta Chiraphadhanakul, Cristian Figueroa, and Matthieu Monsch for the times we spent studying for Quals, having dinner at Pad Thai Cafe, and watching movies during our first few years. Thanks to the rest of my classmates, Dimitrios Bisias, Michael Frankovich, Zach Leung, Xin Lu, Anna Teytelman, Eric Zarybnisky, and Yuan Zhong.

Thanks to Jason Acimovic, Kostas Bimpikis, Shi Cong, Doug Fearing, Dan Iancu, Ilan Lobel, Nikos Trichiakis and Gerry Tsoukalas, who graduated before me but who all have still supported me.

I also thank the friends I made at the ORC, Ross Anderson, Fernanda Bravo, Andre Calmon, Maxime Cohen, Adam Elmachtoub, Matthew Fontana, Paul Grigas, Vishal Gupta, Shubham Gupta, Swati Gupta, Kris Johnson, Nathan Kallus, Phil Keller, John Kessler, Jonathan Kluberg, Angie King, Maokai Lin, Will Ma, Allison O'Hair, Eric Robinson, Joel Tay, Leon Valdes, and Nataly Youssef. I'm glad for all the memories we've made throughout the years of INFORMS, ISMP and MSOM trips, of barbecues, parties and concerts, of camping trips, of C-Functions, of Thirsty Ear karaoke nights, and of many others.

Thanks to my friends outside of the ORC, who have shared experiences with me and who have also been good emotional support. Thanks to Burak Alver, Doug Clendenin, Laura Kligler, Roza Mahmoodian, Karen Nadua, Ferron Ocampo, Nestor Precioso, Nikolas Pyrgiotis, Val San Antonio, Yannis Simiakis, and Melis Tanner, and Derek Yao.

Finally, I am grateful for the ORC Co-directors, Dimitris Bertsimas and Patrik Jaillet, who have truly been committed to the welfare of the students and the Center. Also, thanks to the amazing ORC administrative staff, Andrew Carvalho, Paulette Mosley, and Laura Rose.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In managing the operations of a firm, decisions often have to be made in an environment having some underlying uncertainty. Some examples include inventory management, supply chain coordination, revenue management, and workforce management. In this thesis, we discuss decision making in operations management contexts under uncertainty, with a focus on both theoretical and practical aspects.

In the first part of the thesis, we make the assumption that the probability distribution of the underlying uncertainty is not known as part of the input. This is particularly relevant in real-world applications since the decision maker does not have a complete description of the underlying uncertainty.

In Chapter 2, we consider the well-known newsvendor model, however the only information available is a set of independent samples drawn from the demand distribution. We consider the well-known sample average approximation (SAA) approach, but there is a cost associated with the sampling process. Gathering more samples incurs a sampling cost. On the other hand, gathering less samples incurs an inaccuracy cost. The sample size must be carefully chosen to balance the cost tradeoffs involved with sampling. We model inaccuracy cost as the expected penalty, where a fixed penalty is incurred in the event that the relative regret of the SAA quantity exceeds a threshold. We obtain new tight analytical bounds on the probability of this event which match the empirical performance observed in extensive computational experiments. Unlike previous work, this bound reveals the *weighted mean spread* (WMS) as an underlying property of the demand distribution which impacts the accuracy of the SAA procedure. Furthermore, we are able to derive a bound independent of distribution parameters based on an innovative optimization-based analysis which minimizes WMS over a distribution family.

In Chapter 3, we study a minimax regret approach to the newsvendor problem. Using a distribution statistic, called *absolute mean spread* (AMS), we introduce new families of demand distributions under the minimax regret framework. We propose order policies that require only a distribution's mean and information on the AMS. Our policies have several attractive properties. First, they take the form of simple closed-form expressions. Second, we can quantify an upper bound on the resulting regret. Third, under an environment of high profit margins, they are provably near-optimal under mild technical assumptions on the failure rate of the demand distribution. And finally, the information that they require is easy to estimate with data. We show in extensive numerical simulations that when profit margins are high, even if the information in our policy is estimated from (sometimes few) samples, they often manage to capture at least 99% of the optimal expected profit.

In the second part of the thesis, we demonstrate using a real-world example how analytics and optimization is used for decision-making in a gas utility.

In Chapter 4, we describe a project which addresses a major operational resource allocation challenge that is typical to the industry and to other application domains. We study the resource allocation problem in which some of the tasks are scheduled and known in advance, and some are unpredictable and have to be addressed as they appear. The utility company has maintenance crews that perform both standard jobs (each must be done before a specified deadline) as well as repair emergency gas leaks (that occur randomly throughout the day, and could disrupt the schedule and lead to significant overtime). The goal is to perform all the standard jobs by their respective deadlines, to address all emergency jobs in a timely manner, and to minimize maintenance crew overtime.

We employ a novel decomposition approach that solves the problem in two phases. The first is a job scheduling phase, where standard jobs are scheduled over a time horizon. The second is a crew assignment phase, which solves a stochastic mixed integer program to assign jobs to maintenance crews under a stochastic number of future emergencies. For the first phase, we propose a heuristic based on the rounding of a linear programming relaxation formulation and prove an analytical worst-case performance guarantee. For the second phase, we propose an algorithm for assigning crews to replicate the optimal solution structure.

We used our models and heuristics to develop a decision support tool for the utility which is currently being piloted in one of the company's sites. Using the utility's data, we project that the tool will result in 55% reduction in overtime hours.

Chapter 2

The data-driven newsvendor

2.1 Introduction

In the classical newsvendor problem, a retailer plans to sell a product over a single period to meet a stochastic demand with a known distribution (Zipkin, 2000). She needs to commit to a stocking quantity before observing the actual demand at the end of the sales period. The retailer incurs an *underage cost* for each unit of unsatisfied demand, and an *overage cost* for each unsold unit of product at the end of the period. The goal of the retailer is to choose an order quantity that minimizes the expected cost. The basic assumption of the newsvendor model is that the demand distribution is known.

In reality, managers need to make inventory decisions without having complete knowledge of the demand distribution. Often, the only information available comes from a set of demand data. *Nonparametric data-driven heuristics* are a class of heuristics that assume the demand data is a random sample drawn from the unknown demand distribution. Typically, these data-driven heuristics are more accurate as the sample size increases. If there is no cost incurred for gathering samples, then it is better to gather as many demand samples as possible. However, in some realistic settings, *there is a cost associated with the data-collection or sampling process*. Gathering more samples incurs a *sampling cost*. On the other hand, gathering less samples incurs an *inaccuracy cost*. The sample size must be carefully chosen to balance the cost tradeoffs involved with sampling. In this chapter, we analyze the cost tradeoffs under the popular nonparametric data-driven heuristic called *sample average approximation* (SAA) (Homem-De-Mello, 2000; Kleywegt *et al.*, 2001). The SAA heuristic minimizes the cost averaged over the *empirical distribution* induced by the sample, instead of the true expected cost that cannot be evaluated.

In our model, let D be the stochastic single-period demand. Let $C(q)$ denote the expected underage and overage cost of an inventory level q. If the true distribution of D is known, then the optimal newsvendor quantity $q^* = \min_{q \geq 0} C(q)$ is a well-specified quantile of the distribution (sometimes called the *critical quantile*). In our setting, the distribution of D is unknown, but a random sample of size N drawn from the unknown distribution is available. Let $\hat{Q}_N$ denote the SAA solution, which is stochastic since its value depends on the random sample. We assess the accuracy of the SAA solution by comparing its expected cost $C(\hat{Q}_N)$ against the optimal expected cost $C(q^*)$. The difference, $C(\hat{Q}_N) - C(q^*)$, is often called the error (or the *regret*) of ordering the SAA solution. The regret normalized by $C(q^*)$ is referred to as the *relative regret.*

Let $\theta_S(N)$ denote the sampling cost of a sample of size N. For instance, $\theta_S(N)$ may be linear and increasing in the sample size: $\theta_S(N) = \rho N$, where $\rho > 0$ is the cost incurred for each demand data. Let $\theta_I(N)$ denote the inaccuracy cost of the SAA solution. There are multiple ways to define the inaccuracy cost θ_I depending on the specific problem context. Suppose a penalty K is incurred whenever the relative regret of the SAA solution exceeds a specified threshold ϵ. In this chapter, we choose to model the inaccuracy cost of sampling as the *expected penalty incurred by ordering the SAA solution.* That is, $\theta_I(N) \triangleq E\left(K \cdot \mathbb{1}_{\left[C(\hat{Q}_N) > (1+\epsilon)C(q^*)\right]}\right) = K \cdot \Pr\left(C(\hat{Q}_N) > (1+\epsilon)C(q^*)\right)$, where $\mathbb{1}_{[A]}$ is the indicator function of event A; it takes a value of 1 if event A occurs and zero otherwise.

The optimal choice of the sample size N minimizes the total cost $\theta_S(N)+\theta_I(N)$. Solving this requires evaluating θ_I or, equivalently, having a probabilistic understanding of the relative regret of the SAA heuristic, $(C(\hat{Q}_N)-C(q^*))/C(q^*)$, as a function of the sample size. In this work, we provide an entirely new optimization-based analysis that: (i) Obtains tight analytical probabilistic bounds on SAA accuracy that match the empirical performance. As a result they can be used to accurately estimate the value of additional samples; and (ii) Highlights several new important properties of the underlying demand distribution that drive the accuracy of the SAA heuristic.

2.1.1 Contributions and Insights

Our work has multifold contributions and provides several important insights:

Informative probabilistic bound. We derive a new analytical bound on the probability that the SAA solution has at most ϵ relative regret. This bound depends only on the sample size, the threshold ϵ, the underage and overage cost parameters, as well as a newly introduced property of the demand distribution called *weighted mean spread* (WMS). To the best of our knowledge, the WMS is an entirely new concept first introduced in this thesis. The *absolute mean spread* (AMS) at x, $\Delta(x)$, is the difference between the conditional expectation of demand above x and the conditional expectation below the x (Definition 2.4.1). The WMS at x is simply the AMS weighted by the density function value, i.e. $\Delta(x)f(x)$. Our analysis shows that the WMS is the property that drives the accuracy of the SAA method. Specifically, the probability that the SAA solution has a relative regret greater than ϵ decays exponentially with a constant proportional to $\Delta(q^*)f(q^*)$. Thus, the SAA procedure is more likely to be have smaller regret if samples are drawn from a distribution with a large value for $\Delta(q^*)f(q^*)$.

Tight probabilistic bounds. Regression analysis demonstrates that the probabilistic bounds we derive for SAA accuracy based on our analysis are *tight.* Hence, we are able to quantify the accuracy gained from obtaining additional samples. This is especially valuable in settings where data-collection incurs a cost. Thus, we are able to characterize the optimal sample size that balances the sampling cost and inaccuracy cost.

Probabilistic bounds for log-concave distributions. The new notion of WMS is used to develop a general optimization-based methodology to derive tight probabilistic bounds for the accuracy of the SAA method over any nonparametric family of distributions. This is done through a specified optimization problem that minimizes the WMS, $\Delta(q^*)f(q^*)$, over the family of distributions. We are able to solve this problem in closed form for the important family of *log-concave distributions*, providing a *tight* lower bound on $\Delta(q^*)f(q^*)$ of all log-concave distributions. As a consequence, we obtain a uniform probabilistic bound for the accuracy of the SAA solution under any log-concave demand distribution. This bound is independent of distribution-specific parameters, and only depends on the sample size, the regret threshold, and the underage and overage cost parameters. (Note that many of the common distributions assumed in inventory and operations management are log-concave, e.g., normal, uniform, exponential, logistic, chi-square, chi, beta and gamma distributions.) The new bound is significantly tighter than the bound in Levi *et al.* (2007). The methodology we developed could potentially be used to derive probabilistic bounds for SAA accuracy under other distribution families. We believe this is a promising future research direction.

Comparing SAA vs. traditional fitting approaches. Finally, we conduct an extensive computational study comparing the accuracy of the SAA method against the naive (but commonly used in industry) approach that first fits the samples to a specified distribution and then solves the newsvendor problem with respect to that distribution. The comparison is made based on the average relative regret each method incurs. To implement the fitting approach, we used the distribution-fitting software EasyFit to find the distribution that best describes the samples from its database of more than 50 distributions. We investigate the effect of the sample size and the effect of sampling from nonstandard demand distributions on the magnitude of the relative regret. In most cases, even when the critical quantile is high, the errors of the SAA method are on par or dominate those of the distribution fitting approach. Moreover, when the samples are drawn from a nonstandard distribution (e.g., mixed normals), the distribution fitting method results in huge errors compared to the SAA method.

2.1.2 Literature Review

There exists a large body of literature on models and heuristics for inventory problems that can be applied when limited demand information is known. One may use either a *parametric* approach or a *nonparametric* approach. A parametric approach assumes that the true distribution belongs to a parametric family of distributions, but the specific values of the parameters are unknown. In contrast, a nonparametric approach requires no assumptions regarding the parametric form of the demand distribution. The following are some examples of parametric approaches. Scarf (1959) proposed a Bayesian procedure that updates the belief regarding the uncertainty of the parameter based on observations that are collected over time. More recently, Liyanage & Shanthikumar (2005) introduced *operational statistics* which, unlike the Bayesian approach, does not assume any prior knowledge on the parameter values. Instead it performs optimization and estimation simultaneously. In another recent work, Akcay *et al.* (2009) propose fitting the samples to a distribution in the Johnson Translation System, which is a parametric family that includes many common distributions. Besides SAA, the following are other examples of nonparametric approaches proposed in previous work. Concave adaptive value estimation (CAVE) (Godfrey & Powell, 2001) successively approximates the objective cost function with a sequence of piecewise linear functions. The bootstrap method (Bookbinder & Lordahl, 1989) estimates the critical quantile of the demand distribution. The infinitesimal perturbation approach (IPA) is a sampling-based stochastic gradient estimation technique that has been used to solve stochastic supply chain models (Glasserman & Ho, 1991). Huh & Rusmevichientong (2009) develop an online algorithm for the newsvendor problem with censored demand data (i.e., data is on sales instead of demand) based on stochastic gradient descent. Another nonparametric method for censored demand is proposed by Huh *et al.* (2008), based on the well-known Kaplan-Meier estimator. Robust optimization addresses distribution uncertainty by providing solutions that are robust against different distribution scenarios. It does this by allowing the distribution to belong to a specified family of distributions. Then one can use a *max-min* approach, attempting to maximize the worst-case expected profit over the set of allowed distributions. Scarf (1958) and Gallego & Moon (1993) derived the max-min order policy for the newsvendor model with respect to a family of distributions with the same mean and variance. Another robust approach attempts to minimize the worst-case "regret" (or hindsight cost of suboptimal decision) over the distribution family. Some recent works using a minimax regret criterion include Ball & Queyranne (2009); Eren & Maglaras (2006); Perakis & Roels (2008); Levi *et al.* (2013b).

In general, the sample average approximation (SAA) method is used to solve two types of stochastic optimization problems. The first type of problems are those that are computationally difficult even though the underlying distribution is known (e.g. two-stage discrete problems where the expectation is difficult to evaluate due to complicated utility functions and multivariate continuous distributions). In this case, sampling is used to approximate the complicated (but known) objective function. The resulting sample approximation leads to a deterministic equivalent problem (e.g. an integer program) that is finite, though possibly with a large dimension due to the number of samples. Some analytical results about probabilistic bounds for SAA accuracy have been derived for two-stage stochastic integer programs (Kleywegt *et al.*, 2001; Swamy & Shmoys, 2005; Shapiro, 2008). It was shown by Kleywegt *et al.* (2001) that the optimal solution of the SAA problem converges to the true optimal value with probability 1. They also derive a probabilistic bound on the SAA accuracy that depends on the variability of the objective function and the size of the feasible region, however they observe it to be too conservative for practical estimates. In this first type of problems, understanding the accuracy of SAA is importance since the sample size directly influences the computational complexity of the problem.

The second type of problems SAA is used to solve are problems whose objective functions are easy to evaluate if the distribution is known (like for the newsvendor problem), however the complication is that the distribution is unknown. Sampling is used to estimate the unknown distribution. The problem we are dealing with in this chapter falls under this second category of problems. If there is an explicit cost for sampling, the accuracy of the SAA solution as a function of the sample size needs to be understood to determine the tradeoff between sampling cost and inaccuracy cost. The accuracy of the SAA solution for the newsvendor problem is analyzed by Levi *et al.* (2007) who derive a probabilistic bound on its relative regret. This probabilistic bound is independent of the underlying demand distribution, and only depends on the sample size, the error threshold, and the overage and underage cost parameters. Since it applies to any demand distribution, it is uninformative and highly conservative. It is uninformative since it does not reveal the types of distributions for which the SAA procedure is likely to be accurate. It is conservative because, as we demonstrate in computational experiments later in this chapter, the probabilistic bound in Levi *et al.* (2007) does not match the empirical accuracy of the SAA seen for many common distributions. Since it is not tight, this probabilistic bound is of limited value in a setting where data-collection incurs a cost. We show later in Section 2.6 that the analysis of Levi *et al.* (2007) greatly underestimates the benefit of gathering additional samples. Similar to the probabilistic bounds derived for the first type of problems using SAA, our probabilistic bounds are distribution-specific. However, unlike those bounds, our bounds are tight. Since the demand distribution is unknown, we use an optimization-based framework to derive probabilistic bounds for SAA accuracy that do not depend on any distribution-parameters.

Other works have analyzed the regret of a nonparametric data-driven heuristic under *censored demand data*. When data is censored, the choice of the inventory level affects the demand data for the next period. Thus, the problem of choosing inventory levels is an online convex optimization problem, because the objective function is not known, but an iterative selection of a feasible solution yields some pertinent information. Since the choice of the current period's solution influences the next period's information, the choice of the sample size is not the critical factor; rather, the critical issue is the design of an online policy to ensure that the regret diminishes over time. Recent results in online convex optimization propose algorithms with convergence of $O(1/\sqrt{N})$ for the expected regret averaged over N periods for perishable inventory (Flaxman *et al.*, 2005). Huh & Rusmevichientong (2009) propose an online algorithm for nonperishable inventory which achieves the same convergence rate. The convergence rate can be improved to $O(\log(N+1)/N)$ when the probability density function f has a nonzero lower bound on an interval containing q^* (Huh & Rusmevichientong, 2009). In contrast with these works, the focus of our chapter is on the case of *uncensored* demand. Therefore, since the choice of the inventory level does not affect the next period's data, the accuracy of the heuristic only depends on the regret of the *current* period. Moreover, the focus of our chapter is bounding the *probability that the relative regret exceeds a small threshold*, rather than the more conservative expected relative regret criterion. Finally, we observe if we were concerned with the regret (rather than the *relative* regret) of the SAA solution, then the probability bound on the SAA solution having regret exceeding ϵ decays exponentially with a constant proportional to $f(q^*)$. Therefore, as with Huh & Rusmevichientong (2009), if a nonzero lower bound on $f(q^*)$ exits, then a probability bound independent of the distribution can be derived. However, since our chapter is concerned with *relative regret*, the probability bound decays with a constant of $\Delta(q^*)f(q^*)$. As we show in the chapter, the task of finding a uniform lower bound for $\Delta(q^*)f(q^*)$ is achievable for the class of log-concave distributions.

Finally, our results are also related to quantile estimation literature. This is because the SAA solution for the newsvendor problem is a particular sample quantile of the empirical distribution formed by the demand samples. The confidence interval for the quantile estimator are well-known (Asmussen & Glynn, 2007). However, unlike in quantile estimation, the accuracy of the SAA solution does not depend on the absolute difference between the true quantile and the quantile estimator. Rather it depends on the *cost* difference between the true quantile and the estimator Thus, in our work, we find a relationship between the two types of accuracies.

2.1.3 Outline

This chapter is structured as follows. In §2.2, we describe the data-driven single-period newsvendor problem. We also discuss a general setting where data-collection incurs a cost linear in the sample size. §2.3 briefly discusses the analysis of Levi *et al.* (2007). §2.4 contains the main theoretical contributions of this work. In §2.6, we revisit the problem of choosing a sample size to balance the marginal cost and benefit of sampling. Finally, in §2.7, we perform computational experiments that compare the performance of the SAA approach to other heuristic methods. Unless given, the proofs are provided in Appendix B.

2.2 The Data-driven Newsvendor Problem

In the newsvendor model, a retailer has to satisfy a stochastic demand D for a single product over a single sales period. Prior to observing the demand, the retailer needs to decide how many units q of the product to stock. Only then is demand realized and fulfilled to the maximum extent possible from the inventory on hand. At the end of the period, cost is incurred; specifically, a per-unit *underage cost* $b > 0$ for each unit of unmet demand, and a per-unit *overage cost* $h > 0$ for each unsold product unit. The goal of the newsvendor is to minimize the total expected cost. That is,

$$\min_{q \geq 0} C(q) \triangleq E\left[b(D-q)^+ + h(q-D)^+\right],$$

where $x^+ \triangleq \max(0, x)$. The expectation is taken with respect to the stochastic demand D, which has a cumulative distribution function (cdf) F.

Much is known about the newsvendor objective function and its optimal solution (see Zipkin 2000). In particular, C is convex in q with a right-sided derivative $\partial_+ C(q) = -b + (b+h)F(q)$ and a left-sided derivative $\partial_- C(q) = -b + (b+h)\Pr(D < q)$. The optimal solution can be characterized through first-order conditions. In particular, if $\partial_- C(q) \leq 0$ and $\partial_+ C(q) \geq 0$, then zero is a subgradient, implying that q is optimal (Rockafellar, 1972). These conditions are met by

$$q^* \triangleq \inf\left\{q : F(q) \geq \frac{b}{b+h}\right\},$$

which is the $\frac{b}{b+h}$ quantile of D, also called the *critical quantile* or the *newsvendor quantile*.

The basic assumption of the newsvendor problem is that there is access to complete knowledge of F. If the cdf F of the demand is unknown, then the optimal ordering quantity q^* cannot be evaluated. Let $\{D^1, D^2, \ldots, D^N\}$ be a random sample of size N drawn from the true demand distribution, and let $\{d^1, d^2, \ldots, d^N\}$ be a particular realization. Instead of optimizing the unknown expected cost, the SAA method optimizes the cost averaged over the drawn sample:

$$\min_{q \geq 0} \hat{C}_N(q) \triangleq \frac{1}{N} \sum_{k=1}^{N} \left(b(d^k - q)^+ + h(q - d^k)^+\right). \tag{2.1}$$

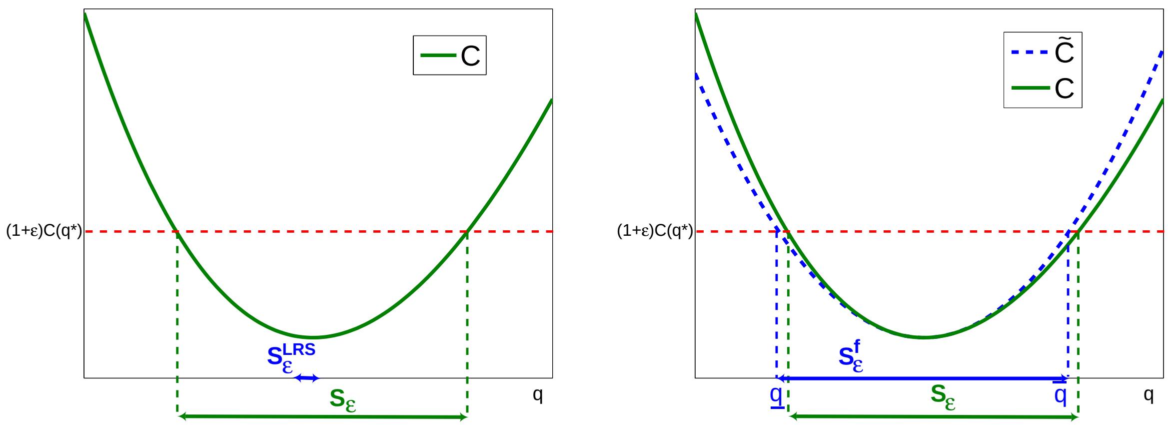
Based on the particular sample, the empirical distribution is formed by putting a weight of $\frac{1}{N}$ on each of the demand values. Note the function $\hat{C}_N$ is the expected cost with respect to the empirical distribution. Hence, the optimal solution to (2.1) is the $\frac{b}{b+h}$ sample quantile. Formally, we denote the empirical cdf as $\hat{F}_N(q) \triangleq \frac{1}{N}\sum_{k=1}^{N} \mathbb{1}_{[D^k \leq q]}$. Let $\hat{Q}_N$ denote the optimal solution to the SAA counterpart with a sample of size N. Thus, $\hat{Q}_N$ is the $\frac{b}{b+h}$ quantile of the random sample:

$$\hat{Q}_N \triangleq \inf\left\{q : \hat{F}_N(q) \geq \frac{b}{b+h}\right\}. \tag{2.2}$$

Note that $\hat{Q}_N$ is a random variable since its value depends on the particular realization of the random sample.

The SAA procedure is more accurate if the sample size N is large. With a larger sample size, the $\frac{b}{b+h}$ sample quantile, $\hat{Q}_N$, is a closer approximation to the true $\frac{b}{b+h}$ quantile, q^*. However, we consider in this work the setting where there is a tradeoff between having too few samples and too many samples. Thus, the choice of the sample size needs to be made carefully. In particular, for some constant $\rho > 0$, we denote by $\theta_S(N) = \rho N$ the *sampling cost* incurred from choosing a sample size N. We denote by $\theta_I(N)$ the *inaccuracy cost* incurred from a sample size N, where $\theta_I(N) = K \cdot \Pr\left(C(\hat{Q}_N) > (1+\epsilon)C(q^*)\right)$ for some $\epsilon > 0$. Note that θ_I represents the expected penalty from ordering the SAA solution

Figure 2-1: The intervals S_ϵ^{LRS} and S_ϵ^f of a newsvendor cost function C.



Note: S_ϵ^{LRS} is defined in (2.3); S_ϵ^f will be defined in Section 4 in (2.6)

if a penalty of K is incurred whenever the relative regret of the SAA solution exceeds a threshold ϵ. We are interested in finding the sample size N for which $\theta_S(N) + \theta_I(N)$ is minimized. This can only be accomplished by developing a probabilistic understanding of how the relative regret of the SAA solution, $(C(\hat{Q}_N) - C(q^*))/C(q^*)$, is influenced by the sample size. In the next section, we discuss a probabilistic bound due to Levi *et al.* (2007). Sections 2.4 and 2.5, we introduce a novel asymptotic analysis of the SAA procedure which results in a tighter probabilistic bound. Armed with this analysis, we will again revisit the problem of sample size selection in Section 2.6.

2.3 Distribution-Free Uniform Probability Bounds

As a background for the new analysis discussed in Section 2.4, the bound of Levi *et al.* (2007) (referred to as the LRS bound) will be discussed first. As part of the discussion, it will be shown that in fact their bound can be improved.

The SAA solution is called ϵ-optimal if its relative regret is no more than ϵ. Let us denote the interval consisting of all ϵ-optimal quantities as S_ϵ, where $S_\epsilon \triangleq \{q : C(q) \leq (1+\epsilon)C(q^*)\}$ (see Figure 2-1). Levi *et al.* (2007) use the left and right one-sided derivatives of C, denoted by $\partial_- C$ and $\partial_+ C$, to define the following interval:

$$S_\epsilon^{LRS} \triangleq \left\{ q : \partial_- C(q) \leq \frac{\epsilon}{3} \min(b,h) \text{ and } \partial_+ C(q) \geq -\frac{\epsilon}{3} \min(b,h) \right\}, \tag{2.3}$$

and then show that $S_\epsilon^{LRS} \subseteq S_\epsilon$ (see left plot of Figure 2-1). Note that S_ϵ^{LRS} consists of all points for which there exists a subgradient with magnitude bounded by $\frac{\epsilon}{3}\min(b,h)$.

Using large deviations results, specifically the Hoeffding inequality (Hoeffding, 1963), Levi *et al.* (2007) derive a bound on the probability that the SAA solution $\hat{Q}_N$ solved with a sample of size N has the properties that $\partial_- C(\hat{Q}_N) \leq \gamma$ and $\partial_+ C(\hat{Q}_N) \geq -\gamma$. The bound depends on N, γ, and the cost parameters b and h. Note that when $\gamma = \frac{\epsilon}{3}\min(b,h)$, the property is equivalent to $\hat{Q}_N \in S_\epsilon^{LRS}$. Then using the fact that $S_\epsilon^{LRS} \subseteq S_\epsilon$, they obtain the following theorem.

Theorem 2.3.1 (LRS bound (Levi *et al.*, 2007))**.** *Consider the newsvendor problem with underage cost $b > 0$ and overage cost $h > 0$. Let $\hat{Q}_N$ be the SAA solution* (2.2) *with sample size N. For a given $\epsilon > 0$, $\hat{Q}_N \in S_\epsilon^{LRS}$ and $C(\hat{Q}_N) \le (1+\epsilon)C(q^*)$ with probability at least*

$$1 - 2\exp\left(-\frac{2}{9}N\epsilon^2\left(\frac{\min\{b,h\}}{b+h}\right)^2\right). \tag{2.4}$$

By using the Bernstein inequality (Bernstein, 1927), we are able to prove a tighter bound than (2.4). Unlike the Hoeffding inequality, the Bernstein inequality uses the fact that we seek to estimate the specific $\frac{b}{b+h}$ quantile. The proof is provided in Appendix B.

Theorem 2.3.2 (Improved LRS bound)**.** *Consider the newsvendor problem with underage cost $b > 0$ and overage cost $h > 0$. Let $\hat{Q}_N$ be the SAA solution* (2.2) *with sample size N. For a given $\epsilon > 0$, $\hat{Q}_N \in S_\epsilon^{LRS}$ and $C(\hat{Q}_N) \le (1+\epsilon)C(q^*)$ with probability at least*

$$1 - 2\exp\left(-\frac{N\epsilon^2}{18+8\epsilon}\cdot\frac{\min\{b,h\}}{b+h}\right). \tag{2.5}$$

The improved LRS bound (2.5) depends on $\frac{\min(b,h)}{b+h}$ rather than on $\left(\frac{\min(b,h)}{b+h}\right)^2$. This is significant because in many important inventory systems, the newsvendor quantile $\frac{b}{b+h}$ is typically close to 1, reflecting high service level requirements. Thus, $\frac{h}{b+h}$ is close to zero, resulting in a very small value for $\frac{\min(b,h)}{b+h}$. Hence, the improved LRS bound gives a *significantly tighter* bound on probability of an ϵ-optimal SAA solution. As an illustration, consider the SAA method applied to a newsvendor problem in which the service level increases from 95% to 99%. In order to maintain the likelihood of achieving the same accuracy, the LRS bound suggests that the sample size needs to be increased by 25 times. In contrast, the improved LRS bound suggests that the accuracy is maintained by a sample size that is only five times as large.

The analysis of Levi *et al.* (2007) yields a probability bound that is general, since it applies to any demand distribution. However, it is uninformative in that it does not shed light on relationship between the accuracy of the SAA heuristic and the particular demand distribution. Furthermore, the LRS bound is very conservative. We demonstrate this empirically through the following experiment. Draw 1000 random samples, each with a sample size of $N = 100$, from a particular distribution. The respective SAA solutions $\{\hat{q}_{100}^1, \ldots, \hat{q}_{100}^{1000}\}$ are computed, where $\hat{q}_{100}^i$ is the SAA solution corresponding to random sample i. We note that although the samples are drawn from specific distributions, the SAA solution is computed purely based on the resulting empirical distribution. The respective

Table 2.1: Theoretical bounds and actual empirical performance of SAA.

Distribution		$\epsilon = 0.02$	$\epsilon = 0.04$	$\epsilon = 0.06$	$\epsilon = 0.08$	$\epsilon = 0.10$
Uniform	Emp conf	81.8%	93.7%	96.6%	99.0%	98.9%
$(A = 0, B = 100)$	N_{LRS}	1,088,200	395,900	209,200	154,300	97,800
	N_f	956	692	544	428	416
Normal	Emp conf	75.8%	89.7%	94.7%	97.3%	99.4%
$(\mu = 100, \sigma = 50)$	N_{LRS}	958,830	339,630	186,370	125,390	109,210
	N_f	3,812	2,676	2,184	1,940	2,096
Exponential	Emp conf	69.6%	84.4%	91.5%	94.0%	98.2%
$(\mu = 100)$	N_{LRS}	855,280	292,090	162,120	102,130	88,560
	N_f	1,472	996	824	684	736
Lognormal	Emp conf	75.1%	90.5%	96.5%	98.2%	98.7%
$(\mu = 1, \sigma = 1.805)$	N_{LRS}	945,890	348,880	207,670	137,190	94,680
	N_f	1,272	932	824	720	616
Pareto	Emp conf	79.1%	92.6%	98.0%	98.1%	99.5%
$(x_m = 1, \alpha = 1.5)$	N_{LRS}	1,025,400	377,500	236,400	135,600	112,600
	N_f	1,152	840	780	592	612

[a] "Emp conf" refers to the empirical confidence, or the fraction of random samples where the SAA procedure achieves relative regret less than ϵ.

relative regret of the SAA solutions are $\{\epsilon^1, \ldots, \epsilon^{1000}\}$ where $\epsilon^i \triangleq \frac{C(\hat{q}^i_{100}) - C(q^*)}{C(q^*)}$. We refer to the fraction of sets that achieve a relative regret less than the target ϵ to be the *empirical confidence*. Using Theorem 2.3.1, we can calculate the minimum sample size predicted by the LRS bound to match the empirical confidence level. If this LRS sample size is significantly greater than 100, then this would imply that the LRS bound is very loose (i.e., conservative) since the same accuracy and confidence probability can be achieved with a smaller number of samples.

Table 2.1 summarizes the results of the outlined experiment for a newsvendor critical quantile of 0.9. Rows labeled N_{LRS} correspond to the LRS sample size. One can see from the table that if, for example, the SAA counterpart is solved with samples drawn from a uniform distribution, the relative errors are less than 0.02 in 81.8% of the sets. However, to match this same error and confidence level, the LRS bound requires 1,088,200 samples. This is almost a thousand times as many samples as the 100 used to generate this confidence probability. We can observe that this large mismatch between the empirical SAA accuracy and the theoretical guarantee by the LRS bound prevails throughout various target errors ϵ and distributions.

One way to understand why the LRS bound is conservative is to observe that the interval S_ϵ^{LRS} is typically very small relative to S_ϵ (see the left plot of Figure 2-1). Naturally, the probability of the SAA solution falling within S_ϵ^{LRS} can be significantly smaller than the probability of it falling within the larger interval S_ϵ.

2.4 New Approximation to the S_ϵ Interval

In this section, we shall develop a tighter approximation of S_ϵ. Specifically, we develop an informative probabilistic bound on the relative regret of the SAA solution that identifies the important properties of the underlying demand distribution that determine the procedure's accuracy.

Suppose the demand is a continuous random variable with a probability density function (pdf) f, which we assume to be continuous everywhere. (Later in §5, we will show that this assumption is automatically satisfied under some simple conditions.) Let $\tilde{C}$ be the second-order Taylor series approximation of C at the point $q = q^*$, where q^* is the optimal newsvendor quantile. Note that since a pdf exists, the cost function is twice differentiable. It is straightforward to verify that

$$\tilde{C}(q) \triangleq \frac{bh}{b+h}\Delta(q^*) + \frac{1}{2}(b+h)(q-q^*)^2 f(q^*),$$

where Δ is the *absolute mean spread* operator defined below.

Definition 2.4.1 (Absolute Mean Spread (AMS)). *Let D be a random variable. We define the* absolute mean spread (AMS) *at x as $\Delta(x) \triangleq E(D|D \geq x) - E(D|D \leq x)$.*

Observe that $\Delta(q^*)$ is simply equal to $\frac{b+h}{bh}C(q^*)$. Consider an approximation to S_ϵ using a sublevel set of $\tilde{C}$ (see the right plot of Figure 2-1):

$$S_\epsilon^f \triangleq \left\{ q : \tilde{C}(q) \leq (1+\epsilon)\tilde{C}(q^*) \right\}. \tag{2.6}$$

The superscript f is to emphasize that the interval is defined by the particular distribution f. The two endpoints of S_ϵ^f are:

$$\underline{q} \triangleq q^* - \sqrt{2\epsilon \frac{bh}{(b+h)^2} \frac{\Delta(q^*)}{f(q^*)}}, \qquad \overline{q} \triangleq q^* + \sqrt{2\epsilon \frac{bh}{(b+h)^2} \frac{\Delta(q^*)}{f(q^*)}}. \tag{2.7}$$

It is clear from Figure 2-1 that S_ϵ^f is not necessarily a subset of S_ϵ. However, by imposing a simple assumption on the pdf, we can guarantee that $S_\epsilon^f \cap [q^*, \infty)$ is a subset of S_ϵ.

Assumption 2.4.1. *The cost parameters (b, h) are such that $f(q)$ is decreasing for all $q \geq q^*$.*

Observe that $\tilde{C}$ matches the first two derivatives of C at the point $q = q^*$. Moreover, $C''(q) = (b+h)f(q)$ and $\tilde{C}''(q) = (b+h)f(q^*)$. Thus, Assumption 2.4.1 implies that $\tilde{C}$ increases faster than C over the interval $[q^*, \infty)$. That is, $\tilde{C}(q) \geq C(q)$ for each $q \in [q^*, \infty)$, implying that $S_\epsilon^f \cap [q^*, \infty)$ is a subset of S_ϵ. Hence, under Assumption 2.4.1, if an order quantity q falls within $S_\epsilon^f \cap [q^*, \infty)$, then this implies that $q \in S_\epsilon$ or, equivalently, $C(q) \leq (1+\epsilon)C(q^*)$.

For distributions that are unimodal or with support $\Re^+$, when the newsvendor quantile $\frac{b}{b+h}$ is sufficiently large, then Assumption 2.4.1 is clearly satisfied. Table D.1 in Appendix D summarizes the range of $\frac{b}{b+h}$ values for which Assumption 1 holds under common demand distributions. In many important inventory systems, the newsvendor quantile is typically large, so Assumption 2.4.1 would hold in a broad set of cases.

Recall that Theorems 2.3.1 and 2.3.2 give lower bounds on the probability that the SAA solution (i.e., the $\frac{b}{b+h}$ sample quantile) lies in S_ϵ^{LRS}, implying that it is ϵ-optimal. However, in our new analysis, the SAA solution is ϵ-optimal if it lies in the interval S_ϵ^f *and* if it is at least as large as q^*. Therefore, instead of taking the $\frac{b}{b+h}$ sample quantile, we bias the SAA solution by a small amount. Later in Theorem 2.4.1, we prove a lower bound on the probability that this *biased* SAA solution lies in $S_\epsilon^f \cap [q^*, \infty)$, implying that it is ϵ-optimal.

Recall that $\hat{F}_N$ is the empirical cdf of a random sample of size N drawn from the demand distribution D. For some $\alpha \geq 0$, define

$$\tilde{Q}_N^\alpha \triangleq \inf\left\{q : \hat{F}_N(q) \geq \frac{b}{b+h} + \frac{1}{2}\frac{\alpha}{b+h}\right\}. \tag{2.8}$$

Note that $\tilde{Q}_N^\alpha$ is a random variable since its value depends on the particular realization of the random sample. The following theorem states that for an appropriately chosen bias factor α, the probability that $\tilde{Q}_N^\alpha$ is ϵ-optimal can be bounded. The proof is found in Appendix B.

Theorem 2.4.1 (Distribution-dependent bound)**.** *Consider the newsvendor problem with underage cost $b > 0$ and overage cost $h > 0$. Let $\tilde{Q}_N^\alpha$ be defined as in (2.8), with $\alpha = \sqrt{2\epsilon bh\Delta(q^*)f(q^*)} + O(\epsilon)$. Under Assumption 1, for a given $\epsilon > 0$, $\tilde{Q}_N^\alpha \in S_\epsilon^f \cap [q^*, \infty)$ and $C(\tilde{Q}_N^\alpha) \leq (1+\epsilon)C(q^*)$ with probability at least $1 - 2U(\epsilon)$, where*

$$U(\epsilon) \sim \exp\left(-\frac{1}{4}N\epsilon\Delta(q^*)f(q^*)\right), \text{ as } \epsilon \to 0. \tag{2.9}$$

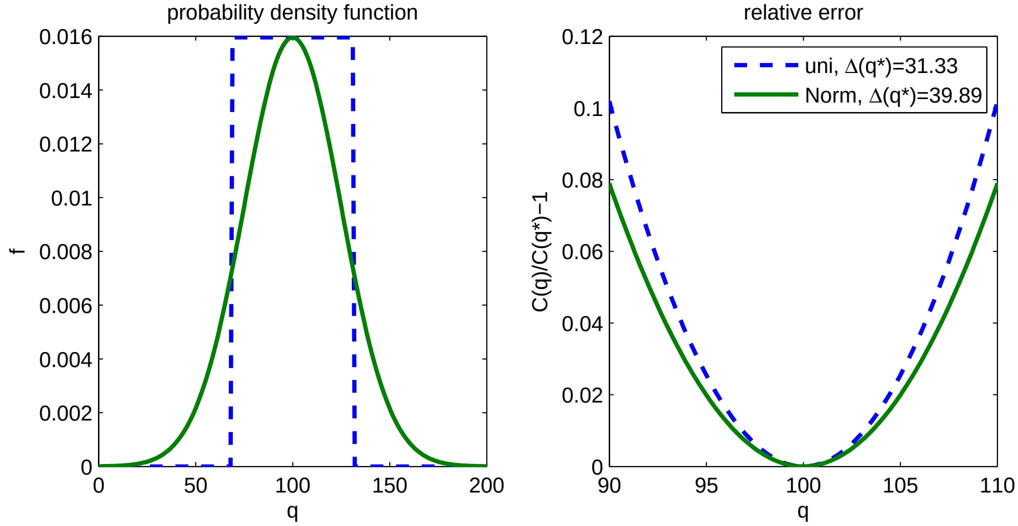
Note that we say that $g_1(x) \sim g_2(x)$ as $x \to 0$ if $\lim_{x\to 0} \frac{g_1(x)}{g_2(x)} = 1$. Thus, in an asymptotic regime as $\epsilon \to 0$, the probability bound in Theorem 2.4.1 only depends on the distribution through the quantity $\Delta(q^*)f(q^*)$. In particular, the data-driven quantity $\tilde{Q}_N^\alpha$ is more likely to be near-optimal when the sample is drawn from a distribution with a high value for $\Delta(q^*)f(q^*)$. Next, we further formalize this insight through the following definition.

Definition 2.4.2 (**Weighted Mean Spread (WMS)**)**.** *Let D be a random variable. We define the* weighted mean spread (WMS) *at x as $\Delta(x)f(x)$.*

From this point onward, any references to the AMS or the WMS are with respect to the $\frac{b}{b+h}$ quantile of the demand distributions, i.e., $\Delta(q^*)$ or $\Delta(q^*)f(q^*)$, respectively.

We briefly discuss the intuition behind the dependence of the bound $\tilde{N}_f$ on the weighted mean spread. The AMS $\Delta(q^*)$ can be thought of as a measure of dispersion around q^*. Note that the slope $C'(q) = -b + (b+h)F(q)$ is zero at q^*. How fast the slope changes depends on how fast the distribution changes around the neighborhood of q^*. In other words, a

Figure 2-2: Probability density function and cost function ($b = 5, h = 5$) for a uniform and normal distribution.



distribution whose mass is concentrated around q^* (i.e., has a small AMS) has a steeper cost function C around q^*. This is illustrated in Figure 2-2. The left plot shows the pdf of a Normal and a uniform distribution. The right plot shows the relative error (as a function of q) when $b = h = 5$ and the optimal order quantity is 100 units for both distributions. The uniform distribution, which has a smaller AMS, has a steeper error function. The decision of ordering 110 units has a larger relative error under the uniform distribution. On the other hand, the size of the confidence interval for the quantile estimator of q^* is inversely proportional to $f(q^*)$ (Asmussen & Glynn, 2007). Thus, if $f(q^*)$ is large, less samples are needed for the quantity $\tilde{Q}_N^\alpha$ to be close (in absolute terms) to q^*. In many distributions the absolute mean spread $\Delta(q^*)$ and the density value $f(q^*)$ exhibit an inverse relationship. Therefore, a large weighted mean spread $\Delta(q^*)f(q^*)$ corresponds to a distribution for which this inverse relationship is balanced.

Recall the empirical experiment in §2.3 that produced Table 2.1. The experiment drew 1000 random samples, each with a sample size of $N = 100$, from a particular distribution. For each of these 1000 sets, the SAA solution was computed and its relative error is noted. (For these experiments, we simply take the $\frac{b}{b+h}$ sample quantile and ignore the bias term. Biasing the SAA solution does not change the insights.) The empirical confidence level is the fraction of the 1000 sets that achieve a relative error less than a target ϵ. Now, we can calculate the sample size predicted by Theorem 2.4.1 to match the empirical confidence level for a target ϵ. The results are reported in Table 2.1 under the rows labeled N_f. Recall that the empirical confidence is generated from an actual sample size $N = 100$. We compare this with N_f and the sample size N_{LRS} predicted by the LRS analysis (Theorem 2.3.1). We find that the sample size predicted by Theorem 2.4.1 using our new analysis are empirically significantly smaller. N_{LRS} typically has an order of magnitude between 100,000 to 1 million

Table 2.2: Regression analysis of $\epsilon = CN^k$

	$\frac{b}{b+h} = 0.8$			$\frac{b}{b+h} = 0.9$			$\frac{b}{b+h} = 0.95$		
	k	C	R^2	k	C	R^2	k	C	R^2
Uniform ($A = 0, B = 100$)	-0.992	2.620	0.994	-1.002	2.807	0.992	-1.051	3.903	0.998
Normal ($\mu = 100, \sigma = 50$)	-1.016	3.155	0.994	-1.026	4.729	0.994	-0.97	5.131	0.994
Exponential ($\mu = 100$)	-0.983	3.066	0.995	-0.979	4.556	0.991	-1.02	9.712	0.998
Lognormal ($\mu = 1, \sigma = 1.805$)	-0.994	1.933	0.994	-1.014	4.384	0.995	-0.948	5.731	0.997
Pareto ($x_m = 1, \alpha = 1.5$)	-1.021	2.977	0.997	-0.999	4.73	0.991	-0.984	7.85	0.992

samples, whereas N_f is typically between 100 to 1,000 samples.

2.4.1 Tightness of distribution-dependent bound

In what follows, we demonstrate through regression analysis that the new probability bound (Theorem 2.4.1) is indeed *tight*; that is, it explains precisely how different factors influence the accuracy of the SAA procedure. First, we verify that it explains how the sample size influences the errors. We estimated empirically the error–sample size relationship by estimating parameters C, k in the equation $\epsilon = CN^k$ through regression. We fix a 90% confidence level and cost parameters $h = 1$ and b. The number of samples N is varied from $\{100, 200, \ldots, 1000\}$. For each N, a total of 1000 independent sets of N independent samples are drawn from a distribution. The SAA solutions $\{\hat{q}_N^1, \ldots, \hat{q}_N^{1000}\}$ are calculated, and the resulting errors are labeled $\{\epsilon_N^1, \ldots, \epsilon_N^{1000}\}$ where $\epsilon_N^k = \frac{C(\hat{q}_N^k) - C(q^*)}{C(q^*)}$. The 90% quantile of the errors is denoted by ϵ_N. We perform the regression using the data $\{(100, \epsilon_{100}), (200, \epsilon_{200}), \ldots, (1000, \epsilon_{1000})\}$. Table 2.2 shows the estimated parameters as well as the R^2 value. The probability bound of Theorem 2.4.1 explains a tight relationship if k is observed to be close to -1. From Table 2.2, all estimates for k are close to -1, and the estimated power function is almost a perfect fit to the data (since R^2 is close to 1).

From Theorem 2.4.1, we can infer that the accuracy of the SAA procedure is only distribution-dependent through the weighted mean spread. We verify this by estimating the relationship $\epsilon = C\{\Delta(q^*)f(q^*)\}^k$ through regression. We fix a 90% confidence level and a sample size N. We consider a pool of ten distributions, i.e., the five distributions in Table 2.2 each under two values of the newsvendor quantile, 0.9 and 0.95. Let ω_i be the weighted mean spread of distribution i. A total of 1000 independent sets of N independent samples are drawn from distribution i. We denote by ϵ_i the 90% quantile of the errors of the 1000 SAA solutions. We perform the regression using the data $\{(\omega_1, \epsilon_1), (\omega_2, \epsilon_2), \ldots, (\omega_{10}, \epsilon_{10})\}$. The results of the regression are reported in Table 2.3 for different values of N. If k is close to -1, then Theorem 2.4.1 precisely explains the relationship between the error and weighted mean spread. From Table 2.3, we observe that all estimates for k are close to -1, and the R^2 value is close to 1 signifying that the estimates are a close fit to the data.

To conclude this section, we note that the probabilistic bound of Theorem 2.4.1 depends on the distribution only through the WMS $\Delta(q^*)f(q^*)$. However, computing it still requires

Table 2.3: Regression analysis of $\epsilon = C\{\Delta(q^*)f(q^*)\}^k$

N	k	C	R^2
100	-0.843	0.0166	0.928
300	-0.939	0.0048	0.990
500	-0.947	0.0029	0.993

knowledge of the specific underlying distribution. In the following section, we shall develop a new optimization framework to get a lower bound on the WMS for a family of distributions. This in turn leads to a uniform probability bound for that family. In particular, we obtain a uniform nonparametric probability bound for all log-concave distributions $\mathbb{L}$.

2.5 Optimization-Driven Bound on WMS

In this section, we assume that the demand distribution f is such that it belongs to $\mathbb{F}$, where $\mathbb{F}$ is a specified family of distributions. Suppose v^* is a lower bound on the weighted mean spread $\Delta(q^*)f(q^*)$ of any distribution in $\mathbb{F}$. Suppose we choose to bias $\tilde{Q}_N^\alpha$, as defined in (2.8), by the factor $\alpha = \sqrt{2\epsilon bhv^*} + O(\epsilon)$. With minor changes to the proof of Theorem 2.4.1, we can show that $C(\tilde{Q}_N^\alpha) \leq (1+\epsilon)C(q^*)$ with probability at least $1 - 2U^*(\epsilon)$, where

$$U^*(\epsilon) \sim \exp\left(-\frac{1}{4}N\epsilon v^*\right), \text{ as } \epsilon \to 0.$$

Note that both the bias factor and this new bound does not depend on specific parameters of the distribution beyond v^*. That is, unlike the bound in Theorem 2.4.1 which depends on the weighted mean spread, *this bound is independent of any distribution parameters.*

Next, we will use an optimization framework to find a lower bound v^* for a family of distributions $\mathbb{F}$. This is accomplished by the following optimization problem:

$$\begin{aligned} \inf_{f,q^*} \quad & \Delta(q^*)f(q^*) \\ \text{s.t.} \quad & f \in \mathbb{F}, \quad \int_{-\infty}^{q^*} f(s)ds = \frac{b}{b+h}. \end{aligned} \tag{2.10}$$

Note that q^* is a decision variable, but because of the second constraint, it is forced to take the value of the $\frac{b}{b+h}$ quantile. Hence, (2.10) finds a distribution in $\mathbb{F}$ with the smallest WMS at the $\frac{b}{b+h}$ quantile. Solving (2.10) or finding a lower bound v^* on its optimal value provides a probability bound for the relative regret of $\tilde{Q}_N^\alpha$ over all demand distributions that belong to $\mathbb{F}$.

In what follows, we will restrict our attention to the family of log-concave distributions $\mathbb{L}$, which includes many of the distributions commonly used in inventory theory (Zipkin, 2000). We shall show that if $\mathbb{F} = \mathbb{L}$, then (2.10) can be solved in closed form. Moreover, the resulting probability bound on the relative regret incurred by the SAA solution is significantly tighter than the LRS bound (2.5).

Definition 2.5.1 (**Log-Concave Distribution**). *A distribution f with support $\mathcal{X}$ is log-concave if* $\log f$ *is concave in* $\mathcal{X}$.

It is known that the *Normal* distribution, the *uniform* distribution, the *logistic* distribution, the *extreme-value* distribution, the *chi-square* distribution, the *chi* distribution, the *exponential* distribution, and the *Laplace* distribution are all log-concave for any respective parameter values. Some families have log-concave density functions for some parameter regimes and not for others. Such families include the *gamma* distribution, the *Weibull* distribution, the *beta* distribution and the *power function* distribution. Note that $\mathbb{L}$ is not characterized through any parameterization (i.e, it does not depend on distributional parameters such as moments that need to be estimated), but rather describes properties satisfied by many common distributions.

Log-concave distributions are necessarily unimodal (Chandra & Roy, 2001). Any distribution in this class must also have monotonic *failure rate* and *reversed hazard rate* (see Definition 2.5.2 below).

Definition 2.5.2 (**Failure Rate and Reversed Hazard Rate**). *The* failure rate *is defined as* $\frac{f}{1-F}$. *The* reversed hazard rate *is defined as* $\frac{f}{F}$.

Log-concave distributions have both an *increasing failure rate* (IFR) and *decreasing reversed hazard rate* (DRHR). Intuitively, this implies that the distribution falls off quickly from its mode.

When we introduced our new analysis in §4, we made the technical assumption that f is continuous everywhere. In fact, if f is log-concave, then it can have at most one jump discontinuity, and the jump can only occur at the left end-point of its support (Sengupta & Nanda, 1997). Therefore, assuming that the demand distribution is log-concave automatically implies that this continuity assumption is also satisfied.

2.5.1 Probability bound for log-concave distributions

To solve (2.10) for log-concave distributions, we first solve a constrained version of (2.10). Specifically, for some $\gamma_0 > 0$ and γ_1, fix the value of q^* and add the constraints $f(q^*) = \gamma_0$ and $\gamma_1 \in \partial \log f(q^*)$, where $\partial \log f(q^*)$ is the set of all subgradients of $\log f$ at q^*. The following optimization problem is obtained:

$$\begin{aligned} \min_f \quad & \frac{b+h}{h}\int_{q^*}^{\infty} sf(s)ds - \frac{b+h}{b}\int_{-\infty}^{q^*} sf(s)(ds) \\ \text{s.t.} \quad & f \in \mathbb{L}, \quad \int_{-\infty}^{q^*} f(s)ds = \frac{b}{b+h}, \\ & f(q^*) = \gamma_0, \quad \gamma_1 \in \partial \log f(q^*). \end{aligned} \tag{2.11}$$

Note that since the density value $f(q^*)$ is fixed, the objective of the constrained problem reduces to minimizing the absolute mean spread $\Delta(q^*)$. The following lemma provides necessary conditions on values of γ_0 and γ_1 for the feasible set of (2.11) to be nonempty. The proof is provided in Appendix B.

Lemma 2.5.1. *Let f be a log-concave pdf. Suppose q^* is the $\frac{b}{b+h}$ quantile with $f(q^*) = \gamma_0$ and $\gamma_1 \in \partial \log f(q^*)$ for some $\gamma_0 > 0$ and γ_1. Then $-\frac{b+h}{h} \leq \frac{\gamma_1}{\gamma_0} \leq \frac{b+h}{b}$.*

Solving the constrained problem (2.11) for log-concave distributions proves to be much simpler than solving (2.10). We shall first show that the optimal value of (2.11) is attained by an exponential-type distribution. As a consequence, we are able to obtain a uniform lower bound for the WMS of *any* log-concave distribution. Particularly, we show that $\Delta(q^*)f(q^*) \geq \frac{\min(b,h)}{b+h}$.

To solve (2.11), we note that for a log-concave distribution f, specifying a value for a subgradient of $\log f$ bounds how fast the pdf f can grow or decay. This is formalized in the next lemma. The proof is provided in Appendix B.

Lemma 2.5.2. *Let f be a log-concave pdf. Suppose that for some t in its support, $f(t) = \gamma_0$ and $\gamma_1 \in \partial \log f(t)$. Then for any x, $f(x) \leq \gamma_0 e^{\gamma_1 (x-t)}$.*

In fact, the upper bound in Lemma 2.5.2 is sufficient to obtain the optimal solution to (2.11). The next lemma characterizes useful conditions that imply the AMS of one random variable D_2 is lower than another random variable D_1. The proof is given in Appendix B.

Lemma 2.5.3 (Domination Lemma)**.** *Let f_1 and f_2 be two pdfs, with respective cdfs F_1 and F_2. Suppose that $f_1(x) \leq f_2(x)$ for all x with $f_2(x) > 0$, and that for some t, $F_1(t) = F_2(t)$. Then $\Delta_1(t) \geq \Delta_2(t)$, where Δ_1 and Δ_2 are the respective AMS of f_1 and f_2.*

Finally, the following proposition constructs the optimal solution to optimization problem (2.11).

Proposition 2.5.4. *Let $\mathbb{L}_{q^*,\gamma_0,\gamma_1}$ be the set of all log-concave distributions with $\frac{b}{b+h}$ quantile q^*, and $f(q^*) = \gamma_0$ and $\gamma_1 \in \partial \log f(q^*)$. The distribution with the smallest AMS $\Delta(q^*)$ in $\mathbb{L}_{q^*,\gamma_0,\gamma_1}$ is:*

$$\tilde{f}(x) = \gamma_0 e^{\gamma_1 (x-q^*)}, \quad \forall x \in [\underline{x}, \overline{x}], \tag{2.12}$$

where $\underline{x} = q^* + \frac{1}{\gamma_1} \log\left(1 - \frac{\gamma_1}{\gamma_0}\frac{b}{b+h}\right)$ and $\overline{x} = q^* + \frac{1}{\gamma_1} \log\left(1 + \frac{\gamma_1}{\gamma_0}\frac{h}{b+h}\right)$.

Proof. Note that $\tilde{f}$ is log-concave and therefore belongs in the set $\mathbb{L}_{q^*,\gamma_0,\gamma_1}$. The range of $\tilde{f}$ is well-defined since by Lemma 2.5.1, we have that $\frac{\gamma_1}{\gamma_0} \in \left[-\frac{b+h}{h}, \frac{b+h}{b}\right]$. From Lemma 2.5.2, we know that $\tilde{f}(x) \geq f(x)$ for all $x \in [\underline{x}, \overline{x}]$ for each $f \in \mathbb{L}_{q^*,\gamma_0,\gamma_1}$. Thus, by the Domination Lemma 2.5.3, $\tilde{f}$ has a smaller AMS than any $f \in \mathbb{L}_{q^*,\gamma_0,\gamma_1}$, and is therefore the optimal solution to problem (2.11). $\square$

A graphical illustration of the distribution $\tilde{f}$ in (2.12) is given in Appendix C (Figure C-1). Finally, using the optimal value for problem (2.11), we are able to derive a uniform lower bound on the WMS of any log-concave distribution. To do this, we will need the following Lemma 2.5.5. The proof is provided in Appendix B.

Lemma 2.5.5. *Let $\beta \in (0,1)$ and $\eta \in \left(-\frac{1}{1-\beta}, \frac{1}{\beta}\right)$. Then, the following relationships are true:*

$$\begin{aligned} \left(\frac{1}{1-\beta} + \eta\right) \log\left(1 + \eta(1-\beta)\right) + \left(\frac{1}{\beta} - \eta\right) \log\left(1 - \eta\beta\right) - \min\{\beta, 1-\beta\}\eta^2 &\geq 0, \\ \frac{\beta}{1-\beta} \log\left(\frac{1}{\beta}\right) - \beta &\geq 0, \\ \frac{1-\beta}{\beta} \log\left(\frac{1}{1-\beta}\right) - (1-\beta) &\geq 0. \end{aligned}$$

Proposition 2.5.6. *Suppose D has a log-concave pdf f, with $\frac{b}{b+h}$ quantile q^* and AMS $\Delta(q^*)$. Then $\Delta(q^*) f(q^*) \geq \frac{\min(b,h)}{b+h}$.*

Proof. Suppose that f belongs to the set $\mathbb{L}_{q^*,\gamma_0,\gamma_1}$, i.e., it has a $\frac{b}{b+h}$ quantile q^* with $f(q^*) = \gamma_0$ and $\gamma_1 \in \partial \log f(q^*)$. Denote the optimal value of problem (2.11) by $z^*_{q^*,\gamma_0,\gamma_1}$. From Proposition 2.5.4, the distribution $\tilde{f}$ defined in (2.12) is the optimal solution of problem (2.11), which achieves the minimum AMS $z^*_{q^*,\gamma_0,\gamma_1}$. We consider three cases. If $\frac{\gamma_1}{\gamma_0} \in \left(-\frac{b+h}{h}, \frac{b+h}{b}\right)$, then

$$z^*_{q^*,\gamma_0,\gamma_1} = \frac{\gamma_0}{\gamma_1^2}\left[\left(\frac{b+h}{h} + \frac{\gamma_1}{\gamma_0}\right) \log\left(1 + \frac{\gamma_1}{\gamma_0}\frac{h}{b+h}\right) + \left(\frac{b+h}{b} - \frac{\gamma_1}{\gamma_0}\right) \log\left(1 - \frac{\gamma_1}{\gamma_0}\frac{b}{b+h}\right)\right].$$

If $\frac{\gamma_1}{\gamma_0} = \frac{b+h}{b}$, then $z^*_{q^*,\gamma_0,\gamma_1} = \frac{1}{\gamma_0}\frac{b}{h}\log\left(\frac{b+h}{b}\right)$. If $\frac{\gamma_1}{\gamma_0} = -\frac{b+h}{h}$, then $z^*_{q^*,\gamma_0,\gamma_1} = \frac{1}{\gamma_0}\frac{h}{b}\log\left(\frac{b+h}{h}\right)$. By applying Lemma 2.5.5 with $\eta = \frac{\gamma_1}{\gamma_0}$ and $\beta = \frac{b}{b+h}$, we find that in all three cases $z^*_{q^*,\gamma_0,\gamma_1} \geq \frac{1}{\gamma_0}\frac{\min(b,h)}{b+h}$. Since f is a feasible solution of problem (2.11), we have that $\Delta(q^*) \geq z^*_{q^*,\gamma_0,\gamma_1} \geq \frac{1}{\gamma_0}\frac{\min(b,h)}{b+h} = \frac{1}{f(q^*)}\frac{\min(b,h)}{b+h}$. ☐

Recall our original objective is to find the smallest WMS among distributions in the set $\mathbb{L}$. To do this we partitioned the set into subsets $\mathbb{L}_{q^*,\gamma_0,\gamma_1}$. The derivation of Proposition 2.5.6 implies that even solving the problem (2.11) restricted to a subset $\mathbb{L}_{q^*,\gamma_0,\gamma_1}$, regardless of the value of γ_1, the minimum absolute mean spread is always bounded below by a term that only depends on b, h and γ_0. Hence, we are able to prove that $\frac{\min(b,h)}{b+h}$ is a uniform lower bound on the weighted mean spread for any log-concave distribution. Finally, this implies a probability bound for log-concave distributions. The proof is in Appendix B.

Theorem 2.5.7 (Log-concave bound)**.** *Consider the newsvendor problem with underage cost $b > 0$ and overage cost $h > 0$. Suppose that D is log-concave and satisfies Assumption 1. Let $\tilde{Q}^\alpha_N$ be defined as in (2.8), with $\alpha = \sqrt{2\epsilon bh\frac{\min\{b,h\}}{b+h}} + O(\epsilon)$. Then for a given $\epsilon > 0$,*

$C(\tilde{Q}_N^\alpha) \leq (1+\epsilon)C(q^*)$ *with probability at least* $1 - 2U^*(\epsilon)$, *where*

$$U^*(\epsilon) \sim \exp\left(-\frac{1}{4}N\epsilon\frac{\min\{b,h\}}{b+h}\right), \text{ as } \epsilon \to 0. \tag{2.13}$$

We point out that there are several nonparametric tests proposed in the literature that check whether a set of samples have been drawn from a log-concave distribution. An (1995) proposes to test two necessary conditions for log-concavity. Another test proposed by Sengupta & Paul (2005) involves finding the Least Concave Majorant (LCM) to the log of empirical probability distribution function. The distribution is log-concave with high probability if the "distance" between the LCM and the log of the distribution function does not exceed a threshold.

Theorems 2.4.1 and 2.5.7 require that $f(q)$ is decreasing for all $q \geq q^*$ (Assumption 2.4.1). We can in fact use the log-concave statistical tests to check if this true. The advantage of the test by Sengupta & Paul (2005) is that it estimates the mode of the (unimodal) log-concave distribution. If the cumulative density at the estimated mode is significantly smaller than $\frac{b}{b+h}$, then Assumption 2.4.1 holds with high probability.

Finally, we would like to stress that the insights from our results are not limited to log-concave distributions. In fact, if we can find a lower bound on the weighted mean spread of *any* distribution family, then we can bound the relative error of the SAA procedure applied to a distribution in that family.

2.6 Balancing the cost and benefit of sampling

Recall that our work is motivated for a setting where ordering decisions are determined completely from data which incur a sampling cost. In this section, we will revisit the problem of choosing the "right" sample size which balances the tradeoff between inaccuracy costs and sampling costs. Suppose there is a cost θ_S incurred from data-collection proportional to the sample size. That is, $\theta_S(N) \triangleq \rho N$ for some $\rho > 0$. Whenever the SAA solution has a relative regret greater than a threshold ϵ, a penalty $K > 0$ is incurred. We let the inaccuracy cost of a sample with size N be the expected penalty of ordering the SAA quantity:

$$\theta_I(N) \triangleq E\left(K \cdot \mathbb{1}_{\left[C(\hat{Q}_N) > (1+\epsilon)C(q^*)\right]}\right) = K \cdot \Pr\left(C(\hat{Q}_N) > (1+\epsilon)C(q^*)\right).$$

We are interested in choosing a sample size N in which the total cost is minimized:

$$\min_{N \geq 0} \theta_S(N) + \theta_I(N). \tag{2.14}$$

Column's labeled 'Exact' in Table 2.4 are sample sizes that equate the marginal cost to the *actual* marginal benefit of an additional sample ρ, with $\rho \in \{0.001, 0.005, 0.01\}$.[1] Note

[1] The actual marginal benefit is estimated from simulations experiments. In the simulation, we draw from the demand distribution 1000 random samples of size N, where $N = 100, 150, 200, \ldots, 1000$. For each sample size N^k, we denote by θ^k the fraction of the 1000 random samples whose SAA solutions have relative

Table 2.4: Sample size equating marginal cost to estimated marginal benefit of additional sample.

	$\rho = 0.001$				$\rho = 0.005$				$\rho = 0.01$			
	Exact	f	$\mathbb{L}$	LRS	Exact	f	$\mathbb{L}$	LRS	Exact	f	$\mathbb{L}$	LRS
Normal	1336	6148	16226	0	922	4213	9466	0	759	3380	6555	0
Exponential	1554	7787	16226	0	1139	5186	9466	0	917	4066	6555	0
Lognormal	1548	7223	16226	0	1086	4857	9466	0	870	3838	6555	0
Pareto	1427	6845	16226	0	1018	4633	9466	0	834	3680	6555	0
Uniform	1038	4417	16226	0	698	3130	9466	0	588	2575	6555	0
Gamma	1531	7379	16226	0	1055	4948	9466	0	885	3902	6555	0
Beta	1392	6690	16226	0	999	4540	9466	0	805	3615	6555	0

that since the true demand distribution is not known, the function θ_I cannot be evaluated. However, from Theorems 2.3.2, 2.4.1 and 2.5.7, we instead have upper bounds on θ_I.

For a fixed threshold ϵ, penalty K, and newsvendor cost parameters b, h, let us define

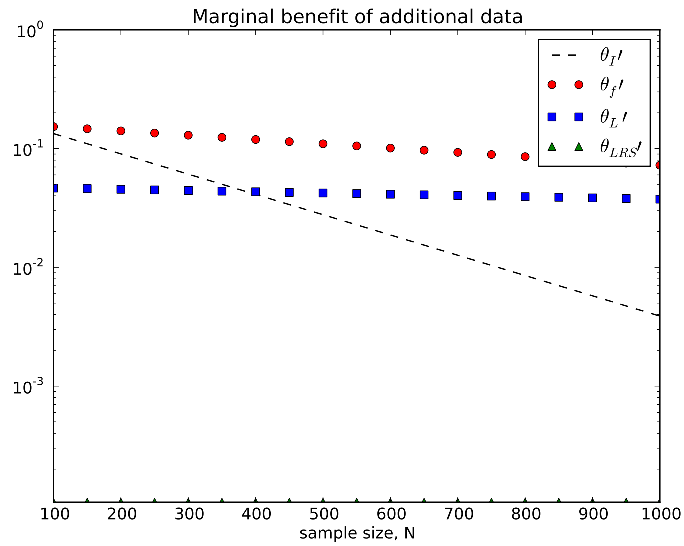
$$\begin{aligned}
\theta_{LRS}(N) &\triangleq 2K \exp\left(-\frac{\epsilon^2}{18+8\epsilon} \cdot \frac{\min\{b,h\}}{b+h} \cdot N\right), \\
\theta_f(N) &\triangleq 2K \exp\left(-\frac{\epsilon}{4}\Delta(q^*)f(q^*)N\right), \\
\theta_{\mathbb{L}}(N) &\triangleq 2K \exp\left(-\frac{\epsilon}{4} \cdot \frac{\min\{b,h\}}{b+h} \cdot N\right).
\end{aligned}$$

Note that each of these functions is an upper bound on the inaccuracy cost, $\theta_I(N)$, of the SAA solution (unbiased and biased). Figure 2-3 plots the derivative of these functions together with the true marginal benefit of additional data, $\theta_I'(N)$, when $b = 95, h = 5, K = 100$. Note that the marginal benefit of an additional data point is decreasing as the sample size increases. However, since the distribution (in this case, a Normal distribution) is unknown, the actual function θ_I' cannot be evaluated. Hence, in practice we cannot determine the exact sample size which equates the marginal benefit θ_I' with the marginal cost ρ.

Suppose we use the functions $\theta_{LRS}, \theta_f, \theta_{\mathbb{L}}$ to replace the unknown θ_I in (2.14). That is, the derivatives of these functions are used to estimate the marginal benefit of additional data (Figure 2-3). In Table 2.4, columns labeled 'LRS', 'f', and '$\mathbb{L}$' are the sample size which equates θ_{LRS}', θ_f', and $\theta_{\mathbb{L}}'$, respectively, to ρ. Note that due to the fact that the LRS probability bound (Theorem 2.3.2) is loose, the derivative θ_{LRS}' greatly underestimates the marginal benefit of additional data. Table 2.4 shows the LRS analysis suggests that a sample size of 0 is optimal, since the benefit gained from each extra sample is negligibly smaller than the cost ρ. Observe that the 'f' sample size is closest to the optimal sample size in magnitude. Moreover, demand distributions that have a larger optimal sample size under 'Exact' also have a larger sample size under 'f'. This demonstrates that the weighted mean spread $\Delta(q^*)f(q^*)$ is the only demand distribution information necessary to infer the

regret larger than ϵ. We use data points $\{(N^k, \theta^k)\}_k$ to estimate the parameters of an exponential function $\theta_I = Ae^{-kN}$.

Figure 2-3: Marginal benefit of additional data as a function of the sample size.



Note: The function θ'_I is computed for a Normal distribution with mean 100 and standard deviation 50. This function is estimated using simulation.

benefit of additional samples. Note that '$\mathbb{L}$' and 'LRS' sample sizes do not depend on the demand distribution, since they only use information of the cost parameters. However, unlike the 'LRS' sample size, the '$\mathbb{L}$' sample size (since it is based on the weighted mean spread) reasonably suggests that additional data decreases the inaccuracy of the data-driven heuristic.

2.7 Computational Experiments

Thus far in this chapter, we have performed an analytical analysis of the accuracy of the SAA procedure. In this section, we conduct computational experiments to analyze its empirical accuracy *vis-à-vis* another widely-used data-driven heuristic. We compare how the accuracy of the two heuristics is affected by the sample size, the critical quantile, and the specific demand distribution.

The SAA heuristic estimates the unknown demand distribution with the *empirical distribution* formed by the samples. Another popular data-driven heuristic is a *distribution-fitting approach*. This heuristic infers the true distribution by fitting the samples to a distribution family the true distribution is assumed to belong. For instance, a common practice is to assume demand is normally distributed, and the sample is used to estimate its parameters. The argument for a distribution-fitting approach is that the tail of the distribution cannot be accurately approximated by the empirical cdf. Hence, when the critical quantile is close to 1, a distribution-fitting approach results in smaller ordering errors by hopefully better approximating the distribution tail. However, the distribution-fitting approach is sensitive to the specification of the distribution family. That is, if the true demand distribution is non-normal, then fitting the sample to a normal distribution might potentially result in a suboptimal order quantity. However, to reduce this risk it is possible to use distribution-fitting softwares, such as EasyFit, which finds the distribution that "best" fits the sample. EasyFit in particular has more than 50 distributions in its database, and chooses the distribution that has the smallest Kolmogorov-Smirnov statistic (based on the largest difference between the fitted distribution and the empirical cdf).

Consider the following experiment. A random sample is drawn from one of the following distributions: (i) an exponential distribution with mean $\mu = 100$, (ii) a Normal distribution with mean $\mu = 100$ and standard deviation $\sigma = 100$, (iii) a Pareto distribution with scale parameter $x_m = 1$ and shape parameter $\alpha = 1.5$, (iv) a scaled beta distribution with range $[0, 50]$ and shape parameters $\alpha = 5, \beta = 1$, and (v) a mixture of three Normal distributions where $\mu = (100, 500, 1000)$, the standard deviation for each is $\sigma = 100$ and the weight vector is $w = (\frac{5}{9}, \frac{1}{3}, \frac{1}{9})$. The drawn sample is used to generate two order heuristics. The first is the SAA order quantity, which is simply the $\frac{b}{b+h}$ quantile of the sample. The second is the Best Fit order quantity, which is the $\frac{b}{b+h}$ quantile of the distribution F chosen by the software EasyFit to be the "best" fit for the sample.

We conduct the experiment for $N \in \{25, 50, 100, 200\}$ and newsvendor quantile $\frac{b}{b+h} \in \{0.1, 0.2, \ldots, 0.9, 0.95, 0.99\}$ (where we fix $h = 1$). For a given $\frac{b}{b+h}$, 100 random samples of size N are drawn from a specific distribution. For the k^{th} sample, the order quantity $\hat{q}^k$ is found through one of the heuristics (SAA or Best Fit). If q^* is the true solution, then the relative error of $\hat{q}^k$ is given by $\epsilon^k = \frac{C(\hat{q}^k) - C(q^*)}{C(q^*)}$. The average relative error of the heuristic is the average of $\{\epsilon^1, \ldots, \epsilon^{100}\}$. Tables D.2, D.3, D.4, D.5 and D.6 in Appendix D present the average errors for different critical quantiles $\frac{b}{b+h}$ and sample sizes N when the samples are drawn from an exponential, a Normal, a Pareto, a beta distribution, and a mixture of Normal distributions, respectively.

We first analyze the effect of the shape of the distribution on the inaccuracy of the SAA method and the Best Fit method. Table D.6 shows the average errors of the methods when the samples are drawn from a nonstandard distribution: a mixture of three Normal distributions. Note that when the sample size is at least 50, the average errors of the SAA heuristic is small (less than 5%, and in most cases less than 1%). The largest average error of the SAA heuristic is about 10%. However, the Best Fit method (which fits the data to standard distributions with at most two modes) show instances when the errors are 25% or even 30%. This suggests that, overall, the SAA method can handle nonstandard distributions better than the Best Fit heuristic, especially if the number of samples available is limited.

We next observe the effect of the sample size on the accuracy of the two heuristics. The cases when the inaccuracy of the solution from the SAA method is most apparent is when the newsvendor quantile is large, e.g. $\frac{b}{b+h} = 0.99$, while the sample size is small. In fact, we observe that when estimating a quantile that corresponds to a rare event (characterized by a small density), the number of samples that need to be taken must be sufficiently large (usually $N = 100$ or more) for the average errors to be small. When $N = 200$, the average errors are uniformly small for all distributions.

2.8 Conclusions

The sample average approximation (SAA) method is a simple and powerful tool for solving inventory problems. It relies only on observing samples of the demand distribution. In many realistic settings, there is a cost incurred in the data-collection process. The sample size must be carefully chosen to minimize the sampling cost and the inaccuracy cost of the SAA procedure. This work derived a bound on the probability that the SAA solution has a relative regret exceeding a specified threshold. This bound reveals the *weighted mean spread* (WMS) as an important property of a distribution which drives the accuracy of the SAA solution to the newsvendor problem. With a fixed sample size, a distribution with a large WMS is more likely to have a small relative regret. The relationship between the error and the sample size and weighted mean spread predicted by our bound is *tight*, as exhibited in our regression analysis. Hence, our work characterizes precisely the additional accuracy gained in collecting additional samples. We introduce an optimization-based framework to derive a uniform lower bound on the WMS of a family of distributions. This results in a probabilistic bound SAA accuracy for that family. We demonstrate this method for log-concave distributions, and derive a probabilistic bound that does not rely on any distribution parameters (such as moments) that need to be estimated. The optimization framework developed seems promising to study and obtain sampling bounds for other interesting classes of distributions.

Chapter 3

Regret optimization with spread information

3.1 Introduction

Inventory decisions often have to be made in an uncertain demand environment. Stochastic inventory models address this demand uncertainty by assuming demand to be stochastic with a known probability distribution (Zipkin, 2000). There are per-unit costs incurred for both understocking and overstocking. The optimal inventory level is highly dependent on the specification of the demand distribution. For example, in the well-known newsvendor problem, the optimal order quantity is a particular quantile of the demand distribution. In reality however, managers have to make inventory decisions using only *limited information* on the demand distribution. This goes against the basic assumption of stochastic inventory models that the complete demand distribution is known. Typically, if a manager only has limited demand information, this often leads him to overstock, resulting in very high inventory costs (Badinelli, 1990).

Often, the only demand information available is *demand data* collected from previous selling periods. An approach widely-used in industry is to chose the demand distribution that best fits the available data. However, this can often result in suboptimal inventory decisions. Levi *et al.* (2013a) demonstrate in simulation experiments that it can be costly to misspecify the demand distribution in the newsvendor problem. They observe that when the demand distribution is chosen by fitting available samples to the best distribution[1], resulting costs can be much greater compared to the optimal cost. With large sample sizes, the cost is between 5% to 10% higher. With small sample sizes, the cost is between 15% to as much as 75% higher.

The focus of this chapter is on *distributionally robust methods* that do not assume one specific demand distribution, but rather assume that the true distribution belongs to a

[1] The best distribution is determined by distribution-fitting software EasyFit which chooses out of a database of more than fifty distributions.

certain *distribution family*. The advantage of such methods is that they mitigate the cost of distribution misspecification. This is because the resulting policies are robust against any distribution that belongs in the family. We focus on one particular such method called the *minimax regret* approach. Regret is defined as the "hindsight cost" of making a suboptimal decision after the parameters of the problem are fully realized. As an example, suppose that the demand is normally distributed, and it is optimal to order 150 units. Due to not having complete information about the demand distribution, a retailer decides to order 300 units. The regret (or hindsight cost) of her decision is the difference between the maximum expected profit (i.e., with 150 units) and the expected profit of ordering 300 units. The minimax regret criterion, first introduced by Savage (1951), attempts to minimize the worst-case regret that a retailer can incur given that she only knows partial information about the distribution. It can be motivated by the following game. Suppose that a retailer needs to commit to an order quantity, but she only knows partial information about the demand distribution. Regardless of which quantity she chooses, nature always tries to hurt her the most by choosing a distribution (out of those that satisfy the known information) which results in the largest regret for her order. Her best strategy is to choose the minimax regret quantity, i.e., the quantity that minimizes the maximum regret over the family of all distributions satisfying the known information. Examples of information that has been used in a minimax regret framework include range, mean, mode, median, variance, skewness or kurtosis (Scarf, 1958; Bertsimas & Popescu, 2004; Yue *et al.*, 2006; Perakis & Roels, 2008).

3.1.1 Contributions

In this chapter, we propose partial-information policies that require only first-order distribution information. In particular, we propose policies that only require information about a distribution's mean and its *absolute mean spread*. Absolute mean spread (AMS) is a first-order measure of a distribution's dispersion around some benchmark value (see Definition 3.2.1). We highlight the advantages of our partial-information policy:

1. *Our policy achieves the minimum worst-case regret over a distribution family.* The policy that we propose can be shown to minimize the maximum regret over any distribution with the same mean and the same absolute mean spread. Therefore, it is robust against a family of distributions.

2. *Our policies require information that is easy to estimate with data.* In reality, the demand information is often inferred from primitive data. Previous work on distributionally robust policies typically ignore the connection of distribution families to data. When the sample size is small, estimating statistics such as skewness or kurtosis can be prone to error. Since the partial information that we require is first-order, we will show (Section 3.5) that they are easy to estimate using data. Therefore, the policies are of practical use when the information has to be estimated from historical demand data.

3. *The minimax regret problem is tractable.* We are able to derive closed-form expressions for the solution of the minimax regret problem using mean and AMS information. In general, the tractability of the minimax regret problem depends on type of demand information. Information such as range, median or mode results in closed-form solutions (Perakis & Roels, 2008). Second-order information such as variance (Yue *et al.*, 2006; Perakis & Roels, 2008) results in a minimax regret problem that can be solved numerically through techniques such as gradient descent. However, with higher order information such as skewness or kurtosis, solving distributionally robust problems is known to be NP-hard (Bertsimas & Popescu, 2004). Since our information is first-order, then the resulting minimax regret problem is tractable and can be solved in closed-form.

4. *Our policy is provably near-optimal when profit margins are high.* Under mild technical conditions on the failure rate of the demand distribution (see Definition 3.3.1), our ordering policies are *near-optimal* in an environment of high profit margins. It can be verified that most common distributions satisfy this technical condition, including those that have an increasing failure rate (IFR). This illustrates the power of our policies in realistic settings, since it has been pointed out by Smith & Agrawal (2000) that the opportunity costs of shortages are quite high in many retail settings. Therefore under scenarios that often lead managers to overstock, our policy can aid managers in making decisions that are almost as good as if there was perfect knowledge of the distribution. We demonstrate through computational experiments in Section 3.6 that this property still holds even when the information has to be estimated from a small number of samples.

3.1.2 Literature Review

Our work belongs under the umbrella of *distribution-free* methods for stochastic inventory models. These are methods that only assume partial information about the distribution. Distribution-free approaches can be either parametric or nonparametric.

Parametric approaches are those in which the distribution is assumed to belong to a parametric family of distributions, but the specific values of the parameters are unknown. Scarf (1959) has analyzed a Bayesian procedure that updates the belief regarding the uncertainty of the parameter based on observations that are collected over time. Liyanage & Shanthikumar (2005) introduced operational statistics which, unlike the Bayesian approach, does not assume any prior knowledge of the parameter values. Instead, it performs optimization and estimation simultaneously. In another recent work, Akcay *et al.* (2009) propose fitting demand samples to a distribution in the Johnson Translation System, which is a parametric family that includes many common distributions.

Nonparametric approaches, on the other hand, require no assumptions regarding the parametric form of the demand distribution. Sample average approximation (SAA) is one such method (Kleywegt *et al.*, 2001; Levi *et al.*, 2007, 2013a) which uses the empirical distribution formed by samples drawn from the true distribution. Concave adaptive value estimation (CAVE) (Godfrey & Powell, 2001) successively approximates the objective cost function with a sequence of piecewise linear functions. The bootstrap method (Bookbinder & Lordahl, 1989) estimates the newsvendor quantile of the demand distribution. The infinitesimal perturbation approach (IPA) is a sampling-based stochastic gradient estimation technique that has been used to solve stochastic supply chain models (Glasserman & Ho, 1991). Huh & Rusmevichientong (2009) develop an adaptive algorithm for inventory planning problems with censored demand data based on stochastic gradient descent. Huh *et al.* (2009) propose an adaptive data-driven policy for censored demand based on the well-known Kaplan-Meier estimator.

Distributionally robust methods address uncertainty about the distribution by providing solutions that are robust against different scenarios. The distribution is allowed to belong to a family of distributions with the same parameters. The minimax regret approach belongs to this category. Another such method is the *max-min* approach (Scarf, 1958; Gallego & Moon, 1993), which attempts to maximize the worst-case expected profit over the set of allowed distributions. Scarf (1958); Gallego & Moon (1993) derived the max-min order quantity for the newsvendor model with respect to a family of distributions with the same mean and variance. However, one major issue with the max-min approach is that it typically leads to policies that are too conservative, whereas the minimax regret approach does not (Perakis & Roels, 2008). Other recent works on robust regret for other revenue management models include Ball & Queyranne (2009); Eren & Maglaras (2006); Perakis & Roels (2008).

To the best of our knowledge, absolute mean spread (AMS) was first introduced in Levi *et al.* (2013a) as a distribution statistic. There, they use it as a tool to analyze the performance of sample average approximation (SAA) method applied to the newsvendor problem. In their analysis, they show that the inaccuracy of the SAA solution is inversely proportional to the AMS value of the demand distribution. Our work, on the other hand, uses AMS in the robust minimax regret framework as a type of information about the demand distribution.

3.1.3 Outline

The chapter is organized as follows. In Section 3.2, we introduce our minimax regret framework using mean and AMS information. Section 3.3 presents optimality results of our policies. Section 3.4 discusses the minimax regret problem under interval AMS information. Section 3.5 introduces a point estimator and a confidence interval estimator for AMS. Finally, in Section 3.6, we conduct computational experiments and compare the empirical performance of our policies to other minimax regret policies.

3.2 Regret optimization under spread information

In this section, we introduce the minimax regret criterion in the context of the single period newsvendor problem. A retailer is selling a product with an uncertain demand D. At the beginning of the sales period, she needs to make a decision on how many units to order, before observing the actual demand. Once the demand occurs, it is satisfied to the maximum extent possible from the units on hand. For simplicity, we assume that the revenue per unit sale is normalized to \$1. Each unit that she purchases costs $\$(1-\beta)$, where $\beta \in (0,1)$ is the per-unit profit margin of the product.

If the demand distribution F is known, the retailer will choose an order quantity y that maximizes her expected profit. The optimal order quantity can be found by solving

$$\max_{y\geq 0} \Pi_F(y) \triangleq E_F\left(\min\{y,D\} - (1-\beta)y\right),$$

where $\Pi_F(y)$ is the expected profit of ordering y units under a distribution F. It is well-known that the optimal order quantity is the β quantile of F, i.e.,

$$F^{-1}(\beta) \triangleq \inf\left\{y \mid F(y) \geq \beta\right\}.$$

Now suppose the true distribution F is unknown, but it is possible to specify a family of distributions $\mathcal{D}$ to which it belongs. The maximum regret (over the distribution family $\mathcal{D}$) of ordering y units is:

$$\rho_{\mathcal{D}}(y) \triangleq \sup_{F\in\mathcal{D}} \left(\max_{z\geq 0} \Pi_F(z) - \Pi_F(y)\right). \tag{3.1}$$

The expression inside the parentheses is the regret (or hindsight cost) of ordering y instead of the optimal quantity after the demand distribution is revealed to be F. The minimax regret criterion chooses the order quantity that minimizes the maximum regret:

$$\rho^*_{\mathcal{D}} \triangleq \min_{y\geq 0} \rho_{\mathcal{D}}(y). \tag{3.2}$$

The family $\mathcal{D}$ is the set of all demand distributions that satisfy the partial information. Information can be a combination of known statistics such as mean, variance, range, mode or median (Yue *et al.*, 2006; Perakis & Roels, 2008). In what follows, we present a novel approach to representing distribution families in a minimax regret framework. Unlike previous distribution families proposed in the literature, ours uses only first-order information, is easy to estimate with data, results in a tractable problem, and is provably near-optimal when profit margins are high.

In Levi *et al.* (2013a), a new distribution statistic was introduced, which they referred to as the *absolute mean spread* or AMS.

Definition 3.2.1 (Absolute mean spread)**.** *For a random variable D with a distribution F, the* absolute mean spread at t *is defined as* $\Delta_F(t) \triangleq E_F(D|D>t) - E_F(D|D\leq t)$.

The absolute mean spread measures the dispersion of the distribution around some point t. It is a first-order measure of spread, in contrast to variance (second-order) or range (zeroth order).

We introduce the family of all distributions with a specified mean and a specified AMS measured at the *newsvendor quantile*:

$$\mathcal{D}_{\mu,\delta} = \left\{ F \mid E_F(\mathbb{1}_{[-\infty,\infty]}) = 1,\ E_F(D) = \mu,\ \Delta_F(F^{-1}(\beta)) = \delta \right\}.$$

We also introduce the family which is a restriction of $\mathcal{D}_{\mu,\delta}$ to distributions with nonnegative support:

$$\mathcal{D}_{\mu,\delta,+} = \left\{ F \mid E_F(\mathbb{1}_{[0,\infty]}) = 1,\ E_F(D) = \mu,\ \Delta_F(F^{-1}(\beta)) = \delta \right\}.$$

When it is clear from the context, we will use the notation $\Delta_F(\beta)$ to denote the absolute mean spread of a distribution around its β quantile, instead of $\Delta_F(F^{-1}(\beta))$. These two families assume exact knowledge of the true value of $\Delta_F(F^{-1}(\beta))$. In Section 3.4, we relax this assumption by admitting *bounds* on the value of $\Delta_F(F^{-1}(\beta))$.

We will show that there exist solutions to the minimax regret problems under $\mathcal{D}_{\mu,\delta}$ and $\mathcal{D}_{\mu,\delta,+}$. Due to the fact that all information required is first-order, the problem has a closed-form solution. Consider the demand distribution families $\mathcal{D}_{\mu,\delta}$ and $\mathcal{D}_{\mu,\delta,+}$, respectively. We first discuss the conditions on μ and δ that guarantee that these families are, respectively, nonempty. For the set $\mathcal{D}_{\mu,\delta}$, since the support is the whole real line, then any combination of values for μ and δ admits a feasible distribution. In particular, note that the following two-point support distribution is an element of $\mathcal{D}_{\mu,\delta}$:

$$D = \begin{cases} \mu - (1-\beta)\delta, & \text{w.p. } \beta, \\ \mu + \beta\delta, & \text{w.p. } 1-\beta. \end{cases} \tag{3.3}$$

However, this is not true once the support is the nonnegative real line, as it is for $\mathcal{D}_{\mu,\delta,+}$. The following proposition states conditions on μ and δ that are necessary and sufficient for the existence of a distribution in $\mathcal{D}_{\mu,\delta,+}$.

Proposition 3.2.1. *The distribution family $\mathcal{D}_{\mu,\delta,+}$ is nonempty if and only if $\mu-(1-\beta)\delta \geq 0$.*

Proof. Suppose $\mu - (1-\beta)\delta \geq 0$. Consider the two-point support distribution (3.3). This distribution is an element of $\mathcal{D}_{\mu,\delta,+}$, proving that $\mathcal{D}_{\mu,\delta,+}$ is nonempty. To prove the reverse implication, suppose that $\mathcal{D}_{\mu,\delta,+}$ is nonempty. Let F be a distribution in $\mathcal{D}_{\mu,\delta,+}$ where $F^{-1}(\beta) = w$ for some $w \geq 0$. Note that

$$\begin{aligned} \mu - w &= E(D-w)^+ - E(w-D)^+ \\ &= (1-\beta)E(D-w|D \geq w) - \beta E(w-D|D \leq w) \\ &= (1-\beta)\left\{E(D|D \geq w) - E(D|D \leq w)\right\} + E(D-w|D \leq w) \end{aligned}$$

$$= (1-\beta)\delta + E(D|D \leq w) - w. \tag{3.4}$$

Thus, this implies that $\mu - (1-\beta)\delta = E(D|D \leq w)$, which is nonnegative (since D has nonnegative support). □

For any values of μ and $\delta > 0$, the minimax regret problem over $\mathcal{D}_{\mu,\delta}$ is well-defined. For the distribution family $\mathcal{D}_{\mu,\delta,+}$ however, we must require that μ and δ satisfy the condition $\mu - (1-\beta)\delta \geq 0$ (Proposition 3.2.1). Denote the minimax regret values under $\mathcal{D}_{\mu,\delta}$ and $\mathcal{D}_{\mu,\delta,+}$ as $\rho^*_{\mu,\delta}$ and $\rho^*_{\mu,\delta,+}$, respectively. In the next two theorems, we will characterize in closed-form an expression for the minimax regret solutions, which we denote as $y^*_{\mu,\delta}$ and $y^*_{\mu,\delta,+}$, respectively.

Theorem 3.2.2. *Consider the nonempty set $\mathcal{D}_{\mu,\delta}$ consisting of all distributions with mean μ and AMS (at the β quantile) δ. Then the minimax regret and minimax regret quantity are:*

$$\rho^*_{\mu,\delta} = \beta(1-\beta)\delta,$$
$$y^*_{\mu,\delta} = \mu + (2\beta - 1)\delta.$$

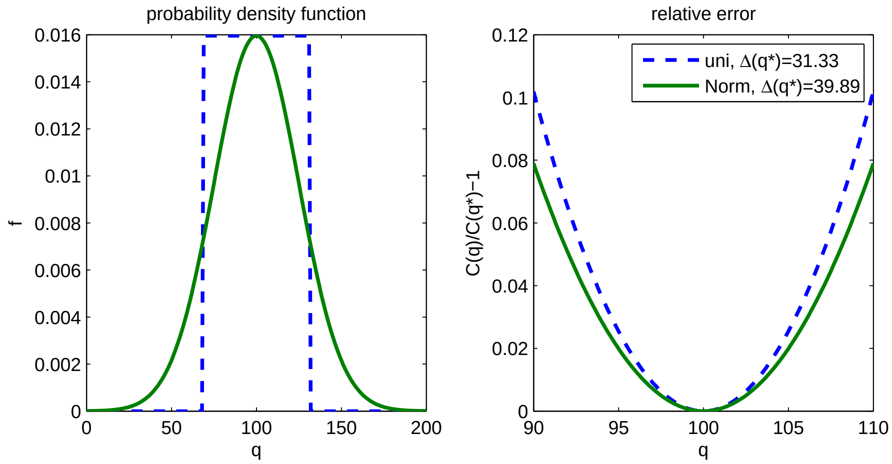
Theorem 3.2.3. *Consider the nonempty set $\mathcal{D}_{\mu,\delta,+}$ consisting of all nonnegative distributions with mean μ and AMS (at the β quantile) δ. Then the minimax regret and minimax regret quantity are:*

$$\rho^*_{\mu,\delta,+} = \frac{1}{\mu}\beta(1-\beta)\delta(\mu - (1-\beta)\delta),$$
$$y^*_{\mu,\delta,+} = \frac{1}{\mu}(\mu - (1-\beta)\delta)(\mu + \beta\delta).$$

The proofs of Theorems 3.2.2 and 3.2.3 are discussed in Appendix B (Sections B.9 and B.10). The following is a sketch of the proofs. For a fixed z and y, the optimization problem in (3.1) reduces to a linear semi-infinite program (LSIP) (Goberna & Lopez, 1998). We take the dual formulation of this LSIP. Unlike for a finite LP, a positive duality gap may exist for an LSIP. However, since we are able to identify primal and dual solutions that achieve the same cost, then weak duality guarantees that these solutions are primal and dual optimal, respectively.

Our choice of distribution families $\mathcal{D}_{\mu,\delta}, \mathcal{D}_{\mu,\delta,+}$ is motivated by the connection between AMS information and the regret function. We observe that *a distribution with a small AMS at the newsvendor quantile, $\Delta_F(\beta)$, typically has a steep expected profit function, $\Pi_F(y)$.* Since the regret function under a distribution F (i.e., $\Pi_F(F^{-1}(\beta)) - \Pi_F(y)$) is simply a transformation of the expected profit function, then this means that a small AMS at the newsvendor quantile also implies a steep regret function. We briefly discuss the basis

Figure 3-1: Regret function under a normal and a uniform distribution when $\beta = 0.5$.



behind this observation. Note that the slope of $\Pi_F(y)$ (assuming a continuous distribution) is $\Pi'(y) = \beta - F(y)$. At the point $y = F^{-1}(\beta)$, the slope is equal to zero, and the rate of change of the slope around this neighborhood depends on how fast the probability density (or the spread) changes at $F^{-1}(\beta)$. As an illustration, see the left plot of Figure 3-1 which shows a uniform and a normal distribution, both with mean 100. If $\beta = 0.5$, the optimal order quantity under both distributions is 100 units. However, under the uniform distribution (which has the smaller value of $\Delta_F(\beta)$), the regret function is steeper, as seen in the right plot of Figure 3-1. Therefore, even simple first-order information already provides strong insight into the impact of suboptimal decisions on the regret.

Due to this strong connection between regret and absolute mean spread information, we can prove attractive optimality properties of the policies using AMS information, as shown in the next section.

3.3 Optimality properties of policies using spread information

In this section, we prove a general result (Theorem 3.3.2 and Theorem 3.3.4) that under a large class of demand distributions, the policies using AMS information are near-optimal when profit margins are high, which is the case in most realistic settings.

Let us first compare the performance of our policies to other distributionally robust policies that use common types of information. One such policy is by Scarf (1958) who uses the mean, variance and nonnegativity information in a max-min framework (we denote this policy as MM-$\mu\sigma+$). Another recent one is Perakis & Roels (2008) who use mean, variance and nonnegativity information in a minimax regret framework (which we denote as MR-$\mu\sigma+$). Through the following experiment, we compare these two policies to the minimax regret policy using mean, AMS and nonnegativity information (which we denote MR-$\mu\delta+$). First, fix a demand distribution F and a profit margin β. Using the distribution's exact values for the mean μ, standard deviation σ, and AMS (at the β quantile) δ, compute the three partial information policies. Since we know the true demand distribution, we know the actual relative regret incurred by each policy. That is, for a given ordering quantity y, the actual relative regret is:

$$\text{Actual relative regret} = \frac{\Pi_F(F^{-1}(\beta)) - \Pi_F(y)}{\Pi_F(F^{-1}(\beta))}. \tag{3.5}$$

Figures 3-2–3-3 compare the actual relative regret of the three policies (MR-$\mu\delta+$, MR-$\mu\sigma+$, M-$\mu\sigma+$) as a function of the profit margin, under fourteen different demand distributions (see Figures C-3 and C-4 in Appendix C for a plot of the demand densities). In almost all of the distributions, the MR-$\mu\delta+$ policy clearly dominates the other two policies. This is because, even though it is using only first-order information, the MR-$\mu\delta+$ policy captures quantile-specific spread information. In contrast, variance is a static information that does not necessarily provide insight into ordering at high profit margins. This experiment highlights that the type of information used in the minimax regret framework is important in having near-optimal policies under high profit margin environments.

In the remainder of this section, we will prove that a mild technical condition on the distribution guarantees the near-optimality of our policies under high profit margins. First, we define the failure rate of a random variable.

Definition 3.3.1 (Failure rate)**.** *Let D be a random variable with distribution F and density f. The failure rate of D is defined as $r(x) \triangleq f(x)/(1 - F(x))$.*

If the failure rate of D satisfies $\lim_{x\to\infty} r(x) > 0$, then we can prove that its AMS value, $\Delta_F(\beta)$, does not grow faster than $\frac{1}{1-\beta}$ as $\beta \to 1$. This is formalized in the following lemma.

Lemma 3.3.1. *If D has a failure rate r such that $\lim_{x\to\infty} r(x) > 0$, then $\lim_{\beta\to 1}(1 - \beta)\Delta_F(\beta) = 0$.*

Proof. Note that for any $t \in \Re$,

$$\int_{-\infty}^{t} F(u)du = F(t)E_F\left(t - D|D \le t\right),$$

$$\int_{t}^{\infty} \left(1 - F(u)\right) du = (1 - F(t))E_F\left(D - t|D \ge t\right),$$

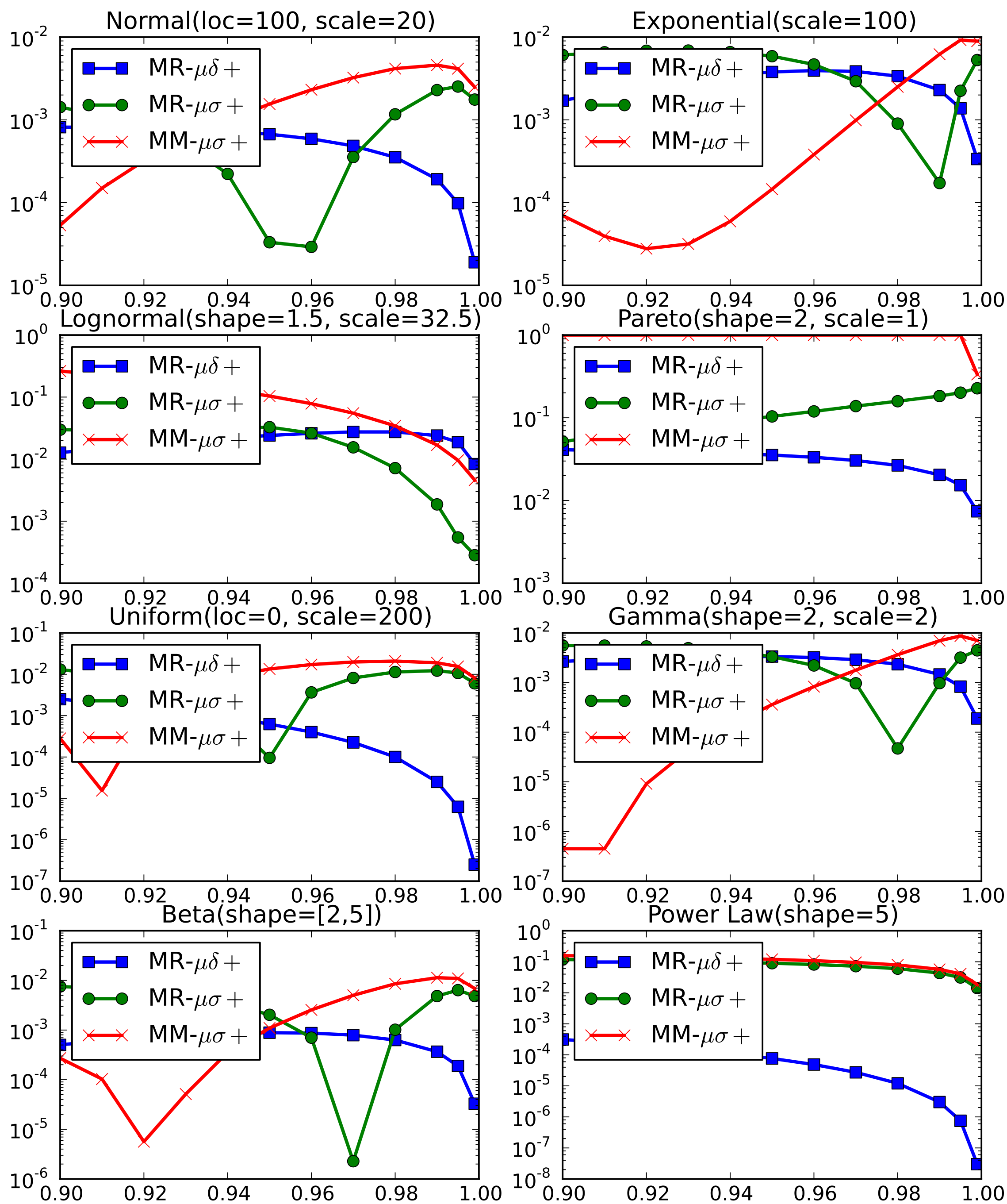
implying that

$$\lim_{\beta\to 1}(1-\beta)\Delta_F(\beta) = \lim_{t\to\infty}\left(\int_t^{\infty}(1 - F(u))du + \frac{1 - F(t)}{F(t)}\int_{-\infty}^{t} F(u)du\right).$$

The limit of the first term in the summation is clearly zero. By L'Hopital's rule, the second term goes to zero if $\frac{F(t)(1-F(t))^2}{f(t)} = \frac{1}{r(t)}F(t)(1 - F(t))$ goes to zero. Since $\lim_{t\to\infty} r(t) > 0$, then $\lim_{t\to\infty} \frac{1}{r(t)}$ exists. Thus,

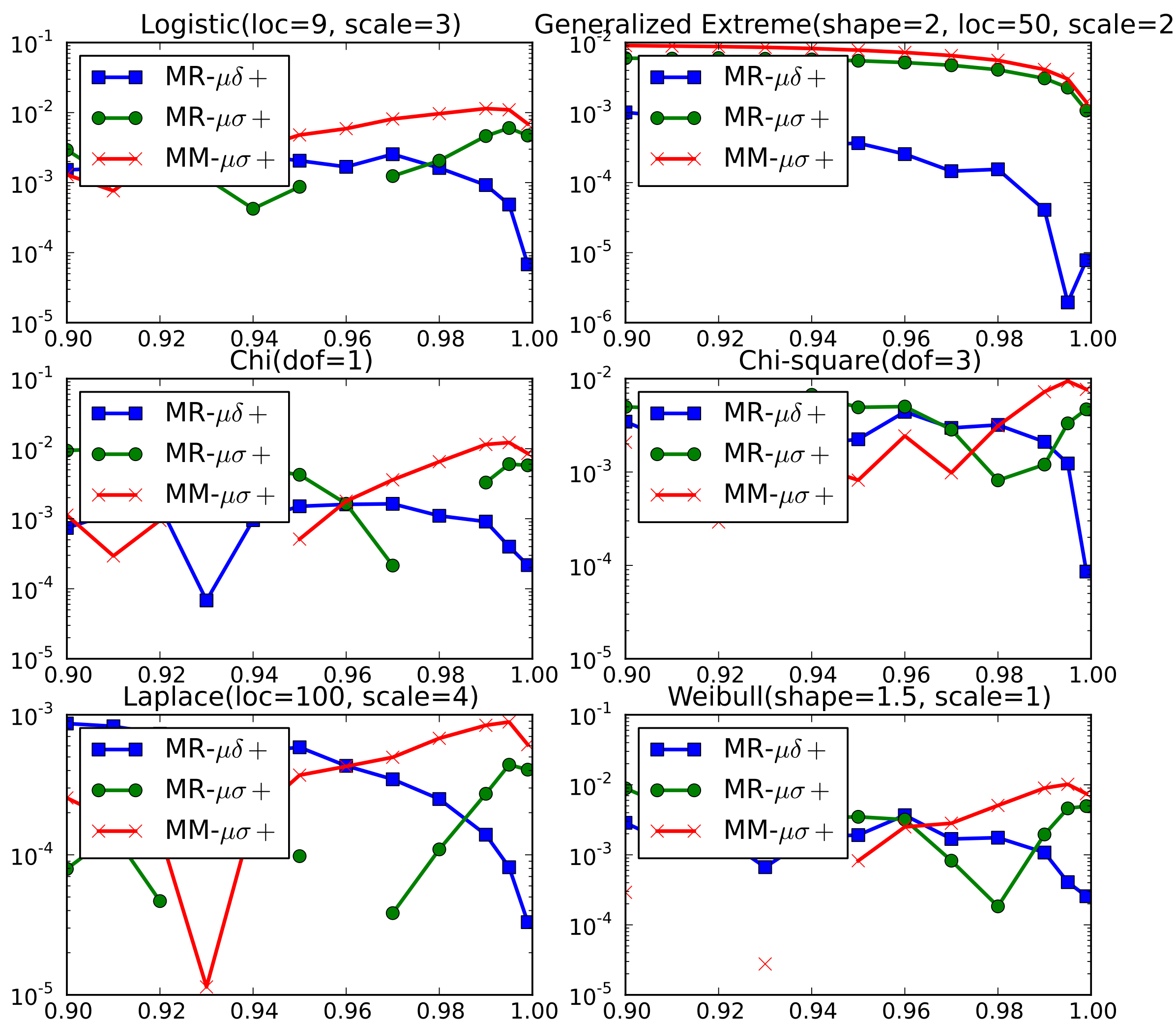
$$\lim_{t\to\infty}\frac{1}{r(t)}F(t)(1 - F(t)) = \lim_{t\to\infty}\frac{1}{r(t)} \times \lim_{t\to\infty} F(t)(1 - F(t)) = 0.$$

Figure 3-2: Actual relative regret of partial information policies plotted against the profit margin β.



The regret incurred by each policy is exact. The mean, variance and AMS for each distribution is known exactly by each policy.

Figure 3-3: Actual relative regret of partial information policies plotted against the profit margin β.



The regret incurred by each policy is estimated using 100,000 sample points. The mean and variance for each distribution is known exactly. The AMS is estimated using 100,000 sample points.

□

Remark Most common demand distributions satisfy the condition of Lemma 3.3.1. That is, the failure rate $r(t)$ is strictly positive in the limit as t approaches infinity. Moreover, if a random variable D has an increasing failure rate (IFR), then this condition is guaranteed to be met. This is because for any point s in the interior of the support of D, $\lim_{t\to\infty} r(t) \geq r(s) = \frac{f(s)}{1-F(s)} > 0$.

In the next theorem, we prove that if the true demand distribution has a failure rate r such that $\lim_{x\to\infty} r(x) > 0$, then the actual regret of ordering using our policies goes to zero as the profit margin approaches 1. This attractive optimality property of our policies is a consequence of its AMS having bounded growth in β (Lemma 3.3.1). Since most common distributions assumed in inventory management have this property, this implies that our policies would achieve near-optimality under most demand distributions.

Theorem 3.3.2. *Let D have a failure rate r such that $\lim_{x\to\infty} r(x) > 0$, and that $E_F(D) = \mu$ and $\Delta_F(\beta) = \delta_\beta$. Then, the actual regret of ordering y^*_{μ,δ_β} goes to zero as $\beta \to 1$. If in addition, D is nonnegative, then the actual regret of ordering $y^*_{\mu,\delta_\beta,+}$ goes to zero as $\beta \to 1$.*

Proof. Let us prove the result for the case without nonnegativity. For a fixed β, the actual regret of ordering y^*_{μ,δ_β} is $\Pi_F(F^{-1}(\beta)) - \Pi_F(y^*_{\mu,\delta_\beta})$. Moreover, since $F \in \mathcal{D}_{\mu,\delta_\beta}$, we have that $\Pi_F(F^{-1}(\beta)) - \Pi_F(y^*_{\mu,\delta_\beta}) \leq \rho^*_{\mu,\delta_\beta} = \beta(1-\beta)\delta_\beta$. From Lemma 3.3.1, we have that $\rho^*_{\mu,\delta_\beta} \to 0$ as $\beta \to 1$, implying that the actual regret goes to zero. Let us now prove the result for nonnegative distributions. The actual regret of ordering $y^*_{\mu,\delta_\beta,+}$ is $\Pi_F(F^{-1}(\beta)) - \Pi_F(y^*_{\mu,\delta_\beta,+})$, which is bounded above by $\rho^*_{\mu,\delta_\beta,+}$. Since $\mathcal{D}_{\mu,\delta_\beta,+} \subset \mathcal{D}_{\mu,\delta_\beta}$, we have that $\rho^*_{\mu,\delta_\beta,+} \leq \rho^*_{\mu,\delta_\beta}$. Following from the first result, this implies that $\rho^*_{\mu,\delta_\beta,+} \to 0$ as $\beta \to 1$. □

In fact, we can prove an even stronger result. Theorem 3.3.4 below states that the *relative* regret of ordering our policies disappears as the profit margin approaches one. The proof requires the following lemma.

Lemma 3.3.3. *If D is a random variable with $E_F(D) = \mu$ and $\Delta_F(\beta) = \delta$, then the optimal newsvendor profit is $\Pi_F(F^{-1}(\beta)) = \beta(\mu - (1-\beta)\delta)$.*

Proof. The result can be established through the following arithmetic arguments.

$$\begin{aligned}
\Pi_F(F^{-1}(\beta)) &= E_F\left(\min\{F^{-1}(\beta), D\}\right) - (1-\beta)F^{-1}(\beta), \\
&= E_F\left(\min\{F^{-1}(\beta), D\} | D \leq F^{-1}(\beta)\right) \Pr(D \leq F^{-1}(\beta)) \\
&\quad + E_F\left(\min\{F^{-1}(\beta), D\} | D > F^{-1}(\beta)\right) \Pr(D > F^{-1}(\beta)) - (1-\beta)F^{-1}(\beta), \\
&= E_F\left(D | D \leq F^{-1}(\beta)\right)\beta + F^{-1}(\beta)(1-\beta) - (1-\beta)F^{-1}(\beta), \\
&= \beta E_F\left(D | D \leq F^{-1}(\beta)\right) = \beta(\mu - (1-\beta)\delta),
\end{aligned}$$

where the last equality is established in (3.4). □

Theorem 3.3.4. *Let D have a failure rate r such that $\lim_{x\to\infty} r(x) > 0$, and that $E_F(D) = \mu$ and $\Delta_F(\beta) = \delta_\beta$. Then, the actual relative regret of ordering y^*_{μ,δ_β} goes to zero as $\beta \to 1$. If in addition, D is nonnegative, then the actual relative regret of ordering $y^*_{\mu,\delta_\beta,+}$ goes to zero as $\beta \to 1$.*

Proof. Let us first prove the result for the case without nonnegativity. If F is in the family $\mathcal{D}_{\mu,\delta_\beta}$, then for a fixed β, the relative regret of ordering y^*_{μ,δ_β} is

$$\frac{\Pi_F(F^{-1}(\beta)) - \Pi_F(y^*_{\mu,\delta_\beta})}{\Pi_F(F^{-1}(\beta))} \le \frac{\rho^*_{\mu,\delta_\beta}}{\Pi_F(F^{-1}(\beta))} = \frac{(1-\beta)\delta_\beta}{\mu - (1-\beta)\delta_\beta},$$

where the last equality follows from Lemma 3.3.3 and Theorem 3.2.2. Moreover, from Lemma 3.3.1 the right-hand side expression goes to zero as $\beta \to 1$. If D is nonnegative, then the actual relative regret of ordering $y^*_{\mu,\delta_\beta,+}$ is

$$\frac{\Pi_F(F^{-1}(\beta)) - \Pi_F(y^*_{\mu,\delta_\beta,+})}{\Pi_F(F^{-1}(\beta))} \le \frac{\rho^*_{\mu,\delta_\beta,+}}{\Pi_F(F^{-1}(\beta))} = \frac{(1-\beta)\delta_\beta}{\mu},$$

which follows from Lemma 3.3.3 and Theorem 3.2.3. The limit of the righthand side expression is zero as $\beta \to 1$, following from Lemma 3.3.1. □

In fact we can use Lemma 3.3.3 to prove that under *any* demand distribution family that is specified by mean and AMS (at the β quantile), there exists an order quantity that: (i) minimizes the maximum regret over the family, (ii) minimizes the maximum *relative* regret over the family, and (iii) maximizes the minimum expected profit over the family. Note that $\mathcal{D}_{\mu,\delta}$ and $\mathcal{D}_{\mu,\delta,+}$ are examples of such families. This is formalized in the following theorem.

Theorem 3.3.5. *Let $\mathcal{D}$ be a distribution set in which all elements have the same mean μ and AMS (at the β quantile) δ. Then there exists an order quantity that, under $\mathcal{D}$, minimizes the maximum regret and maximizes the minimum expected profit. If $\mu - (1-\beta)\delta > 0$, then it also minimizes the maximum relative regret.*

Proof. Denote μ as the common mean and δ as the common AMS (at the β quantile) under family $\mathcal{D}$. Denote by y^* the order quantity that maximizes the minimum expected profit, i.e.,

$$y^* = \arg\max_y \left\{ \min_{F\in\mathcal{D}} \Pi_F(y) \right\}.$$

From Lemma 3.3.3, any distribution in $\mathcal{D}$ has the same optimal newsvendor profit $\beta(\mu - (1-\beta)\delta)$. Therefore, the minimax regret problem is

$$\begin{aligned}\min_y \max_{F\in\mathcal{D}} \left\{\Pi_F(F^{-1}(\beta)) - \Pi_F(y)\right\} &= \min_y \max_{F\in\mathcal{D}} \left\{\beta(\mu - (1-\beta)\delta) - \Pi_F(y)\right\},\\ &= \beta(\mu - (1-\beta)\delta) - \max_y \min_{F\in\mathcal{D}} \Pi_F(y),\end{aligned}$$

which can be solved by choosing y^* as an order quantity. Using similar arguments, we have that the minimax relative regret problem is

$$\min_y \max_{F\in\mathcal{D}} \left\{ \frac{\Pi_F(F^{-1}(\beta)) - \Pi_F(y)}{\Pi_F(F^{-1}(\beta))} \right\} = \min_y \max_{F\in\mathcal{D}} \left\{ \frac{\beta(\mu-(1-\beta)\delta) - \Pi_F(y)}{\beta(\mu-(1-\beta)\delta)} \right\},$$
$$= 1 - \frac{1}{\beta(\mu-(1-\beta)\delta)} \max_y \min_{F\in\mathcal{D}} \Pi_F(y),$$

which is also solved with the order quantity y^*. □

3.4 Interval Information on AMS

Under some types of distribution families, the minimax regret problem becomes intractable under interval information. For instance, to the best of our knowledge, there is no known solution procedure if information about the standard deviation comes in the form of bounds on its value.

In contrast, even if the only information available about $\Delta_F(\beta)$ are *upper and lower bounds* on its value, the minimax regret problem remains tractable. In fact, we can solve the problem in closed-form. Consider the following family of distributions:

$$\mathcal{D}_{\mu,\delta_L,\delta_U,+} = \left\{ F \mid E_F(\mathbb{1}_{[0,\infty]}(D)) = 1, E_F(D) = \mu \text{ and } \Delta_F(F^{-1}(\beta)) \in [\delta_L, \delta_U] \right\}.$$

The next proposition provides conditions on the values of μ, δ_L and δ_U that ensure the set $\mathcal{D}_{\mu,\delta_L,\delta_U,+}$ is nonempty. The proof can be found in Appendix B (Section B.10).

Proposition 3.4.1. *Distribution set $\mathcal{D}_{\mu,\delta_L,\delta_U,+}$ is nonempty if and only if $\mu-(1-\beta)\delta_L \geq 0$.*

The minimax regret problem under $\mathcal{D}_{\mu,\delta_L,\delta_u,+}$ is tractable and has a closed-form expression, which we denote as $y^*_{\mu,\delta_L,\delta_U,+}$ (we provide the expression and its derivation in the appendices, in Sections A.3 and B.10, respectively). From the closed-form expression, we observe that when $\delta_L = 0$ and $\delta_U \geq \frac{\mu}{1-\beta}$, the minimax regret solution $y^*_{\mu,\delta_L,\delta_U,+}$ is in fact equivalent to the minimax regret solution under only mean information (Perakis & Roels, 2008, Theorem 2). This is because under this case, the bounds on the AMS do not give any additional meaningful information.

3.5 Data-driven estimation of AMS

Most partial information policies assume exact knowledge of statistics of a distribution. In reality however, these are rarely if ever available. At most, they have to be estimated from historical demand data or forecasts. In this section, we introduce a data-driven estimator for $\Delta_F(\beta)$. Since estimation is subject to error, we also propose a procedure for providing confidence interval estimates of $\Delta_F(\beta)$ (Algorithm 1 below). These data-driven confidence intervals can then be used as input to the minimax regret policy in the previous section, i.e., as values for δ_L and δ_U.

Suppose $D_1, D_2, \ldots, D_n$ are i.i.d. random variables drawn from the same distribution F. The order statistics of this set is a reordering of the random variables in terms of nondecreasing values, expressed as $D_{(1)} \leq D_{(2)} \leq \cdots \leq D_{(n)}$. We can use the order statistics to estimate the AMS value at the β quantile, $\Delta_F(\beta)$. For any $\beta \in (0,1)$, we propose the following estimator:

$$\Delta_n \triangleq \frac{1}{n} \sum_{i=1}^{n} J_i(\beta, n) D_{(i)}, \tag{3.6}$$

where the weights are defined as:

$$J_i(\beta, n) \triangleq \begin{cases} -\frac{1}{\beta}, & \text{if } i \leq \beta n, \\ \frac{i-1+\beta(1-n)}{\beta(1-\beta)}, & \text{if } \beta n < i \leq \beta n + 1, \\ \frac{1}{1-\beta}, & \text{if } i > \beta n + 1, \end{cases}$$

for $i = 1, 2, \ldots, n$.

Note that Δ_n is a random variable whose realization depends on the specific values taken on by $D_1, D_2, \ldots, D_n$. In fact from (3.6), Δ_n is a linear combination of the order statistics or L-statistic. Govindarajulu & Mason (1983) prove that L-statistics of the form (3.6) can be written as a sum of independent random variables. In particular, we have that

$$\Delta_n = \Delta_F(\beta) + R_n + \frac{1}{n} \sum_{i=1}^{n} Z_n, \tag{3.7}$$

where Z_i are i.i.d. with mean 0, and R_n is such that $\sqrt{n} R_n \to 0$ a.s. as $n \to \infty$. Interested readers are referred to Appendix A (Section A.1) for a further discussion on the distribution of the estimator Δ_n.

From (3.7), Δ_n is a *biased* estimator of the absolute mean spread $\Delta_F(\beta)$, however the bias goes to zero as the sample size goes to infinity. We attempt to numerically estimate the bias $B \triangleq E(\Delta_n) - \Delta_F(\beta)$ through regression analysis. We provide the complete details in Appendix A (Section A.2). In what follows, we sketch the idea behind the analysis. The factors that affect B are the sample size n, the profit margin β, and the demand distribution F. We can vary values for each one of these factors and perform regression to estimate their effect on B. For instance, to estimate the effect of n, we first fix a distribution and profit margin. We estimate the bias for different values of n, giving us multiple bias–sample size pairs. We use these pairs to perform regression to estimate the relationship between the bias and sample size. We find that the relationship $B = \frac{1}{n} \Delta_F(F^{-1}(\beta))$ closely estimates the observed bias in many simulations. Motivated by this, we propose the following scaled estimator of AMS:

$$\tilde{\Delta}_n \triangleq \frac{n+1}{n} \Delta_n = \frac{n+1}{n^2} \sum_{i=1}^{n} J_i(\beta, n) D_{(i)}. \tag{3.8}$$

Stigler (1974) shows that under some technical conditions, L-statistics are asymptotically normally distributed. Thus, we have that

$$\frac{\Delta_n - E(\Delta_n)}{\sqrt{\text{Var}(\Delta_n)}} \to_d N(0,1). \tag{3.9}$$

We use this observation to propose a procedure to build confidence intervals around the AMS estimator. Define the following sample estimate of $n\text{Var}(\Delta_n)$ (see Section A.1 of Appendix A for a motivation behind the choice of estimator):

$$\begin{aligned} s_n^2 = & \frac{1}{n}\frac{1}{(1-\beta)^2}\sum_{i=\lceil (n+1)\beta \rceil}(d_{(i)} - \Delta_n^u)^2 + \frac{1}{n}\frac{1}{\beta^2}\sum_{i=1}^{\lceil (n+1)\beta\rceil - 1}(d_{(i)} - \Delta_n^l)^2 \\ & + \left(\sqrt{\frac{1-\beta}{\beta}}\left[\Delta_n^l - d_{(\lceil (n+1)\beta \rceil)}\right] + \sqrt{\frac{\beta}{1-\beta}}\left[\Delta_n^u - d_{(\lceil (n+1)\beta \rceil)}\right]\right)^2, \end{aligned} \tag{3.10}$$

where

$$\Delta_n^u = \frac{1}{n}\frac{1}{1-\beta}\sum_{i=\lceil (n+1)\beta\rceil}^{n} d(i), \quad \text{and} \quad \Delta_n^l = \frac{1}{n}\frac{1}{\beta}\sum_{i=1}^{\lceil (n+1)\beta\rceil -1} d(i).$$

The following Algorithm 1 builds $(1-\alpha)$-confidence intervals around the sample spread Δ_n, given a significance level $\alpha \in (0,1)$.

Algorithm 1 `Confidence Interval for the Absolute Mean Spread`

Require: A sequence of ranked scalars $(d_{(1)}, \dots, d_{(n)})$ and a significance level $\alpha \in (0,1)$
Ensure: A confidence interval $\Delta_n \pm \theta_n$.

1: Compute $\tilde{\Delta}_n, s_n$, and Z-statistic $z \triangleq Z_{1-\alpha/2}$
2: Return confidence interval $\tilde{\Delta}_n \pm z \times \frac{s_n}{\sqrt{n}}$

To demonstrate the validity of this confidence interval procedure, we perform the following experiment. We draw 10,000 sets of $n \in \{40, 80, 160, 320\}$ samples from some distribution (we use the demand distributions from Figure C-3 in Appendix C). Using Algorithm 1 and with a 5% level of significance, we construct 1,000 sample-based confidence intervals from the sets. The procedure is valid if the fraction of these confidence intervals that contain the true absolute mean spread is close to 95%. Table 3.1 shows this fraction when the samples are drawn from one of six distributions. Observe that for smaller sample sizes ($n = 40$ or 80), the true AMS lies in the confidence interval between 70% to 95% of the time. However for $n = 160$ or 320, the procedure is successful between 90% to 95% of the time.

3.6 Computational Experiments

The goal of the computational experiments in this section is to study the empirical performance of various partial information policies (including our proposed policies) when the information has to be estimated from samples. We will assume that all the methods have

Table 3.1: Fraction (%) of sample-based confidence intervals that contain the true spread.

	Normal(loc=100, scale=20)				Uniform(loc=0, scale=200)				Exponential(scale=100)			
β	$n=40$	$n=80$	$n=160$	$n=320$	$n=40$	$n=80$	$n=160$	$n=320$	$n=40$	$n=80$	$n=160$	$n=320$
0.95	80.2	85.9	89.3	92.9	94.1	95.8	95.5	94.8	70.3	79.8	87.8	91.2
0.9	86.4	91.5	90.9	94.2	95.2	94.5	95.2	93.6	77.9	85.3	89.9	92.5
0.8	92.5	94.4	92.2	94.0	95.6	96.0	95.7	95.5	86.1	90.1	92.0	92.8
0.6	93.7	94.1	94.1	94.4	94.9	94.4	94.6	94.9	89.7	93.2	93.6	93.2
0.4	94.3	93.7	93.7	94.1	94.9	95.9	94.6	96.0	93.4	94.8	94.6	95.3
0.2	91.4	94.2	94.4	95.6	94.0	95.5	95.2	96.0	90.0	92.7	94.8	95.0
	Gamma(shape=2, scale=2)				Beta(shape=[2,5])				Power Law(shape=5)			
β	$n=40$	$n=80$	$n=160$	$n=320$	$n=40$	$n=80$	$n=160$	$n=320$	$n=40$	$n=80$	$n=160$	$n=320$
0.95	72.4	81.8	88.6	92.3	79.6	87.6	90.2	91.9	91.8	93.5	94.4	96.6
0.9	79.9	85.5	90.8	92.1	86.9	89.8	92.8	93.7	92.1	95.3	94.6	95.2
0.8	86.6	90.4	91.5	93.7	91.1	92.6	93.8	94.7	95.1	93.3	95.3	95.7
0.6	90.5	92.7	93.0	93.8	92.7	93.8	93.8	93.6	92.5	94.7	93.0	94.9
0.4	93.1	93.0	93.8	95.9	93.3	93.9	94.6	95.1	91.0	94.1	95.8	94.8
0.2	92.5	93.8	94.9	94.3	94.6	94.1	94.8	94.2	88.1	91.9	93.3	94.5

the same information available to them, particularly, a *set of independent samples drawn from the true distribution*. Using these samples, we will use the following partial information policies to compute order quantities for the newsvendor problem.

MR-$\hat{\mu}$+: A minimax regret approach with only *mean* and *nonnegativity* information (Perakis & Roels, 2008). The sample mean is computed and used as an input.

MR-$\hat{\mu}\hat{\sigma}$: A minimax regret approach with only *mean* and *variance* information (Yue *et al.*, 2006). The sample mean and sample variance is computed and used as an input.

MR-$\hat{\mu}\hat{\sigma}$+: A minimax regret approach with only *mean*, *variance*, and *nonnegativity* information (Perakis & Roels, 2008). The sample mean and sample variance is computed and used as an input.

MR-$\hat{\mu}\hat{\delta}$+: A minimax regret approach with only *mean*, *AMS*, and *nonnegativity* information. The sample mean and the *point* AMS estimate from (3.8) are computed and used as an input to compute the policy, according to Theorem 3.2.3.

MR-$\hat{\mu}\hat{\delta}_L\hat{\delta}_U$+: A minimax regret approach with only *mean*, *AMS* and *nonnegativity* information. The sample mean and the 95% *confidence intervals* for the AMS estimate (using Algorithm 1 above) are computed and used as an input to compute the policy, according to Theorem A.3.1.

The first four policies have been previously proposed from recent works (Yue *et al.*, 2006; Perakis & Roels, 2008). The last policy is our robust policy using first-order information (Theorem 3.2.3 and Theorem A.3.1). Note that once the distribution information is known, policies MR-$\hat{\mu}$+, MR-$\hat{\mu}\hat{\delta}$+ and MR-$\hat{\mu}\hat{\delta}_L\hat{\delta}_U$+ can be computed in closed-form. Policy MR-$\hat{\mu}\hat{\sigma}$, on the other hand, requires optimizing a function using gradient method (Yue *et al.*, 2006). To compute the policy MR-$\hat{\mu}\hat{\sigma}$+ requires solving for the intersection of two functions using bisection method. However, each iteration of the bisection method requires solving an optimization problem using gradient descent. Therefore, it can become computationally inefficient compared to the other methods.

For each ordering policy, we first fix a distribution and a profit margin $\beta \in \{0.9, 0.95, 0.99, 0.995\}$. The distributions we use are Uniform, Normal, Exponential, Lognormal, Pareto, Gamma, Beta, and Power Law distributions (see Figure C-3 in Appendix C for a plot of their densities). We also fix a sample size n, where $n \in \{20, 40, 80, 160\}$. We then draw n independent samples from the distribution, estimate the required information, and compute the corresponding order quantity. We also calculate the actual relative regret (3.5) under the distribution F. Finally, we take the average of the relative regrets over 100 repetitions of the same experiment.

Using the results of our experiments, we would like to address the question of *when should one use point estimates of AMS and when to be conservative and use confidence intervals*. Table 3.2 summarizes the average relative regret of the two AMS-based policies

Table 3.2: Average relative regret (%) of policies using sample estimates of AMS information.

		Uniform				Normal			
		$n = 20$	$n = 40$	$n = 80$	$n = 160$	$n = 20$	$n = 40$	$n = 80$	$n = 160$
$\beta = 0.9$	MR-$\hat{\mu}\hat{\delta}$+	0.6	0.4	0.3	0.3	0.3	0.2	0.1	0.1
	MR-$\hat{\mu}\hat{\delta}_L\hat{\delta}_U$+	0.5	0.2	0.1	0.1	1.0	0.6	0.4	0.3
$\beta = 0.95$	MR-$\hat{\mu}\hat{\delta}$+	0.3	0.1	0.1	0.1	0.2	0.1	0.1	0.1
	MR-$\hat{\mu}\hat{\delta}_L\hat{\delta}_U$+	0.8	0.3	0.1	0.0	0.7	0.5	0.4	0.3
$\beta = 0.99$	MR-$\hat{\mu}\hat{\delta}$+	0.3	0.1	0.5	0.0	0.2	0.1	0.0	0.0
	MR-$\hat{\mu}\hat{\delta}_L\hat{\delta}_U$+	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1
$\beta = 0.995$	MR-$\hat{\mu}\hat{\delta}$+	0.3	0.1	0.0	0.0	0.3	0.1	0.0	0.0
	MR-$\hat{\mu}\hat{\delta}_L\hat{\delta}_U$+	0.1	0.1	0.1	0.0	0.1	0.1	0.0	0.1
		Exponential				Lognormal			
		$n = 20$	$n = 40$	$n = 80$	$n = 160$	$n = 20$	$n = 40$	$n = 80$	$n = 160$
$\beta = 0.9$	MR-$\hat{\mu}\hat{\delta}$+	2.3	1.3	0.7	0.4	7.0	3.7	2.6	1.8
	MR-$\hat{\mu}\hat{\delta}_L\hat{\delta}_U$+	2.7	1.8	1.2	0.9	8.5	5.1	3.2	2.0
$\beta = 0.95$	MR-$\hat{\mu}\hat{\delta}$+	2.2	1.3	0.8	0.6	8.8	5.3	3.8	3.0
	MR-$\hat{\mu}\hat{\delta}_L\hat{\delta}_U$+	4.0	3.1	2.5	2.0	8.5	5.4	3.9	3.0
$\beta = 0.99$	MR-$\hat{\mu}\hat{\delta}$+	2.7	1.1	0.6	0.4	10.8	6.6	4.4	3.4
	MR-$\hat{\mu}\hat{\delta}_L\hat{\delta}_U$+	2.0	1.2	1.4	1.3	9.4	6.3	5.7	5.7
$\beta = 0.995$	MR-$\hat{\mu}\hat{\delta}$+	3.3	1.4	0.6	0.3	12.6	7.1	4.4	2.6
	MR-$\hat{\mu}\hat{\delta}_L\hat{\delta}_U$+	2.3	1.0	0.6	0.8	10.4	6.3	5.0	5.5

MR-$\hat{\mu}\hat{\delta}$+ and MR-$\hat{\mu}\hat{\delta}_L\hat{\delta}_U$+ under four demand distributions. Observe that in majority of the cases, MR-$\hat{\mu}\hat{\delta}$+ clearly has a smaller average relative regret than MR-$\hat{\mu}\hat{\delta}_L\hat{\delta}_U$+. The only instances when MR-$\hat{\mu}\hat{\delta}_L\hat{\delta}_U$+ has a slightly smaller regret is when the sample size is small ($n = 20$ or 40) while the profit margin is extremely high ($\beta = 0.99$ or 0.995). That is, we can infer that confidence interval estimates for AMS produces order quantities that are often conservative compared to the quantities produced from point estimates. Thus, in the remainder of this section, we will only report the results for policy MR-$\hat{\mu}\hat{\delta}$+.

Tables 3.4 and 3.5 report the average relative regret incurred by policies MR-$\hat{\mu}$+, MR-$\hat{\mu}\hat{\sigma}$, MR-$\hat{\mu}\hat{\sigma}$+ and MR-$\hat{\mu}\hat{\delta}$+ under high and extremely high profit margins, respectively. First, we observe (Table 3.4) that when profit margins are high (but not extremely so), the four policies incur average regrets of the same magnitude under many instances. Surprisingly, even MR-$\hat{\mu}$+ which only uses mean information, sometimes incurs relative regrets on par with the other three policies that use additional information. However, this is not the case once profit margins are extremely high (Table 3.5), which is the case in many realistic settings. Using only mean information in this setting results in clearly suboptimal decisions. MR-$\hat{\mu}$+ incurs average relative regrets in the range of 10% to as much as 25%. By including additional variance information, MR-$\hat{\mu}\hat{\sigma}$ usually incurs smaller average relative regrets than

MR-$\hat{\mu}$+, in both environments of high and extremely high profit margins. However, when the demand distribution is skewed (Exponential, Lognormal, Pareto) and profit margins are extremely high, MR-$\hat{\mu}\hat{\sigma}$ can incur very large average relative regrets in the range of 10% to 30%. This is due to the fact that MR-$\hat{\mu}\hat{\sigma}$, even though it accounts for demand variability, ignores the variability around the particular optimal quantile. By adding additional non-negativity information, MR-$\hat{\mu}\hat{\sigma}$+ manages to decrease the regret significantly. Of the four policies, MR-$\hat{\mu}\hat{\sigma}$+ and MR-$\hat{\mu}\hat{\delta}$+ consistently achieve the smallest average relative regret in both profit margin environments. However, because of its computational inefficiency, solving for policy MR-$\hat{\mu}\hat{\sigma}$+ can take significantly longer than solving for policy MR-$\hat{\mu}\hat{\delta}$+. Table 3.3 summarizes the average run times of each of the methods (averaged over distributions, sample sizes, and profit margins) On average, solving for policy MR-$\hat{\mu}\hat{\delta}$+ takes 10 microseconds, whereas solving for MR-$\hat{\mu}\hat{\sigma}$+ takes 3×10^6 microseconds. Finally, we observe that while MR-$\hat{\mu}$+ and MR-$\hat{\mu}\hat{\sigma}$ both appear to incur a significantly larger regret when profit margins are extremely high, the performance of MR-$\hat{\mu}\hat{\delta}$+ seems to be robust with respect to the level of profit margin. Under highly skewed distributions (Exponential, Lognormal, Pareto), it incurs an average regret of 1% to 10%. Under all other demand distributions, its average relative regret is usually less than 1%. This is mostly unsurprising based on Theorem 3.3.4 that for large enough profit margins, the minimax regret policy using AMS information is near-optimal for high profit margins. However, the theorem assumes that exact knowledge of the AMS is known, whereas the AMS information in MR-$\hat{\mu}\hat{\delta}$+ is estimated from samples. Moreover, these estimates are prone to error since in the experiments MR-$\hat{\mu}\hat{\delta}$+ attempts to estimate AMS around a high quantile using only a few samples. However, Tables 3.4 and 3.5 indicate that policies using AMS information remain near-optimal even when information has to be estimated from (sometimes few) data.

3.7 Concluding Remarks

Many inventory management settings require decisions that have to be made in the presence of uncertain demand. Stochastic inventory problems model this uncertainty by assuming that demand is stochastic with a fully-specified probability distribution. Contrary to this assumption, in reality, managers often have to make decisions knowing only limited information about the demand distribution. In this chapter, we have proposed partial-information inventory policies that only require first-order information on demand. In particular, they only require mean and *absolute mean spread* (AMS) information. First, we demonstrated that the resulting policies that we propose are robust, since they are the solution to the minimax regret problem of a family of distributions. Second, we proved that the resulting minimax regret problem is tractable since they only require first-order information. Third, we showed that the policies we proposed are near-optimal for high profit margins under a large class of demand distributions. We showed that other distributionally-robust policies previously proposed in the literature do not exhibit this near-optimal behavior. Finally, we also proposed a sample estimator of a distribution's AMS. In computational experiments,

Table 3.3: Average run times of each method

Method	Average runtime (in microseconds)	Standard deviation (in microseconds)
MR-$\hat{\mu}$+	6.7	6.5
MR-$\hat{\mu}\hat{\sigma}$	283.0	278.6
MR-$\hat{\mu}\hat{\sigma}$+	3,233,427.6	7,361,979.1
MR-$\hat{\mu}\hat{\delta}$+	8.0	47.5
MR-$\hat{\mu}\hat{\delta}_L\hat{\delta}_U$+	10.5	13.6

we find that even though the AMS information is estimated from data, and the minimax regret policy that uses these estimates are near-optimal even when the sample size is small. Moreover, it clearly dominates other minimax regret policies using mean and variance. Through this, we can conclude that there is much value in making inventory decisions that incorporate some information about spread around the optimal quantile.

Table 3.4: Average relative regret (%) of policies using sample estimates of information under high profit margins ($\beta = 0.9, 0.95$).

		$\beta = 0.9$				$\beta = 0.95$			
Distribution	Policies	$n = 20$	$n = 40$	$n = 80$	$n = 160$	$n = 20$	$n = 40$	$n = 80$	$n = 160$
Uniform	MR-$\hat{\mu}$+	7.5	7.5	7.5	7.4	16.9	16.8	17.0	16.9
	MR-$\hat{\mu}\hat{\sigma}$	3.4	3.0	2.7	2.6	4.3	3.8	3.7	3.5
	MR-$\hat{\mu}\hat{\sigma}$+	1.9	1.8	1.5	1.4	0.7	0.3	0.2	0.1
	MR-$\hat{\mu}\hat{\delta}$+	0.6	0.4	0.3	0.3	0.3	0.1	0.1	0.1
Normal	MR-$\hat{\mu}$+	13.3	13.3	13.3	13.3	19.3	19.3	19.3	19.3
	MR-$\hat{\mu}\hat{\sigma}$	0.6	0.5	0.4	0.4	1.1	0.9	0.8	0.8
	MR-$\hat{\mu}\hat{\sigma}$+	0.4	0.3	0.2	0.1	0.1	0.1	0.0	0.0
	MR-$\hat{\mu}\hat{\delta}$+	0.3	0.2	0.1	0.1	0.2	0.1	0.1	0.1
Exponential	MR-$\hat{\mu}$+	2.1	1.2	0.8	0.5	7.7	7.3	7.1	7.1
	MR-$\hat{\mu}\hat{\sigma}$	4.2	2.9	2.2	1.8	7.8	6.2	5.9	5.5
	MR-$\hat{\mu}\hat{\sigma}$+	2.9	1.9	1.2	0.9	2.5	2.0	1.1	0.8
	MR-$\hat{\mu}\hat{\delta}$+	2.3	1.3	0.7	0.4	2.2	1.3	0.8	0.6
Lognormal	MR-$\hat{\mu}$+	7.9	5.3	2.8	2.1	7.8	5.2	3.9	2.7
	MR-$\hat{\mu}\hat{\sigma}$	15.6	13.0	13.3	8.7	13.6	8.6	5.9	4.4
	MR-$\hat{\mu}\hat{\sigma}$+	15.8	6.2	5.1	4.2	11.2	8.1	5.0	3.2
	MR-$\hat{\mu}\hat{\delta}$+	7.0	3.7	2.6	1.8	8.8	5.3	3.8	3.0
Pareto	MR-$\hat{\mu}$+	6.1	5.3	5.4	5.0	10.3	9.8	9.9	9.9
	MR-$\hat{\mu}\hat{\sigma}$	3.7	3.9	2.9	2.8	4.2	3.6	3.7	2.3
	MR-$\hat{\mu}\hat{\sigma}$+	4.4	3.6	2.6	2.8	4.0	1.8	2.4	1.8
	MR-$\hat{\mu}\hat{\delta}$+	4.9	4.2	4.5	4.0	5.0	3.8	3.7	3.7
Gamma	MR-$\hat{\mu}$+	3.3	2.9	2.7	2.6	11.9	12.1	12.0	12.0
	MR-$\hat{\mu}\hat{\sigma}$	2.8	2.1	1.7	1.5	5.2	4.4	3.9	3.8
	MR-$\hat{\mu}\hat{\sigma}$+	1.9	1.4	1.0	0.7	1.9	0.9	0.7	0.5
	MR-$\hat{\mu}\hat{\delta}$+	1.6	0.8	0.6	0.4	1.5	0.9	0.6	0.5
Beta	MR-$\hat{\mu}$+	5.6	5.4	5.4	5.3	15.1	15.4	15.3	15.3
	MR-$\hat{\mu}\hat{\sigma}$	2.5	2.1	1.8	1.7	4.2	3.7	3.4	3.3
	MR-$\hat{\mu}\hat{\sigma}$+	1.7	1.1	1.0	0.9	1.2	0.4	0.4	0.3
	MR-$\hat{\mu}\hat{\delta}$+	0.8	0.4	0.2	0.1	0.7	0.4	0.2	0.1
Power Law	MR-$\hat{\mu}$+	14.9	14.9	14.9	14.9	20.2	20.2	20.2	20.2
	MR-$\hat{\mu}\hat{\sigma}$	0.3	0.3	0.2	0.2	0.4	0.3	0.2	0.2
	MR-$\hat{\mu}\hat{\sigma}$+	0.2	0.1	0.0	0.0	0.3	0.4	0.4	0.4
	MR-$\hat{\mu}\hat{\delta}$+	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0

[a] Highlighted rows correspond to the policies using AMS information.

Table 3.5: Average relative regret (%) of policies using sample estimates of information under extremely high profit margins ($\beta = 0.99, 0.995$).

		$\beta = 0.99$				$\beta = 0.995$			
Distribution	Policies	$n = 20$	$n = 40$	$n = 80$	$n = 160$	$n = 20$	$n = 40$	$n = 80$	$n = 160$
Uniform	MR-$\hat{\mu}$+	23.5	23.6	23.4	23.5	24.4	24.2	24.2	24.2
	MR-$\hat{\mu}\hat{\sigma}$	5.4	4.5	4.4	4.4	5.2	4.8	4.4	4.5
	MR-$\hat{\mu}\hat{\sigma}$+	1.1	1.2	1.2	1.2	1.0	1.0	1.1	1.1
	MR-$\hat{\mu}\hat{\delta}$+	0.3	0.1	0.5	0.0	0.3	0.1	0.0	0.0
Normal	MR-$\hat{\mu}$+	23.9	23.8	23.9	23.8	24.4	24.4	24.4	24.4
	MR-$\hat{\mu}\hat{\sigma}$	1.7	1.6	1.5	1.4	1.8	1.6	1.6	1.5
	MR-$\hat{\mu}\hat{\sigma}$+	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3
	MR-$\hat{\mu}\hat{\delta}$+	0.2	0.1	0.0	0.0	0.3	0.1	0.0	0.0
Exponential	MR-$\hat{\mu}$+	20.4	20.6	20.7	20.6	22.5	22.8	22.5	22.7
	MR-$\hat{\mu}\hat{\sigma}$	13.3	11.8	11.3	11.0	14.3	13.5	12.5	12.2
	MR-$\hat{\mu}\hat{\sigma}$+	0.9	0.4	0.3	0.1	0.5	0.4	0.3	0.2
	MR-$\hat{\mu}\hat{\delta}$+	2.7	1.1	0.6	0.4	3.3	1.4	0.6	0.3
Lognormal	MR-$\hat{\mu}$+	12.1	10.5	10.3	10.0	14.6	14.2	14.3	14.1
	MR-$\hat{\mu}\hat{\sigma}$	24.3	21.6	17.7	15.7	28.7	24.7	22.3	20.0
	MR-$\hat{\mu}\hat{\sigma}$+	7.2	5.2	3.7	2.1	7.0	5.3	2.5	1.7
	MR-$\hat{\mu}\hat{\delta}$+	10.8	6.6	4.4	3.4	12.6	7.1	4.4	2.6
Pareto	MR-$\hat{\mu}$+	17.6	17.6	17.8	18.1	19.5	19.6	19.9	20.1
	MR-$\hat{\mu}\hat{\sigma}$	9.3	8.1	6.6	6.2	11.5	9.8	8.8	7.7
	MR-$\hat{\mu}\hat{\sigma}$+	2.8	2.4	1.3	1.2	2.1	1.6	1.9	0.8
	MR-$\hat{\mu}\hat{\delta}$+	4.2	2.9	2.2	2.0	4.5	2.8	1.8	1.5
Gamma	MR-$\hat{\mu}$+	21.9	22.1	22.0	22.0	23.2	23.4	23.3	23.4
	MR-$\hat{\mu}\hat{\sigma}$	8.8	7.9	7.5	7.3	9.6	8.7	8.1	8.0
	MR-$\hat{\mu}\hat{\sigma}$+	0.4	0.2	0.2	0.1	0.3	0.3	0.3	0.3
	MR-$\hat{\mu}\hat{\delta}$+	1.6	0.7	0.3	0.2	1.9	0.9	0.3	0.2
Beta	MR-$\hat{\mu}$+	22.9	23.0	22.9	23.0	23.9	23.8	24.0	23.9
	MR-$\hat{\mu}\hat{\sigma}$	6.3	5.7	5.5	5.4	6.8	6.2	5.8	5.8
	MR-$\hat{\mu}\hat{\sigma}$+	0.5	0.5	0.5	0.5	0.6	0.7	0.7	0.6
	MR-$\hat{\mu}\hat{\delta}$+	0.9	0.3	0.1	0.1	1.0	0.4	0.1	0.1
Power Law	MR-$\hat{\mu}$+	24.1	24.1	24.1	24.1	24.6	24.6	24.5	24.6
	MR-$\hat{\mu}\hat{\sigma}$	0.4	0.3	0.2	0.2	0.4	0.3	0.2	0.2
	MR-$\hat{\mu}\hat{\sigma}$+	0.4	0.4	0.4	0.4	0.3	0.4	0.4	0.4
	MR-$\hat{\mu}\hat{\delta}$+	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

[a] Highlighted rows correspond to the policies using AMS information.

Table 3.6: Maximum regret achieved by SAA policy out of family $\mathcal{D}_{\mu,\delta,+}$, which includes the distribution used to generate the samples.

Distribution	β	Average max regret of SAA	Minimax regret	% difference
Uniform	0.7	21.1	14.7	43.3
	0.8	16.1	12.8	26.0
	0.9	9.1	8.1	12.7
	0.95	5.0	4.5	10.1
Normal	0.7	7.1	6.3	14.1
	0.8	6.7	5.2	28.2
	0.9	6.0	3.4	76.6
	0.95	5.2	2.0	158.4
Exponential	0.7	22.6	17.5	29.4
	0.8	22.5	19.2	17.1
	0.9	21.5	17.1	25.2
	0.95	20.2	12.6	59.8
Lognormal	0.7	16.1	12.6	27.6
	0.8	20.2	17.4	16.2
	0.9	26.9	22.4	20.4
	0.95	31.8	23.0	38.0
Pareto	0.7	0.445	0.320	39.2
	0.8	0.543	0.341	59.2
	0.9	0.659	0.328	100.8
	0.95	0.720	0.283	154.0
Gamma	0.7	0.767	0.653	17.4
	0.8	0.752	0.646	16.4
	0.9	0.712	0.512	39.0
	0.95	0.660	0.351	87.9
Beta	0.7	0.050	0.042	20.0
	0.8	0.044	0.039	14.6
	0.9	0.036	0.028	30.1
	0.95	0.029	0.017	68.5
Power Law	0.7	0.047	0.037	25.0
	0.8	0.032	0.028	15.3
	0.9	0.017	0.015	8.2
	0.95	0.009	0.008	7.8

Chapter 4

Business analytics for scheduling with random emergencies

4.1 Introduction

Allocating limited resources to a set of tasks is a problem encountered in many industries. It has applications in project management, bandwidth allocation, internet packet routing, job shop scheduling, hospital scheduling, aircraft maintenance, air traffic management, and shipping scheduling. In the past decades, the focus has been primarily on developing methods for optimal scheduling for deterministic problems. These approaches assume that all relevant information is available before the schedule is decided, and the parameters do not change after the schedule is made. In many realistic settings, however, scheduling decisions have to be made in the face of uncertainty. After deciding on a schedule, a resource may unexpectedly become unavailable, a task may take longer or shorter time than expected, or there might be an unexpected release of high-priority jobs (see Pinedo (2002) for an overview of stochastic scheduling models). Not accounting for these uncertainties may cause an undesirable impact, say, in a possible schedule interruption or in over-utilizing some resources. Birge (1997) demonstrated that in many real-world applications, when using stochastic optimization to model uncertainties explicitly, the results are superior compared to using a deterministic counterpart.

In this chapter, we study the problem of scheduling a known set of jobs when *there is an uncertain number of emergency jobs that may arrive in the future*. There are many interesting applications for this type of problem. For instance, Lamiri *et al.* (2008) describe the problem of scheduling surgeries in hospital intensive care units, where operating rooms are shared by two classes of patients: elective patients and emergency patients. Emergency cases arrive randomly but must be served immediately upon arrival. Elective cases can be delayed and scheduled for future dates. In scheduling the elective surgeries, the hospital needs to plan for flexibility (say, by having operating rooms on standby) to handle random arrivals of emergency cases.

This work is motivated by a project with a major electric and gas utility. We worked on improving scheduling of services for the utility's Gas business segment which faces uncertainty in its daily operations. In 2011, the Gas business segment of the utility generated several billion dollars in revenue. The following is a brief description of natural gas transmission and distribution in the United States. Natural gas is either produced (in the US Gulf Coast, midcontinent, and other sources) or imported (from the Middle East or South America). Afterwards, it is delivered to US interstate pipelines to be transmitted across the US. Once it reaches a neighborhood, the gas is delivered by a local gas utility, which owns and operates a network of gas pipelines used to deliver gas to the end customers. The gas utility involved in the project owns several of these local networks.

A major part of daily operations of the gas utility is the maintenance of the large gas pipeline network. This entails executing two types of jobs: (i) *standard jobs* and (ii) *emergency gas leak repair jobs*. The first type of jobs includes new gas pipeline construction, maintenance and replacement of gas pipelines, and customer requests. The key characteristics of standard jobs are that they have deadlines by when they must be finished, they are known several weeks to a few months in advance of their deadlines, and they are often mandated by regulatory authorities or required by customers. The second type of job is to attend to any reports of gas leaks. In the US, more than 60% of the gas transmission pipes are 40 years old or older (Burke, 2010). Most of them are composed of corrosive steel or cast-iron. Gas leaks are likely to occur on corroding bare steel or aging cast iron pipes, which pose a safety hazard especially if they occur near a populated location. If undetected, a gas leak might lead to a fire or an explosion. Such was the case in San Bruno, California in September 2010, where a corrosive pipe ruptured, causing a massive blast and fire that killed 8 people and destroyed 38 homes in the San Francisco suburb (Pipeline & Hazardous Materials Safety Administration, 2011a). To reduce the risk of such accidents occurring, company crews regularly monitor leak prone pipes to identify any leaks that need immediate attention. In addition, the company maintains an emergency hotline for reports of suspected gas leaks. It is the company's policy to attend to a report within 24 hours of receiving it. The key characteristic of emergency gas leak jobs is that they are unpredictable, they need to be attended to immediately, they require several hours to complete, and they happen with frequency throughout a day. The leaks that do not pose significant risk to the public are fixed later within regulatory deadlines dictated by the risk involved. These jobs are part of the standard jobs.

The company keeps a roster of *maintenance crews* to execute both types of jobs. The company has experienced significant crew overtime driven by both controllable factors (such as workforce management, scheduling processes, and information systems) and uncontrollable factors (such as uncertainty related to emergency leaks, diverse and unknown site conditions, and uncertainty in job complexity). Maintenance crews historically worked a significant proportion of their hours on overtime. An average crew member may work between 25% to 40% of his or her work hours on overtime pay. From our analysis, one of the major causes of overtime is suboptimal job scheduling and planning for the occurrence of emergencies. Currently, the company has no standard procedures or does not use quantitative methods for job scheduling and crew assignment. Past studies undertaken by the company suggested that a better daily scheduling process that optimizes daily resource allocation can provide a significant opportunity for achieving lower costs and better deadline compliance.

In this chapter, we study the utility's problem of daily resource allocation along with associated process and managerial factors. However, the models proposed and insights gained from this chapter have wider applicability in settings where resources have to be allocated under stochastic emergencies.

4.1.1 Literature Review and Our Contributions

Our work makes theoretical contributions in several key areas, as well as contributions to the utility's practice. We contrast our contributions with previous work found in the literature.

Modeling and problem decomposition. We develop a multiperiod model for the utility's operations under stochastic emergencies. Before realizing the number of emergencies, the utility has to decide a standard job's schedule (which date it will be worked on) and its crew assignment (the crew assigned to execute it). We model the problem as a stochastic mixed integer program (MIP).

Several practical limitations discussed in Section 4.3.1 (such as computational intractability, the utility's restrictive computing resources, and employees' perception of the model as a "black box") prevented the utility from implementing the multiperiod stochastic MIP model. Therefore, we propose a *two-phase decomposition* which addresses the original model's limitations. The first phase is a *job scheduling phase*, where standard jobs are scheduled so as to meet all the deadlines, but while evenly distributing work over all days (Section 4.4). This scheduling phase solves a deterministic MIP. The second phase is a *crew assignment phase*, which takes the standard jobs scheduled for each day from the first phase and assigns them to the available crews (Section 4.5). Since the job schedules are fixed, the assignment decisions on different days can be made independently. The assignment decisions must be made before arrivals of emergencies, hence, the assignment problem on each day is solved as a two-stage stochastic MIP.

This type of decomposition is similar to what is often done in airline planning problems (see for example Barnhart *et al.*, 2003), which in practice are solved sequentially due to the problem size and complexity. Airlines usually first solve a schedule design problem, which determines the flights flown during different time periods. Then in the next step, they decide which aircraft to assign to each flight depending on the forecasted demand for the flight. Airline planning problems are solved through deterministic models which are intractable due to its problem size. In contrast, the models in this chapter are stochastic in nature, adding a layer of modeling and computational difficulties.

LP-based heuristic for scheduling phase. In this work, we propose a heuristic for the NP-hard job scheduling problem based on solving its linear programming (LP) relaxation, and rounding the solution to a feasible schedule. The scheduling phase problem is equivalent to scheduling jobs on unrelated machines with the objective of minimizing makespan (Pinedo, 2002). In our problem, the dates are the "machines". The makespan is the maximum number of hours scheduled on any day. Note that a job can only be "processed" on dates before the deadline (the job's "processing set"). Scheduling problems with processing set restrictions are known to be NP-hard.

Some other common heuristics in the literature are list scheduling rules (Kafura & Shen, 1977; Hwang *et al.*, 2004). However these are applicable for problems with parallel machines. For the case of unrelated machines, a well-known algorithm by Lenstra *et al.* (1990) performs a binary search procedure, in each iteration solving the LP relaxation of an integer program, and then rounding the solution to a feasible schedule. Our algorithm is also based on solving an LP-relaxation, but it applies for unrelated machines with processing set restrictions. Moreover, it does not require initializing the algorithm with a binary search, therefore only solving an LP once. Since this heuristic is based on linear programming, in practice it solves very fast with commercial off-the-shelf solvers.

Performance guarantee for the LP-based heuristic. We are able provide a data-dependent performance guarantee for our proposed LP-based heuristic (Theorem 4.4.2). Lenstra *et al.* (1990) prove that the schedule resulting from their LP-based algorithm is guaranteed to have a makespan of no more than twice the optimal makespan. Their proof relies on graph theory. On the other hand, the bound we derive uses a novel technique based on *stochastic analysis*. Moreover, when the algorithm is initialized with a binary search, we can prove, using graph theoretic and stochastic arguments, a performance guarantee that is the minimum of 2 and a data-driven factor (Theorem A.7.1). Since, with real utility data, the data-driven factor is less than 2, we improve upon the bound by Lenstra *et al.* (1990) in realistic settings.

Algorithm for crew assignment under a stochastic number of emergencies. The assignment phase problem is a two-stage stochastic MIP. In the first stage the assignment of standard jobs to crews is determined, and in the second stage (after the number of emergencies is known) the assignment of emergencies to crews is decided. Most literature on problems of this type develops iterative methods to solve the problem. For instance, a common method is based on Benders' decomposition embedded in a branch and cut procedure (Laporte & Louveaux, 1993). However, if the second stage has integer variables, the second stage value function is discontinuous and non-convex, and optimality cuts for Benders' decomposition cannot be generated from the dual. Sherali & Fraticelli (2002) propose introducing optimality cuts through a sequential convexification of the second stage problem. There are other methods proposed to solve stochastic models of scheduling under uncertainty. For instance, Lamiri *et al.* (2008) introduce a local search method to plan for elective surgeries in the operating room scheduling problem. Godfrey & Powell (2001) introduce a method for dynamic resource allocation based on nonlinear functional approximations of the second-stage value function based on sample gradient information. However, since they are developed for general two-stage stochastic problems, these solution methods do not give insights on how resources should be allocated in anticipation of an uncertain number of emergencies.

We exploit the structure of the problem and of the optimal assignment and propose a simple and intuitive algorithm for assigning the standard jobs under a stochastic number of emergencies (Algorithm Stoch-LPT). This algorithm can be thought of as a generalization of the Longest-Processing-Time First (LPT) algorithm in the scheduling literature (Pinedo, 2002). We prove that this algorithm terminates with an optimal crew assignment for some special cases.

Models and heuristics for resource allocation with random emergencies. Our work is motivated by the specific problem of a gas distribution company. However, the models and algorithms we develop in this chapter are also applicable to other settings where resources need to be allocated in a flexible manner in order to be able to handle random future emergencies. As a specific example, in the operating room planning problem described in the introduction, the resources to be allocated are operating rooms. Elective surgeries and emergency surgeries are equivalent to standard jobs and gas leak repair jobs, respectively, in our problem.

Business analytics for a large US utility. We collaborated with a large multi-state utility on improving the scheduling of operations in its Gas business segment. The job scheduling and crew assignment optimization models described above are motivated by the company's resource allocation problem under randomly occurring emergencies. The job scheduling heuristic and the crew assignment heuristics described earlier are motivated from practical requirements, including the company's need for fast solution methods. We developed a Web-based planning tool based on these heuristics which is being piloted in one of the company's sites.

We also used our models to help the utility make strategic decisions about its operations. In simulations using actual data and our models, we highlight how different process changes impact crew-utilization and overtime labor costs. In this work, we analyzed three process changes: (i) maintaining an optimal inventory of jobs ready to be scheduled, (ii) having detailed crew productivity information, and (iii) increasing crew supervisor presence in the field. We demonstrate the financial impact of these new business processes on a hypothetical utility.

4.1.2 Outline

In Section 4.3, we present the job scheduling and crew assignment problem, and motivate the two-stage decomposition. In Section 4.4, we discuss the job scheduling phase, introduce an LP-based heuristic, and develop a data-driven performance guarantee of the heuristic. Section 4.5 discusses the crew assignment phase. In this section, we prove structural properties of the optimal solution, propose a crew assignment heuristic, and show that it terminates with an optimal solution for some cases. In Section 4.6, we discuss the development of planning tool, the pilot project, and how we used simulation and the models we developed for business analytics at the Gas business of a large multi-state utility company. Proofs not shown in the chapter can be found in Appendix B.

4.2 Gas utility operations and background

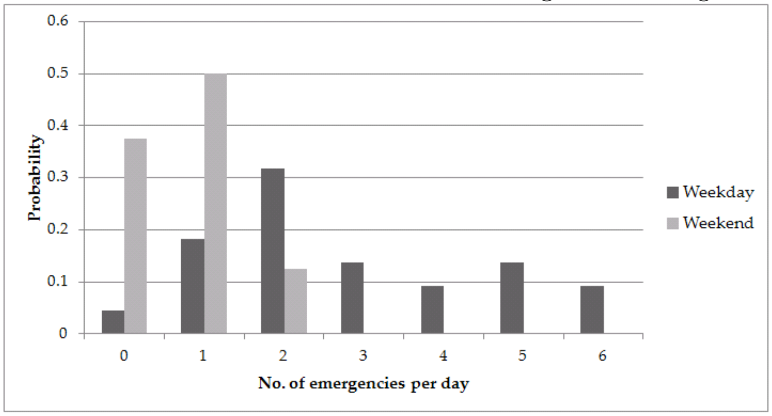
In this section, we give a background of the company operations and organization. The discussion serves to motivate the model, assumptions, and our choice of heuristics later in the chapter.

Gas utilities in the US operate large networks of gas pipelines. Some of these pipelines are aging and are composed of corrosive material. Some gas pipes that are still in service in many cities today are composed of corrosive cast iron that have been installed since the 1830's (Pipeline & Hazardous Materials Safety Administration, 2011b). Gas utility companies are involved in a government-mandated cast iron main replacement program which aims to replace many of the cast iron pipes into more durable steel or PVC pipes. To meet the requirements of this program, the company has a dedicated department called the *Resource Management Department* that sets yearly targets for standard jobs to be performed in the field and monitors the progress relative to these targets throughout the year. All targets are yearly and company-wide.

Standard jobs that occur within a geographical region (usually a town or several neighboring towns) are assigned to a *yard*. A yard is the physical company site which houses maintenance crews who are dispatched to complete the standard jobs. After the Resource Management Department decides on a company-wide target for standard jobs, it is translated into monthly targets for each yard based on yard size, number of workers available, and other characteristics of the region the yard serves. Several years ago, the utility expanded in the US from a string of mergers of small independent local utilities operating in towns. As a result, even today, separate yards belonging to the company operate independently. Small yards can have 10 crews, while large yards can have up to 30 crews, with each crew composed of two or three crew members. Each standard job has a deadline set by the Resource Management Department to ensure that the targets are met and the company does not incur the heavy regulatory fines for not meeting the requirements of the main replacement program. The company maintains a centralized database of standard jobs which lists each job's deadline, status (e.g. completed, pending or in progress), location, job type, other key job characteristics, and also information on all past jobs completed. A large yard can complete close to 500 standard jobs in one month. The focus of the project and hence of this chapter is on yard-level operations, which we describe below.

Daily yard operations. Each yard has a *resource planner* who is charged with making decisions about the yard's daily operations. At the start of each day, the resource planner reviews the pending standard jobs and their upcoming deadlines, and decides which jobs should be done by the yard that day. He or she also determines which crews should execute

Figure 4-1: Historical distribution of the number of emergencies in a given yard for April.



Note: Since most emergencies are found by monitoring, there are often more emergencies discovered during weekdays when more monitoring crews are working.

these jobs. Shortly after, the maintenance crews are dispatched to their first assignments. Throughout the day, the yard might receive reports of emergency gas leaks that also need to be handled by maintenance crews. These gas leaks are found by dedicated company crews (operated by a department independent from the yards) monitoring leak prone pipes to identify any leaks that need immediate attention. Leaks found that do not pose significant risk to the public are fixed later within regulatory deadlines (usually within 12 months) dictated by the risk involved. These less severe leaks are categorized as standard jobs. Emergencies are highly unpredictable; a given yard can have between zero to six emergencies per day (Figure 4-1). They also are long duration jobs since, for regulatory compliance, the utility requires its crews to dedicate 8 hours (equal to a crew's shift) for attending to emergencies.

In compliance with regulation, the yard needs to dispatch a crew to an emergency within 24 hours of receiving the report of the gas leak. The resource planner typically dispatches an idle crew to an emergency, if they can. However, if all crews are working when an emergency arrives, they continue their work until the first crew finishes. Only then is that crew dispatched to the emergency. Once started, a standard job will not be paused even when an emergency arrives, due to the significant startup effort for the job. Startup activities include travel to the site, digging the street to access the gas pipe, and arranging mandatory police presence at the site.

Resource planners make crew assignment decisions at the beginning of the day. Then they monitor the arrival of emergencies throughout the day. However, once the crew assignment is made, it is usually fixed for the rest of the day. They do not reassign a standard job to another crew once it has been initially assigned to a crew. Yards rarely postpone standard jobs in the case of multiple emergencies. This is because the set of jobs to be done in a given day needs to be known ahead of time in order to arrange for work permits, police detail protection at the work site, crew equipment, and other logistical requirements for performing the job.

Costs of operations at yards. Maintenance crews have eight hour shifts, but can work beyond their shifts if necessary. Any hours worked in excess of the crew's shift is billed as *overtime*, and costs between 1.5 to 2 times as much as the regular hourly wage. Discussions with management reveal that it is preferable for maintenance crews to work overtime to complete standard job assignments, rather than postpone standard jobs and risk incurring any regulatory fines for not meeting deadlines. Based on data from the company's yards, crews in each yard have been working a significant proportion of their hours at overtime. An average crew member works 25% to 40% of his or her hours on overtime. Figure 4-2 shows the actual crew-hours worked in April 2011 for one of the company's average-sized yards (with 25 weekday crews and 4 weekend crews). From Figure 4-2, we observe that even without the randomness introduced by the emergencies, the hours spent working on standard jobs are unevenly divided among the workdays. We observed that one of the major causes of overtime is suboptimal job scheduling and planning for the occurrence of emergencies. Currently, the company has no guidelines or does not use quantitative methods for job scheduling and crew assignment. Instead, resource planners depend on their experience and feedback from supervisors. The company does not currently measure and analyze crew productivity. This results in resource planners relying on subjective input from supervisors on crew assignment decisions. Also, resource planners do not provide slack capacity (i.e., idle crew hours) to attend to any emergencies that might occur later in the day. The variability of emergencies put resource planners in a reactive mode to meet deadlines as well as to handle emergencies, resulting in suboptimal resource allocation.

4.3 Modeling and Problem Decomposition

In this section, we discuss how we developed a stochastic optimization model for multiperiod planning of yard operations. Under a random number of emergencies, the model decides the job schedule (i.e, determining which date each standard job is done) and, at the same time, decides the crew assignment (i.e., once a standard job is scheduled on a date, determining which crew is assigned to complete the job). Later in this section, we discuss a novel decomposition motivated by the practical limitations encountered during the project.

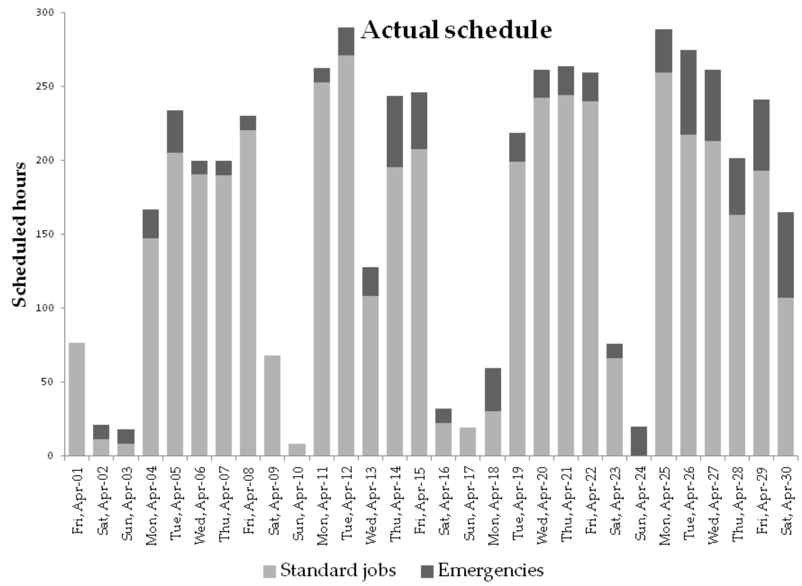
In what follows, we discuss all the assumptions in our model, motivated from the yard operations.

A1. The number of crews available per day is deterministic, although this number can vary daily.

A2. There is no preemption of standard jobs.

A3. Standard jobs have deterministic durations. They do not necessarily have equal durations.

Figure 4-2: Current system: Actual crew-hours worked in April 2011 in an average-sized yard.



A4. The number of emergencies per day is stochastic. Emergencies have equal durations.

A5. Crew assignment does not take distances (geography) into consideration.

A6. The day can be divided into two parts (pre-emergency and post-emergency). In pre-emergency, the standard jobs are assigned to the available crews. Then, the number of emergencies are realized. In post-emergency, these emergencies are assigned to the crews.

Some of these assumptions were imposed for simplicity of the model. One of the requirements of the utility was to have a simple model for various reasons which we discuss later in Section 4.3.1.

Assumption A1 is due to staffing decisions not being part of yard operations, since they are being made by the Resource Management Department based on company-wide projections of work for the year. A2 reflects the actual situation in yard operations due to significant startup effort for standard jobs (see "Daily yard operations" in Section 4.2 for further discussion). A3 is because standard job durations are accurately predicted using factors such as the job type, the size and diameter of the pipe, the age of the pipe, and whether or not the job is on a main street. Based on historical job data, we built a simple regression model which predicts job durations based on job characteristics. We observe minimal variation between predicted values and actual values of job durations. A4's assumption of emergencies having equal durations is due to the utility requiring its crews to devote a fixed amount of time on emergencies. However, the total number of emergencies in each day is stochastic based on the variation seen in yards (Figure 4-1). A5 and A6 are reasonable from a practical point of view since the factors they ignore are second-order in the model. Assumption A5 is made since travel time between jobs is usually much less compared to the duration of jobs. A6 means that we ignore the specific time that an emergency arrives. This is a reasonable assumption since regulation only requires a crew to be dispatched to an emergency within 24 hours (and not immediately as soon as the emergency arrives). Therefore, when crews are already working on standard jobs when an emergency arrives, the emergency does not have to be attended to until a crew finishes its current job. On the other hand, it is possible that a crew might be idling if an emergency arrives after the crew finishes all its standard job assignments. Therefore, the model can incorporate dynamics in a given day and account for specific arrival times of emergencies. In Section 4.5.4, we provide a rolling horizon implementation for a dynamic assignment of standard jobs and emergencies which depends on specific arrival times of emergencies.

Next, we present our model for yard operations. Consider a set of standard jobs that need to be completed within a time horizon (e.g. one month). Each standard job has a known duration and a deadline. Without loss of generality, the deadline is assumed to be before the end of the planning horizon. Within a given day, a random number of emergencies may be reported. Reflecting actual yard operations, the number of emergencies is only realized once the standard job schedule and crew assignments for that day have been made. The following is a summary of the notation used in our model.

T	length of planning horizon
K_t	number of crews available for work on day t, where $t = 1, \ldots, T$
n	total number of known jobs
d_i	duration of job i, where $i = 1, \ldots, n$
τ_i	deadline of job i, with $\tau_i \leq T$, where $i = 1, \ldots, n$
d_L	duration of each emergency
$L(\omega)$	number of emergencies under scenario ω
Ω_t	(finite) set of all scenarios in day t, where $t = 1, \ldots, T$
$P_t(\cdot)$	probability distribution of scenarios on day t, $P_t : \Omega_t \mapsto [0, 1]$

We can estimate the probability distribution of the number of emergencies, which is different for each yard and each month, from historical yard data. For example, Figure 4-1 can be used as the probability distribution for a yard on the month of April.

At the start of the planning horizon, the job schedule has to be decided. At the start of each day, the crew assignments need to be decided before the number of emergencies is known. This is because the calls for emergencies occur later in the day, but the crews must be dispatched early in the morning to their assigned standard jobs before these reports are received. After the number of emergencies is realized, the model decides on an assignment of the emergencies to the crews.

Let the binary decision variable X_{it} take a value of 1 if and only if the job i is scheduled to be done on day t. Let the binary decision variable Y_{itk} take a value of 1 if and only if job i is done on day t by crew k. If scenario ω is realized on day t, let $Z_{tk}(\omega)$ be the second-stage decision variable denoting the number of emergencies assigned to crew k. It depends on the number of standard jobs that have already been assigned to all the crews on day t. The variables $\{X_{it}\}_{it}, \{Y_{itk}\}_{itk}$ are the first-stage decision variables. The variables $\{Z_{tk}(\omega)\}_{tk\omega}$ are the second-stage decision variables.

For each day t, a recourse problem is solved. In particular, given the day t crew assignments, $Y_t \triangleq (Y_{itk})_{ik}$, and the realization of the number of emergencies, $L(\omega)$, the objective of the day t recourse problem is to choose an assignment of emergencies, $Z_t(\omega) \triangleq (Z_{tk}(\omega))_k$, so as to minimize the *maximum number of hours worked over all crews*. Thus, the day t recourse problem is:

$$\begin{aligned} F_t(Y_t, L(\omega)) \ \triangleq \ \underset{Z_t(\omega)}{\text{minimize}} \quad & \max_{k=1,\ldots,K_t} \left\{ d_L Z_{tk}(\omega) + \sum_{i=1}^{n} d_i Y_{itk} \right\} \\ \text{subject to} \quad & \sum_{k=1}^{K_t} Z_{tk}(\omega) = L(\omega) \\ & Z_{tk}(\omega) \in \mathbb{Z}^+, \qquad k = 1, \ldots, K_t, \end{aligned} \tag{4.1}$$

where the term in the brackets of the objective function is the total hours (both standard jobs and emergencies) assigned to crew k. We refer to F_t as the day t recourse function. The constraint of the recourse problem is that all emergencies must be assigned to a crew.

In developing the model, we originally considered an objective of minimizing total expected labor cost (which equivalent to minimizing total expected overtime, since straight hours are a fixed cost). However, due to internal company reasons, they chose not to have a monetary objective in the model. Moreover, in solving the recourse problem with a cost objective, the resulting solution did not correspond to a solution acceptable to the company. In addition, the root cause of the problem that the company has been facing is an uneven distribution of both planned and unplanned work to the yard's crews. In discussions with the company, it has been decided that high overtime labor costs is a *symptom* of this problem, and not the root cause that they wanted to solve. As a result, we chose the objective of minimizing the *maximum work hours* over all the crews. This results in a solution with slightly higher expected total overtime hours, but it is "fair" in that it distributes overtime evenly over the crews.[1]

The objective of the first-stage problem is to minimize the *maximum expected recourse*

[1]To illustrate this, consider a recourse problem with two emergencies (of 8 hour durations each), two standard jobs (of 8 hour durations each), and two maintenance crews. Under an objective of minimizing overtime, an optimal solution is to assign one emergency to the first crew (8 hours), and assign the remaining work to the second crew (24 hours). Under a min-max objective, an optimal solution is to assign one emergency and one standard job to each of the crews (16 hours). Under both solutions, the total overtime is 16 hours. However, the overtime is shared by the two crews under a min-max objective.

function over all days in the planning horizon:

$$
\begin{aligned}
\underset{X,Y}{\text{minimize}} \quad & \max_{t=1,\dots,T} \; E_t\left[F_t\left(Y_t, L(\omega)\right)\right] \\
\text{subject to} \quad & \sum_{t=1}^{\tau_i} X_{it} = 1, \qquad i = 1,\dots,n, \\
& \sum_{k=1}^{K_t} Y_{itk} = X_{it}, \qquad i = 1,\dots,n,\; t = 1,\dots,T, \\
& X_{it} \in \{0,1\}, \qquad i = 1,\dots,n,\; t = 1,\dots,T, \\
& Y_{itk} \in \{0,1\}, \qquad i = 1,\dots,n,\; t = 1,\dots,T,\; k = 1,\dots,K_t,
\end{aligned}
\tag{4.2}
$$

where $F_t(Y_t, L(\omega))$ is described in (4.1). The constraints are: (i) job i must be scheduled before its deadline τ_i, and (ii) if a job is scheduled for a certain day, a crew must be assigned to work on it. The optimization problem (4.2) can be rewritten as a mixed integer program. Section A.4 in Appendix A provides the MIP formulation.

In our model, we assume that standard jobs cannot be postponed if multiple emergencies appear in one day. However, it is possible to explicitly incorporate job postponement using a dynamic model. Note that such models are difficult to solve computationally (see a discussion by Godfrey & Powell (2001) on difficulties of solving multistage problems). In practical applications, the most natural solution strategy is to use a rolling-horizon procedure, solving the static problem at each time period using what is known at that period and a forecast of future events over some horizon. Later in Section 4.4.3, we compare our rolling-horizon procedure to other dynamic models for job scheduling.

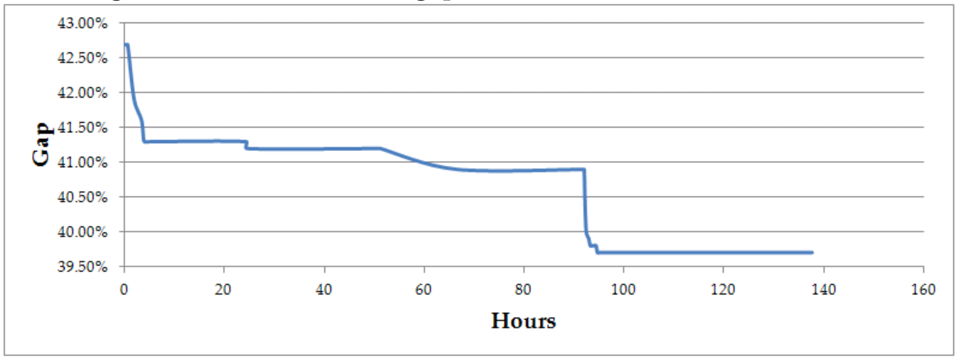
4.3.1 Practical limitations of the joint job scheduling and crew assignment problem

The job scheduling decisions and crew assignment decisions in (4.2) are made jointly. However, there were several practical issues that prevented the implementation of the joint job scheduling and crew assignment problem in yard operations which we discuss below.

Firstly, the full optimization problem is intractable to solve for actual yard problems within a reasonable amount of time. For actual yard settings, crew assignments need to be determined within at most a few minutes. If there are no emergencies, the problem is known to be NP-hard (Pinedo, 2002). Additionally, the presence of a stochastic number of emergencies makes the problem even more computationally intractable when solving the deterministic equivalent problem using commercial off-the-shelf solvers. This is due to the structure of a stochastic MIP (Ahmed, 2010). We demonstrate this by solving the deterministic equivalent problem with actual yard data using Gurobi.[2] Figure 4-3 shows the gap between the current upper and lower bounds on the optimal cost in Gurobi's branch-and-bound search. A gap of zero means that the current solution is optimal. Note that

[2] Actual yard data had 481 standard jobs, 20 crews per weekday, 5 crews per weekend, 0 to 6 emergencies per weekday, 0 to 3 emergencies per weekend.

Figure 4-3: Relative MIP gap in Gurobi's branch and bound.



Note: The gap is the difference between the current upper and lower bounds on the optimal cost in the branch-and-bound procedure. When the gap is zero, the current solution is optimal.

even after 140 hours, Gurobi still is only able to reduce this gap to about 40%.

Secondly, the yard employees and resource planners required transparency in how decisions are being made. There are general-purpose computational methods that solve stochastic optimization problems efficiently such as the integer L-shaped method (Laporte & Louveaux, 1993), scenario decomposition (Carøe & Schultz, 1999), or cutting plane approaches with sequential convexification of the second stage problem (Sherali & Fraticelli, 2002). However, since yard decisions are traditionally being made by resource planners without guidance from any quantitative models or data, resource planners were naturally suspicious of "black box" decision models that do not give insights as to how decisions are being made.

Finally, due to issues about integration with the company's current databases and other strategic issues, the company chose not to invest in a commercial integer programming solver for a implementation of the project throughout the whole company. Therefore, a limitation faced in the project was the fact that our models and heuristics needed to be solved using Excel's Solver or Premium Solver.

These practical issues motivated us to consider a decomposition of the joint problem, into one in which the two decisions (job scheduling and crew assignment) are made sequentially. First is the job scheduling phase, which crudely schedules the jobs on the planning horizon assuming only an *average number of emergencies* on each day. The goal is to meet all the standard job deadlines while evenly distributing work (i.e., the ratio of scheduled work hours to the number of crews) over the planning horizon. Once the schedule of jobs is fixed, then the crew assignment problem can be solved independently for each day. In the crew assignment phase, the standard jobs are assigned to crews under a *stochastic number of emergencies*. The goal is to minimize the expected maximum hours worked by any crew. Note that the two-phase decomposition results in two layers of resource allocation problems. The first layer is a longer term problem where the "resources" are the days that needed to be allocated to the standard jobs. The second layer is a one-day planning problem where the "resources" are the crews that needed to be allocated to the standard jobs and the random emergencies. The decomposition provides transparency in how decisions about job schedules and crew assignments are made. Moreover, the problem is more tractable due to the smaller problem dimensions. Sections 4.4–4.5 provide more details on the two phases of the decomposition.

4.4 Phase I: Job scheduling

In this section, we discuss the job scheduling phase, where standard jobs of varying durations and deadlines have to be scheduled on a planning horizon. We present a deterministic mixed integer program (MIP) whose solution is a feasible job schedule that evenly distributes work over the horizon. We also present a tractable algorithm for producing a job schedule. The algorithm is based on solving the LP-relaxation which, based on actual problem sizes, can be solved using Excel Premium Solver. The schedule resulting from the heuristic is near-optimal in computational experiments and in actual yard problems.

In yard operations, there is a random number of emergencies per day, and the number of crews can change for different days. For instance, yards usually have less crews working during weekends compared to weekdays. Moreover, there are less company crews monitoring gas leaks during weekends, so there are usually less emergencies discovered during weekends. We chose to model the job scheduling phase to schedule standard jobs assuming a deterministic number of emergencies (equal to the average). That is, the standard jobs are scheduled to meet all the deadlines, while balancing (over all the days), the *average* hours scheduled scaled by the number of crews. The job scheduling phase solves the following optimization problem:

$$\begin{aligned}
\underset{X}{\text{minimize}} \quad & \max_{t=1,\dots,T} \left\{ \frac{1}{K_t} \left(d_L E_t[L(\omega)] + \sum_{i=1}^{n} d_i X_{it} \right) \right\} \\
\text{subject to} \quad & \sum_{t=1}^{\tau_i} X_{it} = 1, \qquad i = 1, \dots, n, \\
& X_{it} \in \{0, 1\}, \qquad i = 1, \dots, n, \; t = 1, \dots, T.
\end{aligned}$$

The motivation behind scaling the average scheduled hours per day by the number of crews is so that the optimal solution will schedule less hours on days when there are only a few crews.

Note that the scheduling decisions are made without a detailed description of the uncertainties. Rather, this phase simply takes the expected value of the number of emergencies per day. The stochasticity in emergencies will be handled in the crew assignment phase described in Section 4.5. Due to these modeling assumptions, the problem can be cast as an MIP with only a small number of variables and constraints.

Proposition 4.4.1. *Scheduling phase problem* (4.4) *can be cast as the mixed integer pro-*

gram:

$$
\begin{aligned}
\underset{C,X}{\textit{minimize}} \quad & C \\
\textit{subject to} \quad & d_L E_t[L(\omega)] + \sum_{i=1}^{n} d_i X_{it} \le K_t C, \qquad t = 1, \dots, T, \\
& \sum_{t=1}^{\tau_i} X_{it} = 1, \qquad i = 1, \dots, n, \\
& X_{it} \in \{0, 1\}, \qquad i = 1, \dots, n, \ t = 1, \dots, T.
\end{aligned}
\tag{4.3}
$$

This problem is related to scheduling jobs to *unrelated* machines with the objective of minimizing makespan when there are processing set restrictions (Pinedo, 2002). The makespan is the total length of the schedule when all machines have finished processing the jobs. In our setting, "machines" are equivalent to the dates $\{1, 2, \dots, T\}$. Each job i is restricted to be only scheduled on dates (or "machines") before the deadline, i.e., on "machines" $\{1, 2, \dots, \tau_i\}$. In our setting, the makespan of machine t is the ratio of scheduled hours to number of crews for day t.

Note that even the simpler problem of scheduling jobs on *parallel* machines is well-known to be NP-hard (Pinedo, 2002). List scheduling heuristics (where standard jobs are sorted using some criterion and scheduled on machines one at a time) are commonly used to approximately solve scheduling problems with parallel machines (Kafura & Shen, 1977; Hwang *et al.*, 2004; Glass & Kellerer, 2007; Ou *et al.*, 2008). For the case of unrelated machines, a well-known algorithm by Lenstra *et al.* (1990) performs a binary search procedure, in each iteration solving the linear programming relaxation of an integer program, and then rounds the solution to a feasible schedule. Using a proof based on graph theory, they show that the schedule resulting from their algorithm is guaranteed to have a makespan of no more than twice the optimal makespan.

In what follows, we introduce a heuristic for approximating the solution for the job scheduling problem (4.4). Similar to Lenstra *et al.* (1990), our algorithm is also based on solving the LP-relaxation and rounding to a feasible schedule. However, we do not require initializing the algorithm with a binary search procedure, therefore only solving the LP-relaxation only once. We are able to provide a data-dependent performance guarantee for the heuristic (Theorem 4.4.2) which we derive using a novel technique based on *stochastic analysis.*

4.4.1 LP-based job scheduling heuristic

The details of the job scheduling algorithm, which we call Algorithm LP-schedule, are found in Appendix A. The idea is to solve the linear programming relaxation of the scheduling phase MIP (4.3). The LP solution is rounded into a feasible job schedule by solving a smaller scale MIP.

Consider the LP relaxation of the scheduling phase MIP (4.3) where all constraints of the form $X_{it} \in \{0,1\}$ are replaced by $X_{it} \geq 0$. Denote the solutions to the LP relaxation by C^{LP} and X^{LP}. The algorithm takes the LP solution and converts it into a feasible job schedule using a rounding procedure. The idea in the rounding step is to fix the jobs that have integer solutions, while re-solving the scheduling problem to find schedules for the jobs that have fractional solutions. However, a job i with a fractional solution can now only be scheduled on a date t when the corresponding LP solution is strictly positive, i.e. $X_{it}^{LP} \in (0,1)$. The rounding step solves an MIP, however it only has $O(n+T)$ binary variables, instead of the original scheduling phase integer problem which had $O(nT)$ binary variables (the proof of this is similar to that in Lenstra *et al.* (1990)).

The following theorem states that the schedule resulting from Algorithm LP-schedule is feasible (in that it meets all the deadlines), and its maximum ratio of hours scheduled to number of crews can be bounded.

Theorem 4.4.2. *Let C^{OPT} be the optimal objective cost of the scheduling phase problem* (4.3)*, and let C^{LP} be the optimal cost of its LP relaxation. If X^H is the schedule produced by Algorithm LP-schedule, then X^H is feasible for the scheduling phase problem* (4.3)*, and has an objective cost C^H where*

$$\frac{C^H}{C^{OPT}} \leq 1 + \frac{1}{C^{LP}} \left(\min_{t=1,\ldots,T} K_t \right)^{-1} \sqrt{\frac{1}{2} \left(\sum_{i=1}^{n} d_i^2 \right) (1 + \ln \delta)}, \tag{4.4}$$

where $\delta = \max_{t=1,\ldots,T} \delta_t$ and $\delta_t \triangleq \left\{ r = 1,\ldots,T : X_{ir}^{LP} > 0 \text{ and } X_{it}^{LP} > 0 \right\}$.

The stochastic analysis based proof is in Appendix B. Outline of the proof: Introduce $\tilde{X}$ as the randomized schedule derived by interpreting the LP solution X^{LP} as probabilities. For example, if $X_{i1}^{LP} = X_{i2}^{LP} = 0.5$, then job i is equally likely to be scheduled on day 1 and day 2 in the random schedule. Note that all realizations of $\tilde{X}$ are all the possible roundings of X^{LP}. Moreover, the algorithm produces the rounding X^H with the smallest cost (the maximum ratio of scheduled hours to number of crews). Define B_t as the "bad" event that $\tilde{X}$ has a day t ratio of scheduled hours to number of crews greater than the right-hand side of (4.4). To prove the theorem, we need to show that there is a *positive probability* that none of the bad events $B_1, B_2, \ldots, B_T$ occur. Note that each bad event is mutually dependent on at most δ other bad events. Then if there exists a bound on $\Pr(B_t)$ for all t, we can use Lovász's Local Lemma (Erdős & Lovász, 1975) to prove that the event that none of these "bad" events occur is strictly positive. Since B_t is the event that a function of independent random variables deviates from its mean, $\Pr(B_t)$ can be bounded using the large deviations result McDiarmid's inequality (McDiarmid, 1989).

Let us try to gain some intuition on (4.4). Suppose that there are K crews per day. Note that C^{LP} takes its smallest value when all jobs are due on the last day, with $C^{LP} = \left(\sum_{i=1}^{n} d_i\right)/(KT)$. On the other extreme, if all the deadlines are on the first day, C^{LP} takes its largest value with $C^{LP} = \left(\sum_{i=1}^{n} d_i\right)/K$. Hence, the bound is smaller under more restrictive deadlines. Furthermore, note that δ_t represents (based on the LP solution) the number of days that share a fractional job with day t. With more restrictive deadlines, we would expect δ_t to be smaller, implying that $\delta = \max_t \delta_t$ is smaller. Finally, consider the case where $C^{LP} = \left(\sum_{i=1}^{n} d_i\right)/(\alpha K)$, for some constant $\alpha > 0$ (note that this is the case when deadlines are either all on the first day or all on the last day). Then the bound simplifies to $1 + \alpha \frac{||d||_2}{||d||_1} \sqrt{\frac{1}{2}(1 + \ln \delta)}$, where d is the vector of job durations. Interpreting $\frac{||d||_2}{||d||_1}$ as the coefficient of variation in job durations, we can infer that the bound is smaller if there is less variance in the job duration data.

In both randomly generated job scheduling problem instances as well as actual yard problems, we observe that the data-dependent bound (4.4) is less than 2. But in some cases, the bound might become large, for instance as T increases. However, we can modify the algorithm ensuring that the resulting schedule has a cost of no more than αC^{OPT}, where α is the minimum of 2 and a data-dependent expression (Theorem A.7.1 in Appendix A). Hence, the bound will not explode in asymptotic regimes. The modification is to initialize the algorithm with binary search procedure (described in Section A.6 of Appendix A).

4.4.2 Computational experiments comparing to a sensible resource planner

We implement Algorithm LP-schedule to solve randomly generated problem instances. The size of the problem instance is chosen so that the job scheduling problem solves to optimality within a reasonable amount of time. In these experiments, we randomly generate 100 problem instances and compare the schedule resulting from Algorithm LP-schedule to a schedule that a sensible resource planner might otherwise produce following some rules-of-thumb (we call this second heuristic SRP). SRP's rules have been determined after consulting with several resource planners of the utility we worked with. In SRP, the standard jobs are sorted with increasing deadlines so that the job with the earliest deadline comes first in the list. Then, SRP will determine a cutoff value for work hours per day. Starting from the first day in the horizon, SRP will go through the sorted list of jobs. If the current job has a deadline of today *or* if the current work hours scheduled for today is less than the cutoff, SRP will schedule the current job for today and remove it from the list. Otherwise, it does not schedule it today and moves on to the next day. The cutoff used by SRP is the total average work hours divided by the number of days.

In each problem instance, there are 7 days in the planning horizon, and 3 crews available each day. There are 70 standard jobs to be scheduled (with durations randomly generated between 0 to 8 hours, and deadlines randomly chosen from the 7 days). For each problem instance, we apply both LP-schedule and SRP, noting the cost of both schedules, i.e., maximum ratio of average scheduled hours to number of crews. A schedule is near-optimal if its cost is close to the optimal cost from solving the scheduling problem (4.3). For each problem instance, we compute the percentage difference of the heuristics' cost to the optimal cost. Algorithm LP-schedule has a sample mean (taken over 100 instances) for the percentage difference equal to 3.6%. The 95% confidence interval for this sample mean is

[2.94%, 4.25%]. On the other hand, SRP has a sample mean for the percentage difference equal to 9.7%. The 95% confidence interval for this sample mean is [9.13%, 10.21%].

Additionally, Algorithm LP-schedule manages to improve computational efficiency. Solving for the optimal schedule in (4.4) often requires several hours, which is not viable in actual yard operations. On the other hand, Algorithm LP-schedule only takes a few seconds to solve.

We also implemented the algorithm on actual yard data for one month. During that month, there were 481 standard jobs with durations ranging between 3 hours to 9 hours. On weekdays, there were 20 crews available per day, and the number of emergencies ranged between 0 to 6 per day. On weekends, there were 5 crews available per day, and emergencies ranged between 0 to 3 per day. Due to the size of the problem, the job scheduling problem (4.3) even implemented in Gurobi does not solve to optimality within several days. However, since C^{OPT} is bounded below by C^{LP}, we used the LP relaxation solution to determine that our algorithm results in a schedule that is at most 5.6% different from the optimal job schedule.

4.4.3 Dynamic job scheduling

The Phase I model results in a static job schedule. In the case of the utility we have been working with, yards rarely postpone standard jobs in the case of multiple emergencies (see discussion in Section 4.2). However, one can potentially solve the job scheduling problem with a rolling horizon so that standard jobs can be rescheduled as more information is revealed. In practice, static models are often solved with a rolling horizon rather than solving a dynamic program. This is because dynamic resource allocation models are computationally intractable, with solution methods often only approximating the value function (Godfrey & Powell, 2001; Huh *et al.*, 2013).

We compare our job scheduling model solved using a rolling horizon to the *perfect hindsight* job schedule, which is the optimal job schedule after knowing the sequence of emergencies occurring each day. The cost of the perfect hindsight job schedule is clearly smaller than *any* dynamic job schedule, since it has the benefit of complete information. The cost of the perfect hindsight model is unachievable in reality. However, we use it to evaluate the performance of our model.

We conduct experiments using actual yard data for one month (481 standard jobs, 20 weekday crews, 5 weekend crews). To implement the rolling horizon schedule, the job scheduling problem is re-solved every day, with a horizon starting from the current day until the end of the month. On the current day, the number of emergencies is known (for our experiments, it is randomly drawn from the empirical probability distribution shown in Figure 4-1). The job scheduling model is solved using the known number of emergencies for the current day and an expected number of emergencies for the remaining days. We compare the rolling horizon job schedule to the perfect hindsight schedule where the sequence of emergencies is known.

Figure 4-4: Cost difference between rolling horizon job scheduling model and perfect hindsight model.

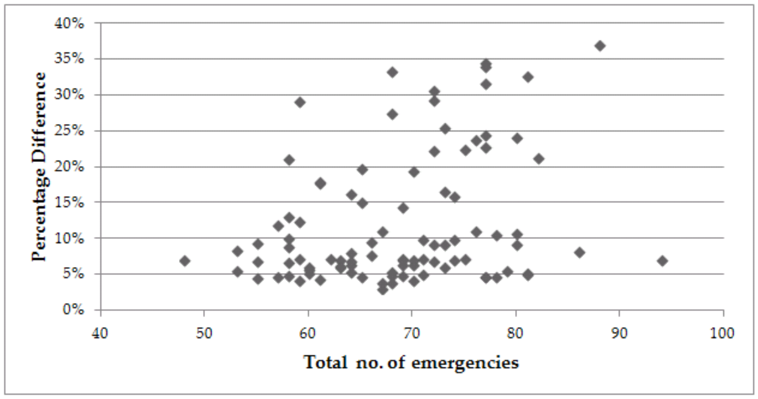


Figure 4-4 demonstrates the percentage cost difference between the rolling horizon schedule and the perfect hindsight schedule in 100 simulation trials. In the experiments, the total number of emergencies for the month varied between 48 to 94. The average cost difference between the rolling horizon schedule and the perfect hindsight schedule is only 12%. Moreover, for 80% of the simulation trials, the cost difference is no more than 20%.

4.5 Phase II: Crew Assignment

In this section, we focus on the second phase of the decomposition, i.e., the crew assignment problem within one day. The basic problem is determining which crews should execute which standard jobs, and which crews to reserve for emergencies, given the stochastic number of emergencies. We develop a heuristic for crew assignment (Algorithm Stoch-LPT) motivated from a property of the optimal crew assignment. Later in this section, we prove that Algorithm Stoch-LPT terminates with an optimal crew assignment for certain special cases. We also modify the algorithm for a multiperiod setting which allows reassignments and evolution of forecasts for the emergencies within the day.

Denote by I the set of standard job indices to be assigned in one day (as determined in the job scheduling phase). Let L be the stochastic number of emergencies on that day. Suppose there are K crews available on that day. The crew assignment problem assigns all standard jobs in set I to the available crews. However, these assignments must be made before the number of emergencies for the day is realized. After the number of emergencies is known, all emergencies must be assigned to the crews. The objective is to minimize the expected maximum hours worked on that day.

The crew assignment problem for one day solves a *two-stage stochastic mixed integer* *program.* The first stage problem is:

$$
\begin{aligned}
\underset{Y}{\text{minimize}} \quad & E\left[F\left(Y, L(\omega)\right)\right] \\
\text{subject to} \quad & \sum_{k=1}^{K} Y_{ik} = 1, \qquad i \in I, \\
& Y_{ik} \in \{0,1\}, \qquad i \in I,\ k = 1,\ldots,K,
\end{aligned}
\tag{4.5}
$$

where $F(Y, L(\omega))$ is defined as:

$$
\begin{aligned}
F(Y, L(\omega)) \triangleq \underset{Z}{\text{minimize}} \quad & \max_{k=1,\ldots,K} \left\{ d_L Z_k + \sum_{i\in I} d_i Y_{ik} \right\} \\
\text{subject to} \quad & \sum_{k=1}^{K} Z_k = L(\omega) \\
& Z_k \in \mathbb{Z}^+, \qquad k = 1,\ldots,K.
\end{aligned}
\tag{4.6}
$$

Note that the term in the brackets of the objective function is the number of hours assigned to crew k during scenario ω and under the standard job assignments Y. The assignment phase problem can also be rewritten as a mixed integer program (see Section A.8 of Appendix A).

A limitation encountered in our project was that resource planners, who traditionally made daily yard operations decisions without the use of models, were resistant of "black box" decision models which did not give insight into how crew assignment decisions are being made (see Section 4.3.1). Therefore, problem (4.5) cannot be solved using IP solvers or computational techniques aimed for solving general stochastic optimization problems. Motivated by this, we developed a crew assignment algorithm which exploits the specific structure of the crew assignment problem, which we will introduce later in Section 4.5.2. This algorithm is simple and intuitive since it can viewed as a stochastic variant of the Longest Processing Time First (LPT) rule. The algorithm we developed also resulted in natural guidelines for resource planners to follow in making yard operations under a stochastic number of emergencies.

4.5.1 Stochastic model compared to using averages

The stochasticity of the number of emergencies makes (4.5) makes computationally intractable. In what follows, we compare the two-stage stochastic model (4.5) to a natural heuristic which ignores stochasticity. In particular, this heuristic solves a deterministic model assuming that *the number of emergencies is equal to the expectation* $E[L(\omega)]$. We now demonstrate that solving (4.5) results in more robust assignments than the deterministic model ignoring stochasticity.

We refer to the deterministic heuristic as AVG, and the two-stage stochastic model as OPT. In the following example, we compare the cost (i.e., maximum work hours in each emergency scenario) under AVG and OPT. Suppose there are 7 crews available, and 15

Table 4.2: Maximum work hours in different scenarios under optimal assignment and assignment based on average number of leaks.

Scenario	Probability	OPT max hours	AVG max hours
0 leaks	0.4	11.76	10.66
1 leak	0.2	11.77	10.66
2 leaks	0.4	11.83	18.28
Expected maximum hours		11.79	13.70

[a] OPT is the optimal solution to (4.5). AVG optimizes assuming an average number of leaks.

standard jobs need to be assigned. Standard job durations are between 1 hour and 7 hours. The emergency gas leak job duration is 8 hours. The probability of 0 leaks is 40%, the probability of 1 leak is 20%, and the probability of 2 leaks is 40%. The average number of leaks is 1.

Table 4.2 summarizes the work hours with assignments from AVG and OPT. Based on the table, if there are no leaks, all crews work less than 11.76 hours under OPT, whereas all crews work less than 10.66 hours under AVG. Note that, regardless of the number of leaks, all the crews work less than 11.83 under OPT. But under AVG, at least one crew is working 18.28 hours if there are 2 leaks. Hence, with 40% probability, a crew under the AVG assignment works 18.28 hours. Since OPT results in a crew assignment where all crews work less than 11.83 hours on any leak scenario, it is more robust to stochasticity of gas leaks. These results agree with Birge (1997) who demonstrated that in many real-world applications stochastic optimization models are superior to their deterministic counterparts.

4.5.2 Crew assignment heuristic

We conducted computational experiments on several examples in order to gain insight into the structure of the optimal crew assignment solution to (4.5). The appendix (Section A.9) explains in detail the experiments we conducted. An observation we make from the experiments is that in the optimal solution, if a crew is assigned to work on an emergency in a given scenario, that crew should also be assigned to work on an emergency in a scenario with more emergencies. This is formalized in the following proposition.

Proposition 4.5.1. *There exists an optimal solution* $(Y^*, Z^*(\omega), \omega \in \Omega)$ *to the stochastic assignment problem* (4.5) *with the property that if* $L(\omega_1) < L(\omega_2)$ *for some* $\omega_1, \omega_2 \in \Omega$, *then* $Z^*_k(\omega_1) \leq Z^*_k(\omega_2)$ *for all* $k = 1, \ldots, K$.

This proposition motivates our heuristic for crew assignment under stochastic emergencies. The heuristic aims to mimics the property of the optimal crew assignment in Proposition 4.5.1. We refer to the heuristic as Algorithm Stoch-LPT, since it is is a variant of the Longest-Processing-Time First (LPT) algorithm under a stochastic number of emergencies. Recall that LPT applies when there are no emergencies, and the objective is to minimize the maximum work hours of the crews. In each iteration of LPT, it keeps track of the current number of assigned work hours (current load) for each crew. LPT initializes the current load for each crew to be zero. Then it sorts the standard jobs in decreasing duration. Starting with the longest duration job, each iteration of LPT assigns the current standard job to the crew with the smallest current load, updating the current load after an assignment is made.

In what follows, we describe the proposed Algorithm Stoch-LPT for stochastic emergencies, with the objective of minimizing the *expected* maximum work hours of the crews. The algorithm begins by first making assignments of emergencies in each scenario. For example, in a scenario with two emergencies, the algorithm needs to assign two emergencies to the crews. For each scenario, the algorithm assigns emergencies, starting with the scenario with the least emergencies, then the one with the second least, continuing until it assigns all emergencies under all scenarios. For the current scenario, Algorithm Stoch-LPT uses a procedure for assigning the emergencies that preserves the monotonicity property described in Proposition 4.5.1. It assigns the emergencies in the current scenario to the crews by the LPT rule. But in case of ties (where more than one crew has the smallest current load), it chooses a crew whose current load is strictly smaller than its load in the previous scenario's assignment.

After emergencies have been assigned for all scenarios, the next step in Stoch-LPT is to assign the standard jobs. The algorithm keeps track of the current load of each crew *in each scenario*, which is initialized after the crews' emergency job assignments. Then, like LPT, the algorithm sorts the standard jobs in decreasing order of duration. Starting with the longest duration job, each iteration of Stoch-LPT assigns the current standard job to a crew according to the following rule. Under each crew, determine the increase in expected maximum load that results from assigning the current job to that crew. Note that different assignments result in different loads for each crew in each scenario. The expected maximum load is computed by summing over all scenarios the maximum load in the scenario multiplied by the probability of the scenario. The standard job is assigned to the crew that has the smallest amount of increase in the expected maximum load. If there are any ties, the standard job is assigned to the crew with the smallest expected current load. After a standard job is assigned, the current load of each crew in each scenario is updated.

We conduct computational experiments comparing the crew assignment produced by Algorithm Stoch-LPT to the optimal crew assignment. Table 4.3 compares expected maximum hours worked under both crew assignments. Each experiment uses the same set of 15 standard jobs and 7 crews, but a different probability distribution for the number of emergency gas leaks (see Table D.23–D.24 in Appendix D for the data). Note that under all probability distributions, Algorithm Stoch-LPT results in expected maximum hours no more than 8.25% of the optimal.

Finally, we would like to discuss another implication of Proposition 4.5.1. In reality, the leak scenario reveals itself over time since leaks are discovered throughout the day. However,

Table 4.3: Expected maximum hours worked under the optimal crew assignment and the assignment from Algorithm Stoch-LPT.

	E[no. leaks]	Stdev[no. leaks]	Expected maximum hours		% Difference
			Optimal	Algorithm Stoch-LPT	
Leak distribution 1	1.0	0	10.66	11.50	7.96%
Leak distribution 2	1.0	0.45	11.41	12.06	5.72%
Leak distribution 3	1.0	0.63	11.78	12.75	8.22%
Leak distribution 4	1.0	0.89	11.79	12.67	7.50%
Leak distribution 5	1.0	0.89	12.18	12.19	0.11%
Leak distribution 6	1.0	1	12.50	12.95	3.62%
Leak distribution 7	1.0	1.18	12.85	13.34	3.79%

as the proposition states, if a crew is assigned to an emergency for a scenario with one leak, then this same crew is assigned at least one emergency for scenarios with two, three, and more leaks. Therefore, the first leak that appears is always assigned to that crew, under any scenario. This way, one can "rank" the crews that handle the emergencies. Thus, the crew assignment solution of the static model can be easily implemented in a real-time setting where leaks arrive throughout the day. This ranking of crews based on the optimal leak assignment is formalized in the following proposition.

Proposition 4.5.2. *Suppose* $(Y, Z(\omega), \omega \in \Omega)$ *is a feasible solution to the stochastic assignment problem* (4.5) *such that, if* $L(\omega_1) < L(\omega_2)$ *for some* $\omega_1, \omega_2 \in \Omega$*, then* $Z_k(\omega_1) \leq Z_k(\omega_2)$ *for all* $k = 1, \ldots, K$*. The crews can be relabeled as* $k_1, k_2, \ldots, k_K$ *so that* $Z_{k_{j-1}}(\omega) \geq Z_{k_j}(\omega)$ *for all* $\omega \in \Omega$*.*

Corollary 4.5.3. *There exists an optimal solution* $(Y^*, Z^*(\omega), \omega \in \Omega)$ *to the stochastic assignment problem* (4.5) *where the crews can be relabeled as* $k_1, k_2, \ldots, k_K$ *so that* $Z^*_{k_{j-1}}(\omega) \geq Z^*_{k_j}(\omega)$ *for all* $\omega \in \Omega$ *and* $\sum_{i \in I} d_i Y^*_{i,k_{j-1}} \leq \sum_{i \in I} d_i Y^*_{i,k_j}$*.*

4.5.3 Special case: Two crews and two emergency scenarios

In what follows, we derive some results about how our heuristic performs relative to the optimal crew assignment solution when there are only two crews and two scenarios (either no leak, or one leak). For this section, we assume that the probability of no leak is p, and the probability of one leak is $1 - p$. The duration of an emergency is d_L. Let x and y be the number of standard job hours that are assigned to crew A and crew B, respectively. Without loss of generality, we assume that leaks are assigned to crew A. Then it is easy to verify that the expected makespan is given by $p \max(x, y) + (1 - p) \max(d_L + x, y)$.

Proposition 4.5.4. *In the optimal crew assignment, the standard jobs assigned to crew A has a shorter total duration than the standard jobs assigned to crew B.*

Proof. We prove this by contradiction. Consider an assignment where we swap the standard job assignments of the crews. Then in both leak scenarios, the maximum work hours (makespan) is no greater under this new assignment. This implies that the new assignment has an expected makespan that is less than or equal to that of the original assignment. □

Proposition 4.5.5. *If $p \leq \frac{1}{2}$, then Algorithm Stoch-LPT will begin by assigning the longest duration standard job. It will then assign subsequent jobs to crew B until the duration of the emergency is shorter than the cumulative duration of the assigned jobs.*

Proof. Let D be the cumulative duration of assigned standard jobs by Algorithm Stoch-LPT, where $D \leq d_L$. We will show that it is optimal to assign all D hours to crew B. Let $x + y = D$, and consider two cases. First, note that if $x \leq y$, then the expected makespan is $py + (1-p)(d_L + x) = (2p-1)y + (1-p)(D + d_L)$. Therefore, since $p \leq \frac{1}{2}$, it is optimal to assign the most amount of standard job hours to crew B. For the second case, if $x > y$, then the expected makespan is $px + (1-p)(d_L + x) = D - y + (1-p)d_L$. Therefore, even in this case, it is optimal to assign the most amount of standard job hours to crew B. □

The next proposition considers the special case of standard jobs with equal durations. The proposition states that, under this special case, Algorithm Stoch-LPT terminates with an optimal crew assignment.

Proposition 4.5.6. *Suppose all standard jobs have equal duration d. Then the following statements are true for the crew assignment resulting from Algorithm Stoch-LPT:*

1. *If $p \leq \frac{1}{2}$, then Stoch-LPT will assign the first $\lfloor \frac{d_L}{d} \rfloor$ longest duration standard jobs to crew B. The algorithm will assign the subsequent jobs alternately between crew A and crew B.*
2. *If $p > \frac{1}{2}$, then Stoch-LPT will assign the longest duration job to crew B. The algorithm will assign the subsequent jobs alternately between crew A and crew B.*
3. *For any value of $p \in (0, 1)$, Stoch-LPT terminates with an optimal crew assignment.*

4.5.4 Dynamic crew reassignment

Motivated by yard operations where assignment of standard jobs is determined once in the beginning of the day and cannot be changed later, our model for the crew assignment is static. In what follows, we demonstrate a modification of Algorithm Stoch-LPT where standard jobs that have not yet been started can be re-assigned later in the day, as more information about the emergencies becomes available.

We assume that the standard jobs can be reassigned every hour. We also assume that the arrival of emergencies follow a Poisson process with an arrival rate λ. Any arrival process can be used, however we chose a Poisson process for the purpose of illustration. Recall that the emergencies are found by dedicated company crews that monitor for gas leaks in a shift of 8 hours. Then there is a natural update rule for the belief on the number of emergencies. Suppose there are s hours remaining until company crews stop monitoring for leaks. Then the probability that n emergencies are found within s hours is $P(L=n) = \frac{(\lambda s)^n}{n!} e^{-\lambda s}$, for $n = 0, 1, 2, \ldots$

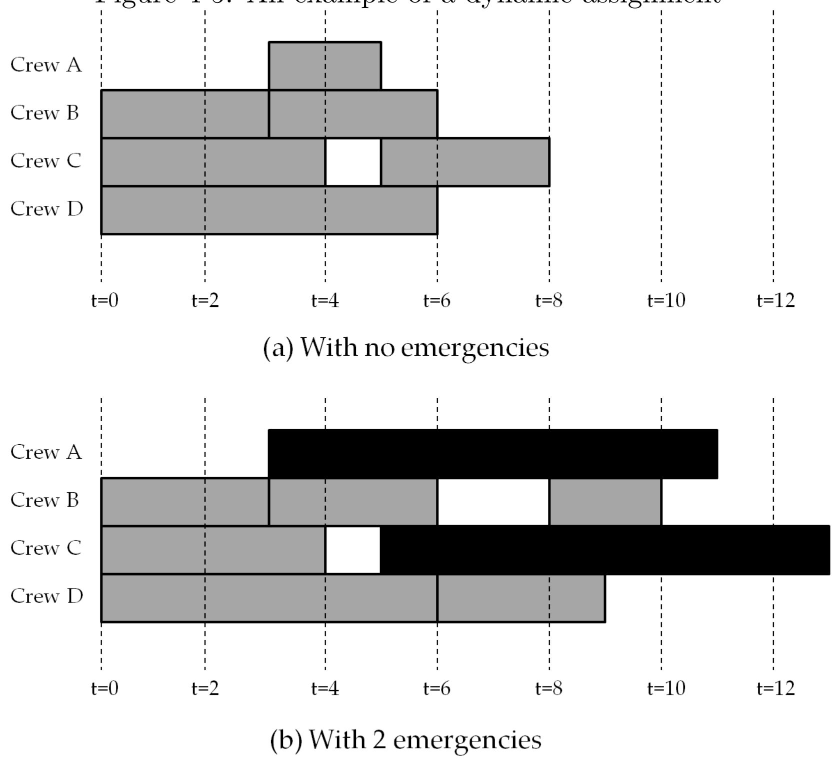
What we refer to next as the dynamic crew reassignment is the following. At the start of the day, determine the standard job and emergency assignments according to Algorithm Stoch-LPT. We make a distinction between Stoch-LPT's assignments and the *actual* assignments that are based on the realization of emergencies. In the first hour, the number of emergencies are realized according to the Poisson distribution. Make the actual emergency assignment based on Algorithm Stoch-LPT. For example, if there is one emergency, assign that emergency based on the algorithm's emergency assignment under the scenario with one emergency. If a crew has an actual emergency assignment, it starts work on that emergency in the current hour. Otherwise, choose an actual standard job assignment from the set of standard jobs that Stoch-LPT assigns to that crew. We choose the longest duration job in the set.[3] The crew starts work on the chosen standard job (if any) in the current hour. Moving to the next hour, we again apply Algorithm Stoch-LPT, but (i) with only the standard jobs not yet started, (ii) with the current load of some crews reflecting the job they started from the previous hour, and (iii) with an updated probability distribution of the number of emergencies. Emergencies are realized for that hour according to the Poisson distribution, and actual assignments of emergencies and standard jobs are determined as before. Then, continue the re-assignment at the beginning of each hour until either there are no more standard jobs left or the end of 8 hours is reached. At the end of the last hour, the LPT rule is applied for the remaining standard jobs (if any).

Figure 4-5 illustrates through an example how the crew assignment evolves as the emergencies arrive. The example uses four crews, six jobs, and an emergency arrival rate of 0.2 per hour. Grey rectangles represent standard jobs. Black rectangles represent emergency jobs. Figure 4-5(a) shows the job assignment when there are no emergencies. Figure 4-5(b) shows the job assignment when two emergencies arrive at $t = 3$ and $t = 5$. White space between jobs shows that the crew is idle during that period. Note that, depending on the arrival of emergencies, the standard jobs assignments are different.

We compare the dynamic reassignment solution to the static solution using simulation experiments. Consider a yard with 17 crews and 21 standard jobs with durations varying from 3 to 9 hours. An emergency has a duration of 8 hours. Emergency arrivals follow a Poisson process with rate 0.352 per hour. Emergencies can only arrive in an 8 hour period, during which there is an expected number of 2.8 emergencies. We simulate 100 sequences of emergency arrivals. For each sequence, we apply both the static Stoch-LPT and the dynamic Stoch-LPT. Figure 4-6 shows the number of overtime hours saved by the dynamic reassignment plotted against the number of emergencies in the sequence. The static assignment (which is made based on an expected number of 2.8 emergencies) is conservative, and reassignment can adjust this conservative solution as more information

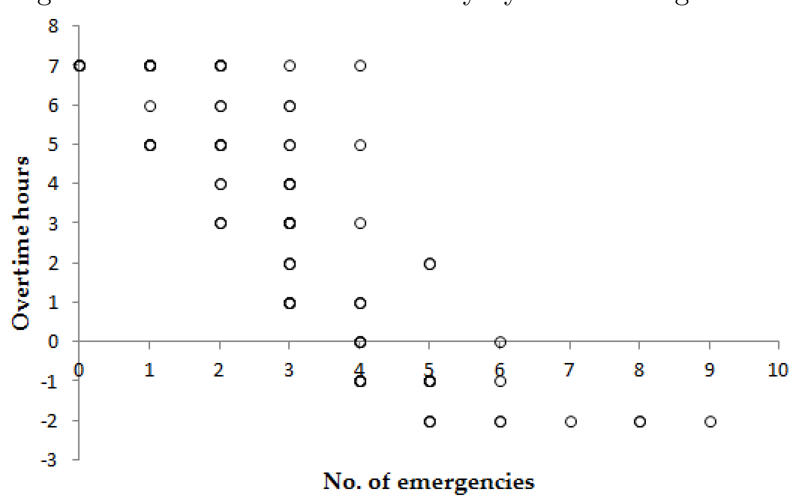
[3]It is possible to have a different rule for choosing the actual standard job assignments. For instance, one can choose the shortest duration job in the set.

Figure 4-5: An example of a dynamic assignment



The horizontal axis represents time. A rectangle represents a job; its duration is proportional to the length of the rectangle. Grey rectangles are standard jobs. Black rectangles are emergency jobs.

Figure 4-6: Overtime hours saved by dynamic reassignment



Each data point corresponds to a different sequence of emergency arrivals.

is revealed. Note that the benefits of dynamic reassignment decreases if there are more emergency arrivals during the 8 hour period.

4.6 Business analytics for a utility's Gas business

In this section, we describe how the research above applies to the scheduling of operations at the Gas business of a large multi-state utility. This is based on a joint project between the research team and the company that gave rise to the results of this chapter.

The company maintains a network of gas pipeline. It keeps a roster of maintenance crews who have two types of tasks: to execute standard jobs by their deadlines, and to respond to emergencies. We discuss how we used the optimization models and heuristics described in this chapter so that the company could develop better strategies to create flexibility in its resources to handle emergencies.

4.6.1 Overview of the project

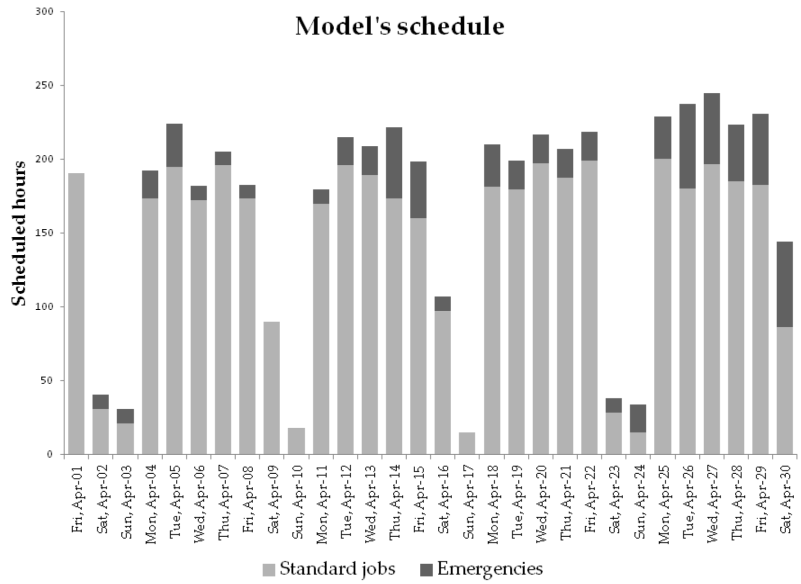
At the onset of the project, our team analyzed sources of inefficiency in yard operations by mapping in detail the existing yard processes. We visited several company yards and interviewed a number of resource planners, supervisors and crew leaders, as well as members of the Resource Management Department. We also did extensive job shadowing of crews from multiple yards performing different types of jobs, and documented the range of processes followed. We also constructed historical job schedules based on data gathered from the company's job database (see Figure 4-2 for a schedule of a yard's one month operations).

Our project with the utility company had three main objectives. The first was to develop a tool that can be used with ease in the company's daily resource allocation. Based on the models and heuristics we discuss in this chapter, we created a tool – the Resource Allocation and Planning Tool (RAPT) – to optimally schedule jobs and to assign them to crews while providing flexibility for sudden arrival of emergencies. RAPT has access to the job database and the time-sheet database, and uses this information to estimate leak distributions and job durations. The resource planner can view a webpage showing RAPT's output of the weekly schedule for each crew and detailed plans under different gas leak scenarios. This tool is being piloted in one of the company's yards.

The second objective was to create and improve processes related to daily resource allocation so that the tool could be easily embedded into daily scheduling process. We observed that a lot of the data in the database was either missing, inappropriately gathered or not vetted before entry into the system. Having missing or inaccurate data makes it very difficult to apply a data-driven tool such as RAPT and makes it even more difficult to address the right issues. Processes were created to ensure that when new jobs were added to the database, they had the right database fields set in a consistent manner across all jobs and yards.

The third objective was to analyze the impact of key process and management drivers on operating costs and the ability to meet deadlines using the optimization model we developed.

Figure 4-7: Hypothetical scenario: Crew-hours worked if optimization model is used to schedule jobs.



Results from this analysis will help the company deploy the optimization model with all the necessary process and management changes in order to capture the potential benefits outlined in this chapter. The key process and management drivers selected for the study are work queues of available jobs for scheduling, availability of detailed productivity data (down to crew level) and supervisor presence in the field. These are discussed in Section 4.6.2.

Finally, we set out to determine the potential impact of the RAPT tool to the company. Recall Figure 4-2 which shows the actual one-month profile of work hours in an average-sized yard. Figure 4-7 shows the profile for the same set of jobs if RAPT is used to schedule jobs and assign them to crews. The result is a 55% decrease in overtime crew-hours for the month. Clearly, the schedule and crew assignments produced by RAPT is superior to those produced previously by the resource planner. However, even if compared to the best possible schedule where uncertainty is removed, the decisions produced by RAPT compare favorably. The "perfect hindsight" scheduling and assignment decisions are based on complete knowledge of the realizations of emergencies that occur in the month. The "perfect hindsight" model results in the maximum possible reduction in overtime since the yard can plan completely for emergencies. However, even though the RAPT model assumes a random number of emergencies, it still is able to capture 98.6% of the maximum possible overtime reduction by "perfect hindsight".

4.6.2 Using the model to recommend changes

Using our models from this chapter, we conducted a study to understand the impact of changes in yard operations on productivity. Based on past studies the company had conducted, the company understands that yard productivity is a complex phenomenon driven by process settings such as the size of work queues (i.e., jobs available for scheduling), effective supervision, incentives, and cultural factors. The research team and the company agreed to analyze three specific drivers of productivity using RAPT: work queue level, use of crew-specific productivity data, and the degree of field supervision.

Optimal work queue level.

Jobs need to be in a "workable" state before crews can begin to execute them. For example, the company needs to apply for a permit with the city for the job. Jobs in a "workable jobs queue" are jobs ready to be scheduled by RAPT. A queue is maintained since "workable" jobs are subject to expiration and require maintenance to remain in a workable state (e.g., permits need to be kept up-to-date). We observed some yards kept a low level of jobs in the workable jobs queue. The low workable jobs queue adversely impacted the RAPT output by not fully utilizing the tool's potential. The team decided to run simulations to determine the strategic target level for the workable jobs queue to maximize the impact of RAPT while minimizing the efforts to sustain the workable jobs queue level.

For our simulations we used actual data from one of the company's yards. Table D.39 in Appendix D provides the data used in the experiment. Five crews, each with 8 hour shifts, are available to work in each simulated day. There are ten different job types to be done. On each day, the Resource Management Department announces a minimum quota of the number of jobs required to be done for each type. These quotas are random and depend on various factors beyond the yard's control. Based on historical quotas, we estimate the probability distribution of daily quotas for each job type. The yard maintains a workable jobs queue for each job type. Suppose today the quota for CMP jobs is 10, however there are only 6 jobs in the workable CMP job queue. Then, today, the yard will execute 6 CMP jobs, and will carry over the remaining 4 CMP jobs as a backlog for the next day.

The yard uses a continuous review policy for the workable jobs queue specified by a *reorder point* and an *order quantity.* Each time the total workable jobs (both in the queue and in the pipeline) drops below the reorder point, the yard requests new workable jobs. The size of the request is equal to the order quantity. The request is added to the pipeline and arrives after a lead time of 3 days. For instance, this lead time may include time used for administrative work to apply for a permit. Suppose the yard chooses a reorder point of 2 and an order quantity of 10 for the CMP workable jobs queue. Then, each time the total CMP workable jobs drops below 2, the yard places an additional request for 10 workable CMP jobs. In our simulations, the order quantity is set for each job type queue so that, on average, new requests are made every week. The reorder point is determined from a service level the yard chooses, where the service level is the probability that there is enough jobs in the workable jobs queue to meet new quotas during the lead time period (i.e., while waiting for new workable jobs to arrive).

For each simulated day, quotas are randomly generated and met to the maximum extent possible from the workable jobs queue. The jobs are assigned to the 5 crews using the RAPT

Figure 4-8: Workable jobs queue over one simulated month with 50% service level.

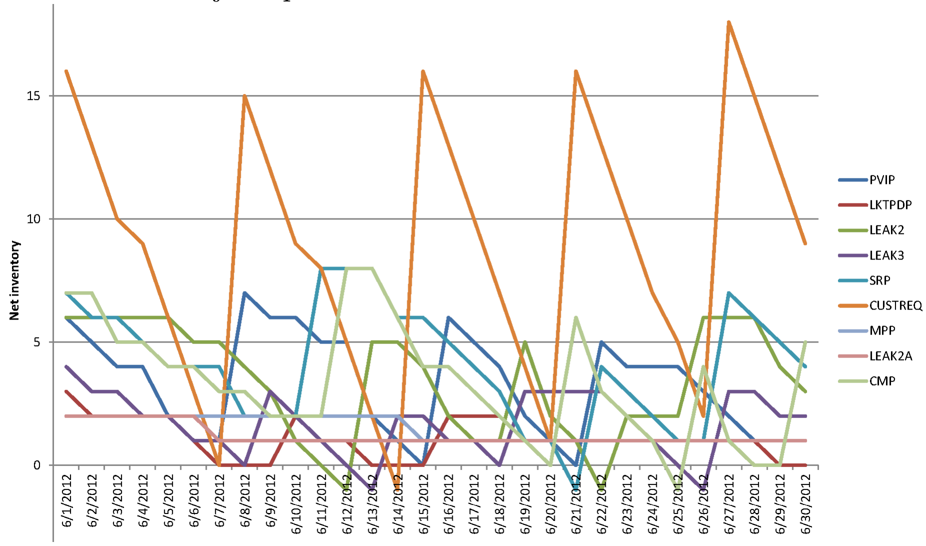


Table 4.4: Effect of service levels on average workable jobs inventory, backlogged jobs, and overtime crew-hours for one simulated month.

Service Level	50%	75%	90%	99%
Average inventory per day	28.8	35.6	37.6	50.6
Total backlogged jobs	7	0	0	0

crew assignment model. Figure 4-8 shows the evolution of the workable jobs queue in one simulated month for a 50% service level. The net inventory level corresponds to the total number of workable jobs currently in the queue. When the net inventory is negative, then there is a backlog of workable jobs for that job type (i.e., there are not enough workable jobs to meet the quotas). Table 4.4 summarizes the results of the simulation for different service levels. Note that increasing the service level increases the average size of the workable jobs queues. With 50% service level, the average inventory per day in the workable jobs queue is 28.8. However, a total of 7 quotas have not been met in time. Increasing the service level to 75% requires increasing the average inventory per day to 35.6, resulting in eliminating any backlogged jobs. Increasing the service level further to 90% or 99% results in higher average inventories of workable jobs, but with essentially the same effect on backlogged jobs as the smaller service level 75%.

Using crew productivity data.

Presently, detailed crew productivity is not available in the company's database. As such, it is not possible to make crew assignments to take advantage of the inherent job-specific productivity differences between crews in the crew assignment phase. We used our model to aid company management in understanding the impact of using job-specific productivity data in crew assignment versus assigning jobs based on average productivity. The hope

Table 4.5: Total expected overtime crew-hours for different expertise factors.

	Base case	$\gamma = 5\%$	$\gamma = 10\%$
Total expected overtime crew-hours	340.4	329.4	302.6
% Improvement over base case	—	3.23%	11.1%

is this work can motivate upper management to provide the appropriate resources to keep track of crew productivity.

In the simulation, we assume that each of the five crews are "experts" in one of the job types. We let $\gamma \in [0, 1)$ be an *expertise factor* which is the percentage reduction in a job's duration if an expert works on it. Larger values of γ mean that experts are more productive relative to regular crews. In our experiments, we let $\gamma = 0\%$ (base case), 5%, and 10%. We run the simulation for 30 days. In each day, the work that has to be assigned is randomly generated (the distribution of quotas is given in Table D.39 of Appendix D). We observe that the assignment model assigns most jobs to crews that have expertise in them. Table 4.5 shows the total expected overtime crew-hours over a one month period. By having expert crews who work with 5% reduced durations, overall overtime hours can decrease by 3.23%. The decrease in overtime hours is nonlinear, since if expert crews can work with 10% reduced durations, the total overtime hours in one month are reduced by as much as 11.1%.

Increasing supervision over crews.

A prior study conducted by the company observed that crew productivity is directly related to field supervision. More time spent overseeing crews in the field results in more productive crews. The team used RAPT to validate and measure the appropriate level of supervision to maximize productivity since field supervision has a cost.

In these simulations, we compare the effect of having an increased supervisor presence in the field to the average expected overtime incurred by crews. Consider the work types given in Table D.39 of Appendix D. Assume that by having increased supervisor presence, the durations of work types can be decreased. We will compare different cases: the base case (no reduction), 5% reduction, 10% reduction and 25% reduction. Let there be 5 crews, and the daily quotas are randomly generated based on Table D.39. Unlike the previous simulations, we assume that there is an infinite supply of permitted work (so the inventory policy is not a factor). Each day, we assign the work to the 5 crews using RAPT and note the total expected overtime incurred by the five crews during that day. For the different cases, we run this simulation for 30 days and calculate the total expected overtime averaged over 30 days.

Table 4.6 reports the result of the simulation for the different cases. We can infer that each 5% decrease in job durations (by increasing supervisor presence) results in a reduction of 1.6 overtime crew-hours each day for the five-crew yard. Therefore, assuming that there are 3 members in a crew, a 5% increase in productivity results in reducing a total of 143 overtime hours charged for the yard in one month.

Projected financial impact from changes.

In our project, we used our models to illustrate projected financial impact of implementing process changes in the utility. We illustrate this with a hypothetical utility that has an operating profit of $3.5 billion per year. The hypothetical utility employs 10,000 field personnel. The straight-time hours per person per year are 2,000, with an additional 500 overtime hours per person per year. The average wage of a field personnel is $50 per hour. Overtime is paid out at $75 per hour. The hypothetical utility spends $1 billion in straight-time labor costs (20 million hours), with an additional $375 million in overtime labor costs (5 million hours). Table 4.7 summarizes the projected financial impact to this hypothetical utility of introducing the business process changes described earlier in this section. The percentage savings in overtime costs are all based on the analyses in Sections 4.6.2–4.6.2.

If the utility were to keep crew-specific productivity data as described in Section 4.6.2, we would anticipate annual savings of about $12 million, which represents 0.3% of the utility's annual operating profit. Suppose the company were to increase crew supervision as described in Section 4.6.2. Based on previous company studies, increased supervisor presence reduces job durations by at least 10%. This results in annual savings of $74 million (or 2% of the annual operating profit). If the company is able to implement both changes, this has a cumulative savings of about $84 million per year which represents 2.4% of the annual operating profit.

4.7 Conclusions

In many industries, a common problem is how to allocate a limited set of resources to perform a specific set of tasks or jobs. However, sometimes these resources are also used to perform emergencies that randomly arrive in the future. For example, in hospitals, operating rooms are used both for elective surgeries (that are known in advance) and emergency surgeries (which need to be performed soon after they arrive). Another example which in fact motivated this chapter is scheduling crews in a gas utility company. Maintenance crews have to perform both standard jobs (pipeline construction, pipe replacement, customer service) as well as gas leak repair jobs. The second type of jobs arrive randomly throughout the day. With randomly arriving emergencies, the problem becomes more complicated since the resources need to be allocated before realizing the number of emergencies that have to be performed. Thus, a schedule needs to be flexible in that there must be resources available to perform these future emergencies.

We use stochastic optimization to model the problem faced by the gas utility. The problem is decomposed into two phases: a job scheduling phase and a crew assignment phase. The optimization problems resulting from each phase are computationally intractable, but

Table 4.6: Simulation results for increasing supervisor presence in the field.

	Base case	5% reduction in job durations	10% reduction in job durations	25% reduction in job durations
Average overtime per day per crew (crew-hours)	3.09	2.81	2.48	1.51
% Improvement over base case	—	9.2%	19.7%	51.1%

Table 4.7: Projected financial impact of business process changes in a hypothetical utility with a $3.5 billion annual operating profit.

	Base case	Have expert crews	Increase supervisor presence	Have expert crews and increase supervisor presence
Annual overtime hours	5 million	4.84 million	4.02 million	3.89 million
Annual overtime labor cost	$375 million	$363 million	$301 million	$291 million
% Savings in overtime labor cost	—	3.23%	19.7%	22.3%
Savings in overtime labor cost	—	$12 million	$74 million	$84 million

[a] Assumptions: (i) expert crews are 5% more productive, (ii) increased supervisor presence results in a 10% increase in productivity for field personnel.

we provide tractable heuristics for solving each of them. The job scheduling phase heuristic solves a mixed integer program, for which we propose an LP-based heuristic. We are able to prove a data-driven performance guarantee for this heuristic. The crew assignment phase solves a two-stage stochastic mixed integer program. Here, we propose an algorithm which replicates the structure of the optimal crew assignment. We demonstrate how the two heuristics can be implemented in a rolling horizon for rescheduling and reassignment in response to the state of emergencies.

We used our models and algorithms to improve job scheduling and crew assignment in the Gas business of a large multi-state utility company which faced significant uncertainty in its daily operations. Our models were also used to help the utility make strategic decisions about changes in its business and operations. In simulations using actual data and our models, we project the impact of different process changes to crew utilization and overtime labor costs.

4.7.1 Future Directions

There are several future directions that go beyond the scope of this chapter that could be pursued.

In this chapter, we focused on the job scheduling and crew assignment problems assuming that there is no travel time between two jobs. In the real-world application, this simplifying assumption makes sense due to small distances between jobs. However, a future direction might be considering geography in making decisions. For instance, the job scheduling model can include a penalty for two jobs of long distances scheduled for the same day.

Another possible direction is to have emergencies with random durations. This is related to literature on scheduling under stochastic job durations, where jobs need to be processed on parallel machines without preemption. The number of jobs is known (unlike our setting), but the processing time of each job is an independent random variable. The objective is to minimize expected makespan (like our setting). It is known that the longest-expected-processing-time (LEPT) rule minimizes the expected makespan for exponential jobs or for remainders of i.i.d. decreasing hazard rate jobs (Pinedo & Weiss, 1979; Weber, 1982). In general, LEPT is a good but not optimal heuristic (Pinedo & Weiss, 1979). For this reason, and based on preliminary experiments, we believe that Algorithm Stoch-LPT would perform well under the case where emergency durations are random.

Another potential direction is an analytical performance guarantee for the crew assignment heuristic, Stoch-LPT. We are able to prove that Stoch-LPT terminates with the optimal crew assignment under special cases. However, establishing a guarantee for the general case is an interesting direction.

We demonstrated the potential impact of the Resource Allocation Planning Tool (RAPT) in managing uncertainty in yard operations and decreasing labor costs. However, there is still some further work to be done in order for the yards to achieve these results. These include gaining grassroot support from the workers' union and continuing with strong management leadership. Some new processes need to be also introduced in all of the company's yard to ensure that the tool can be implemented successfully. The purpose of these new processes is to ensure integrity of the data fed into the model, and to create multiple levels of accountability for better oversight and cost control.

Appendix A

Miscellaneous

A.1 Distribution of estimator Δ_n

We use the strong representation for L-statistics by Govindarajulu & Mason (1983) to derive the distribution of Δ_n. We can write the weight functions in (3.6) as

$$J_i(\beta, n) = n \int_{(i-1)/n}^{i/n} J(u)du, \quad \text{for } i = 1, 2, \dots, n,$$

where $J : [0, 1] \mapsto \mathbb{R}$ is the score function

$$J(u) = \begin{cases} -\frac{1}{\beta}, & \text{if } u \in [0, \beta], \\ \frac{1}{1-\beta}, & \text{if } u \in (\beta, 1]. \end{cases}$$

Note that the score function has only a single point of discontinuity at $u = \beta$.

Define

$$\mu(J, F) \triangleq \int_0^1 J(u)F^{-1}(u)du,$$
$$\sigma^2(J, F) \triangleq \frac{1}{2}E \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} J(F(u))J(F(v)) \left[I(u, D_1) - I(u, D_2)\right] \left[I(v, D_1) - I(v, D_2)\right] dudv,$$

where $I(x, D)$ is an indicator function which takes a value 0 if $x \leq D$ and 1 otherwise. Govindarajulu & Mason (1983) show that

$$\Delta_n = \mu(J, F) + R_n + \frac{1}{n}\sum_{i=1}^{n} Z_n, \tag{A.1}$$

where Z_i are i.i.d. with mean 0, and R_n is such that $\sqrt{n}R_n \to 0$ a.s. as $n \to \infty$. Stigler (1974) shows the limiting values of the mean and variance of an L-statistic. We have that

$$\lim_{n\to\infty} E(\Delta_n) = \mu(J, F), \tag{A.2}$$
$$\lim_{n\to\infty} n\text{Var}(\Delta_n) = \sigma^2(J, F), \tag{A.3}$$

where

$$\mu(J,F) = \frac{1}{1-\beta}E(D|D \geq F^{-1}(\beta)) - \frac{1}{\beta}E(D|D \leq F^{-1}(\beta)) = \Delta(F^{-1}(\beta)),$$

$$\sigma^2(J,F) = \frac{1}{\beta}\text{Var}(D|D \leq F^{-1}(\beta)) + \frac{1}{1-\beta}\text{Var}(D|D \geq F^{-1}(\beta))$$
$$+ \left(\sqrt{\frac{1-\beta}{\beta}}\left[E(D|D \leq F^{-1}(\beta)) - F^{-1}(\beta)\right] + \sqrt{\frac{\beta}{1-\beta}}\left[E(D|D \geq F^{-1}(\beta)) - F^{-1}(\beta)\right]\right)^2.$$

Moreover, if the distribution F is such that the score function J is continuous a.e. F^{-1} and $E(D^2) < \infty$, then from Theorem 2 of Stigler (1974), Δ_n is asymptotically normally distributed. In particular,

$$\frac{\Delta_n - E(\Delta_n)}{\sqrt{\text{Var}(\Delta_n)}} \to_d N(0,1). \tag{A.4}$$

A.2 Regression analysis to estimate the bias of Δ_n

In this section, we estimate numerically the bias $B \triangleq E(\Delta_n) - \Delta_F(\beta)$ using regression analysis. The factors that affect B are the sample size n, the profit margin β, and the demand distribution F. We vary values for each one of these factors and perform regression to estimate B.

First, we estimate its dependence on n. For a given n, we fix a distribution and profit margin, generate n samples from the distribution and compute Δ_n. We do this for 10,000 repetitions and take the average difference of $\Delta_n - \Delta_F(\beta)$. We denote this average difference as b_n. We can generate (n, b_n) pairs for $n \in \{20, 40, 80, 160, 320, 640, 1280, 2560, 5120\}$. We now use this pairwise data to estimate a relationship $B = \hat{C}_1 n^{\hat{k}_1}$ using power regression. Table D.7 summarizes the results for various distributions and profit margins. The results seem to suggest that the bias is inversely proportional to the sample size, i.e., $B \propto \frac{1}{n}$.

Estimating how the bias is affected by the distribution and profit margin is more complicated. This is because it is unclear which property of the distribution directly influences the bias. We fix a sample size $n = 100$ and vary the profit margin and distribution. Same as before, we generate n samples to compute Δ_n. We take the average difference of $\Delta_n - \Delta_F(\beta)$ over 10,000 repetitions. From analyzing the results, we observe that distributions with larger values for $\Delta_F(\beta)$ tend to have a larger bias (see Figure C-2 for pairwise values). Therefore, we use pairs of AMS values and bias values to estimate the relationship $B = \hat{C}_2 \Delta_F(\beta)^{\hat{k}_2}$. Figure C-2 plots the pairwise values and the regression equation. Based on the regression analysis, $B \propto \Delta_F(\beta)$. Moreover, the coefficient of regression equation is almost entirely explained by the factor $\frac{1}{n}$.

Combining these observations, we conjecture that $B = -\frac{1}{n}\Delta_F(\beta)$. This motivates the following modified estimator of spread: $\tilde{\Delta}_n = \Delta_n + \frac{1}{n}\Delta_n$.

A.3 Theorem A.3.1

Theorem A.3.1. *Consider the set $\mathcal{D}_{\mu,\delta_L,\delta_U,+}$ consisting of all nonnegative distributions with common mean μ and AMS (at the β quantile) in the range $[\delta_L, \delta_U]$. Then the minimax regret and minimax regret quantity are:*

1. *If $\delta_L < \frac{(\beta-1/2)\mu}{\beta(1-\beta)}$, then*

$$y^*_{\mu,\delta_L,\delta_U,+} = \begin{cases} \frac{1}{\mu}\left(\mu-(1-\beta)\delta_U\right)\left(\mu+\beta\delta_U\right), & \text{if } \delta_U \in \left[\delta_L, \frac{\beta\mu}{1-\beta^2}\right), \\ \frac{(1-\beta)}{4\mu}\left(\frac{\mu}{1-\beta}+\mu-(1-\beta)\delta_U\right)^2, & \text{if } \delta_U \in \left[\frac{\beta\mu}{1-\beta^2}, \frac{\mu}{1-\beta}\right), \\ \frac{\mu}{4(1-\beta)}, & \text{if } \delta_U \in \left[\frac{\mu}{1-\beta}, \infty\right), \end{cases}$$

$$\rho^*_{\mu,\delta_L,\delta_U,+} = \begin{cases} \frac{1}{\mu}\beta(1-\beta)\delta_U\left(\mu-(1-\beta)\delta_U\right), & \text{if } \delta_U \in \left[\delta_L, \frac{\beta\mu}{1-\beta^2}\right), \\ \frac{1}{4\mu}\left(\beta\mu+(1-\beta)^2\delta_U\right)^2, & \text{if } \delta_U \in \left[\frac{\beta\mu}{1-\beta^2}, \frac{\mu}{1-\beta}\right), \\ \frac{\mu}{4}, & \text{if } \delta_U \in \left[\frac{\mu}{1-\beta}, \infty\right), \end{cases}$$

2. *If $\frac{(\beta-1/2)\mu}{\beta(1-\beta)} \le \delta_L < \frac{\beta\mu}{(1-\beta)(1+\beta)}$, then*

$$y^*_{\mu,\delta_L,\delta_U,+} = \begin{cases} \frac{1}{\mu}\left(\mu-(1-\beta)\delta_U\right)\left(\mu+\beta\delta_U\right), & \text{if } \delta_U \in \left[\delta_L, \frac{\beta\mu}{1-\beta^2}\right), \\ \frac{(1-\beta)}{4\mu}\left(\frac{\mu}{1-\beta}+\mu-(1-\beta)\delta_U\right)^2, & \text{if } \delta_U \in \left[\frac{\beta\mu}{1-\beta^2}, \frac{\beta\mu-2\beta(1-\beta)\delta_L}{(1-\beta)^2}\right), \\ \frac{\mu+\beta\delta_L}{\mu}\left(\mu-(1-\beta)\delta_U+\beta(1-\beta)(\delta_U-\delta_L)\right), & \text{if } \delta_U \in \left[\frac{\beta\mu-2\beta(1-\beta)\delta_L}{(1-\beta)^2}, \frac{\mu}{1-\beta}\right), \\ \frac{\beta}{\mu}\left(\mu-(1-\beta)\delta_L\right)\left(\mu+\beta\delta_L\right), & \text{if } \delta_U \in \left[\frac{\mu}{1-\beta}, \infty\right) \end{cases}$$

$$\rho^*_{\mu,\delta_L,\delta_U,+} = \begin{cases} \frac{1}{\mu}\beta(1-\beta)\delta_U\left(\mu-(1-\beta)\delta_U\right), & \text{if } \delta_U \in \left[\delta_L, \frac{\beta\mu}{1-\beta^2}\right), \\ \frac{1}{4\mu}\left(\beta\mu+(1-\beta)^2\delta_U\right)^2, & \text{if } \delta_U \in \left[\frac{\beta\mu}{1-\beta^2}, \frac{\beta\mu-2\beta(1-\beta)\delta_L}{(1-\beta)^2}\right), \\ \frac{\beta(1-\beta)}{\mu}\left(\mu-(1-\beta)\delta_L\right)\left((1-\beta)\delta_U+\beta\delta_L\right), & \text{if } \delta_U \in \left[\frac{\beta\mu-2\beta(1-\beta)\delta_L}{(1-\beta)^2}, \frac{\mu}{1-\beta}\right), \\ \frac{\beta(1-\beta)}{\mu}\left(\mu-(1-\beta)\delta_L\right)\left(\mu+\beta\delta_L\right), & \text{if } \delta_U \in \left[\frac{\mu}{1-\beta}, \infty\right) \end{cases}$$

3. *If $\frac{\beta\mu}{(1-\beta)(1+\beta)} \le \delta_L \le \frac{\mu}{1-\beta}$, then*

$$y^*_{\mu,\delta_L,\delta_U,+} = \begin{cases} \frac{\mu+\beta\delta_L}{\mu}\left(\mu-(1-\beta)\delta_U+\beta(1-\beta)(\delta_U-\delta_L)\right), & \text{if } \delta_U \in \left[\delta_L, \frac{\mu}{1-\beta}\right), \\ \frac{\beta}{\mu}\left(\mu-(1-\beta)\delta_L\right)\left(\mu+\beta\delta_L\right), & \text{if } \delta_U \in \left[\frac{\mu}{1-\beta}, \infty\right) \end{cases}$$

$$\rho^*_{\mu,\delta_L,\delta_U,+} = \begin{cases} \frac{\beta(1-\beta)}{\mu}\left(\mu-(1-\beta)\delta_L\right)\left((1-\beta)\delta_U+\beta\delta_L\right), & \text{if } \delta_U \in \left[\delta_L, \frac{\mu}{1-\beta}\right), \\ \frac{\beta(1-\beta)}{\mu}\left(\mu-(1-\beta)\delta_L\right)\left(\mu+\beta\delta_L\right), & \text{if } \delta_U \in \left[\frac{\mu}{1-\beta}, \infty\right) \end{cases}$$

A.4 Deterministic equivalent of the joint job scheduling and crew assignment problem

Proposition A.4.1. *The deterministic equivalent of the optimization problem* (4.2) *is the following mixed integer program.*

$$
\begin{aligned}
\underset{C,V,X,Y,Z}{\textit{minimize}} \quad & C \\
\textit{subject to} \quad & \sum_{\omega\in\Omega_t} P_t(\omega)V_t(\omega) \le C, \qquad t=1,\dots,T, \\
& d_L Z_{tk}(\omega) + \sum_{i=1}^{n} d_i Y_{itk} \le V_t(\omega), \qquad t=1,\dots,T,\ \omega\in\Omega_t,\ k=1,\dots,K_t, \\
& \sum_{t=1}^{\tau_i} X_{it} = 1, \qquad i=1,\dots,n, \\
& \sum_{k=1}^{K_t} Y_{itk} = X_{it}, \qquad i=1,\dots,n,\ t=1,\dots,T, \\
& \sum_{k=1}^{K_t} Z_{tk}(\omega) = L(\omega), \qquad t=1,\dots,T,\ \omega\in\Omega_t, \\
& X_{it}\in\{0,1\}, \qquad i=1,\dots,n,\ t=1,\dots,T, \\
& Y_{itk}\in\{0,1\}, \qquad i=1,\dots,n,\ t=1,\dots,T,\ k=1,\dots,K_t, \\
& Z_{tk}(\omega)\in\mathbb{Z}^+, \qquad t=1,\dots,T,\ \omega\in\Omega_t,\ k=1,\dots,K_t.
\end{aligned}
\tag{A.5}
$$

Proof. Let us denote by $\mathcal{F}$ the feasible region of (4.2). We can write the first-stage problem as:

$$
\begin{aligned}
\underset{W,X,Y}{\text{minimize}} \quad & W \\
\text{subject to} \quad & E_t\left[F_t\left(Y_t, L(\omega)\right)\right] \le W, \qquad t=1,\dots,T, \\
& (X,Y)\in\mathcal{F}.
\end{aligned}
\tag{A.6}
$$

Using the probability distribution P_t for the gas leak scenarios Ω_t, we can rewrite constraint set (A.6) for each t as $\sum_{\omega\in\Omega_t} P_t(\omega)\times F_t(Y_t, L(\omega)) \le W$.

Similarly, we can rewrite the second-stage recourse problem $F_t(Y_t, L(\omega))$ as an MIP:

$$
\begin{aligned}
\underset{V,Z}{\text{minimize}} \quad & V_t(\omega) \\
\text{subject to} \quad & d_L Z_{tk}(\omega) + \sum_{i=1}^{n} d_i Y_{itk} \le V_t(\omega), \qquad k=1,\dots,K_t, \\
& \sum_{k=1}^{K_t} Z_{tk}(\omega) = L(\omega) \\
& Z_{tk}(\omega)\in\mathbb{Z}^+, \qquad k=1,\dots,K_t.
\end{aligned}
$$

Combining the above two reformulations results in the optimization problem (A.5) in the proposition. Since Ω_t is finite, there is a finite number of constraints, and problem (A.5) is a MIP. □

A.5 Job scheduling LP-based heuristic

The following describes the job scheduling algorithm (Algorithm LP-schedule) in detail. Consider the following linear programming (LP) relaxation of the job scheduling MIP:

$$\begin{aligned}
\underset{C,X}{\text{minimize}} \quad & C \\
\text{subject to} \quad & d_L E_t[L(\omega)] + \sum_{i=1}^{n} d_i X_{it} \leq K_t C, \qquad t = 1, \ldots, T, \\
& \sum_{t=1}^{\tau_i} X_{it} = 1, \qquad i = 1, \ldots, n, \\
& X_{it} \geq 0, \qquad i = 1, \ldots, n, \; t = 1, \ldots, T.
\end{aligned} \tag{A.7}$$

Denote the optimal solution as X^{LP} and the optimal cost as C^{LP}. To round X^{LP} into a feasible job schedule, define the following sets:

$$\begin{aligned}
I^s(t) &= \left\{ i : 0 < X_{it}^{LP} < 1 \right\}, \\
I^f(t) &= \left\{ i : X_{it}^{LP} = 1 \right\}, \\
T_i &= \left\{ t : 0 < X_{it}^{LP} < 1 \right\}.
\end{aligned}$$

The rounding step of the algorithm consists of solving the following MIP:

$$\begin{aligned}
\underset{W,X}{\text{minimize}} \quad & W \\
\text{subject to} \quad & d_L E_t[L(\omega)] + \sum_{i \in I^f(t)} d_i + \sum_{i \in I^s(t)} d_i X_{it} \leq K_t \cdot W, \qquad t = 1, \ldots, T, \\
& \sum_{t \in T_i} X_{it} = 1, \qquad i \in I^s(1) \cup \cdots \cup I^s(T), \\
& X_{it} \in \{0, 1\}, \qquad t = 1, \ldots, T, \; i \in I^s(t).
\end{aligned} \tag{A.8}$$

Note that any set of variables $\{X_{it}\}$ satisfying the last two constraints in (A.8) is a rounding of the fractional variables of the LP solution $\{X_{it}^{LP}\}$. Let us denote by $\{X_{it}^{R}\}$ the solution to the MIP (A.8). For all $i = 1, \ldots, n$ and $t = 1, \ldots, T$, set the rounded solution X_{it}^{H} by the following equation:

$$X_{it}^{H} = \begin{cases} 0, & \text{if } X_{it}^{LP} = 0, \\ 1, & \text{if } X_{it}^{LP} = 1, \\ X_{it}^{R}, & \text{otherwise.} \end{cases}$$

Note that X^H is a feasible solution to the original job scheduling problem (4.3).

Algorithm 2 [LP-schedule] LP-based job scheduling algorithm

Require: Planning horizon $\{1, \ldots, T\}$, and standard jobs indexed by $1, \ldots, n$, where job i has deadline $\tau_i \leq T$ and duration d_i
Ensure: Feasible schedule X^H with $\sum_{t=1}^{\tau_i} X_{it}^H = 1$, for all $i = 1, \ldots, n$, and $X_{it}^H \in \{0, 1\}$

$X^{LP} \leftarrow$ solution to the linear relaxation (A.7)
Initialize $X_{it}^H \leftarrow 0$ for all i, t
$I^f(t), I^s(t) \leftarrow \varnothing$ for all t, and $T_i \leftarrow \varnothing$ for all i
for $i = 1$ **to** n **do**
    for $t = 1$ **to** T **do**
        if $X_{it}^{LP} = 1$ **then**
            $X_{it}^H \leftarrow 1$
            $I^f(t) \leftarrow I^f(t) \cup \{i\}$
        else if $X_{it}^{LP} \in (0, 1)$ **then**
            $I^s(t) \leftarrow I^s(t) \cup \{i\}$
            $T_i \leftarrow T_i \cup \{t\}$
$X^R \leftarrow$ solution deterministic rounding MIP (A.8)
for $t = 1$ **to** T **do**
    for $i \in I^s(t)$ **do**
        $X_{it}^H \leftarrow X_{it}^R$

A.6 Binary search initialization for the job scheduling LP-based heuristic

Here, we describe the details of the job scheduling algorithm (Algorithm BinLP-schedule) which is initialized by a binary search procedure, then solves an LP relaxation of a MIP, and rounds the LP solution to a feasible schedule. The binary search procedure is adapted from Lenstra *et al.* (1990). For a fixed parameter C, define the following set of job-date pairs:

$$\Gamma_C \triangleq \left\{ (i, t) : t \leq \tau_i \text{ and } \frac{d_i}{K_t} \leq C \right\}.$$

Consider the solving the following linear optimization problem $LP(C)$:

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & C \\ \text{subject to} \quad & d_L E_t[L(\omega)] + \sum_{i:(i,t) \in \Gamma_C} d_i X_{it} \leq K_t C, \qquad t = 1, \ldots, T, \\ & \sum_{t:(i,t) \in \Gamma_C} X_{it} = 1, \qquad i = 1, \ldots, n, \\ & X_{it} \geq 0, \qquad (i, t) \in \Gamma_C. \end{aligned} \tag{A.9}$$

Using binary search, find the smallest value of C for which the LP- problem $LP(C)$ is feasible. Let C^B be this value, and X^B the corresponding optimal solution. We round X^B into a feasible job schedule in the same manner described in Section A.5 for rounding X^{LP}.

Algorithm 3 [BinLP-schedule] LP-based job scheduling algorithm with binary search initialization procedure

Require: Planning horizon $\{1, \ldots, T\}$, and standard jobs indexed by $1, \ldots, n$, where job i has deadline $\tau_i \leq T$ and duration d_i

Ensure: Feasible schedule $X^{H'}$ with $\sum_{t=1}^{\tau_i} X_{it}^{H'} = 1$, for all $i = 1, \ldots, n$, and $X_{it}^{H'} \in \{0, 1\}$

Initialize $u \leftarrow$ makespan of arbitrary feasible job schedule
Initialize $l \leftarrow 0$
while $l < u$ **do**
    $C \leftarrow \left\lfloor \frac{1}{2}(l + u) \right\rfloor$
    Solve $LP(C)$ in (A.9)
    if $LP(C)$ is feasible **then**
        $u \leftarrow C$
    else
        $l \leftarrow C$
$C^B \leftarrow C$
$X^B \leftarrow$ solution of $LP(C^B)$
Initialize $X_{it}^{H'} \leftarrow 0$ for all i, t
$I^f(t), I^s(t) \leftarrow \varnothing$ for all t, and $T_i \leftarrow \varnothing$ for all i
for $i = 1$ **to** n **do**
    for $t = 1$ **to** T **do**
        if $X_{it}^B = 1$ **then**
            $X_{it}^{H'} \leftarrow 1$
            $I^f(t) \leftarrow I^f(t) \cup \{i\}$
        else if $X_{it}^B \in (0, 1)$ **then**
            $I^s(t) \leftarrow I^s(t) \cup \{i\}$
            $T_i \leftarrow T_i \cup \{t\}$
$X^{R'} \leftarrow$ solution deterministic rounding MIP (A.8)
for $t = 1$ **to** T **do**
    for $i \in I^s(t)$ **do**
        $X_{it}^{H'} \leftarrow X_{it}^{R'}$

A.7 Performance guarantee for job scheduling heuristic with binary search initialization

Theorem A.7.1. *Let C^{OPT} be the optimal objective value of the scheduling phase problem* (4.3)*. If C^B is the result of the binary search in Algorithm BinLP-schedule and $X^{H'}$ is the schedule. Then $X^{H'}$ is feasible for the scheduling phase problem* (4.3)*, and has a cost $C^{H'}$ such that*

$$\frac{C^{H'}}{C^{OPT}} \leq \min\left\{2,\ 1+\frac{1}{C^B}\left(\min_{t=1,\dots,T} K_t\right)^{-1}\sqrt{\frac{1}{2}\left(\sum_{i=1}^{n} d_i^2\right)(1+\ln\delta)}\right\},$$

where $\delta = \max_{t=1,\dots,T} |\delta_t|$ and $\delta_t \triangleq \left\{r = 1,\dots,T : X_{ir}^{LP} > 0 \text{ and } X_{it}^{LP} > 0\right\}$.

Proof. Using the binary search procedure in Section A.6, Lenstra *et al.* (1990) show that the LP solution has a rounding in which the makespan is at most 2 times the optimal makespan C^{OPT}. Since the rounding procedure of Algorithm BinLP-schedule results in the rounding $X^{H'}$ with the smallest makespan, then the makespan of $X^{H'}$ is at most $2C^{OPT}$. Moroever, with a minor modification of the proof of Theorem 4.4.2 (see Section B.11), we can prove that $C^{H'} \leq C^B\left(1+H\left(C^B, \frac{1}{e\delta}\right)\right)$. Thus, since $C^B \leq C^{OPT}$, we have that

$$\frac{C^{H'}}{C^{OPT}} \leq \min\left\{2,\ 1+\frac{1}{C^B}\left(\min_{t=1,\dots,T} K_t\right)^{-1}\sqrt{\frac{1}{2}\left(\sum_{i=1}^{n} d_i^2\right)(1+\ln\delta)}\right\}.$$

□

A.8 Deterministic equivalent of the assignment phase problem

Proposition A.8.1. *The deterministic equivalent of the day t two-stage assignment phase problem* (4.5) *is the following mixed integer program.*

$$\begin{aligned}
\underset{V,Y,Z}{\text{minimize}} \quad & \sum_{\omega\in\Omega_t} P_t(\omega)V(\omega) \\
\text{subject to} \quad & d_L Z_k(\omega) + \sum_{i\in I_t} d_i Y_{ik} \leq V(\omega), \qquad \omega\in\Omega_t,\ k=1,\dots,K_t, \\
& \sum_{k=1}^{K_t} Y_{ik} = 1, \qquad i \in I_t, \\
& \sum_{k=1}^{K_t} Z_k(\omega) = L(\omega), \qquad \omega\in\Omega_t, \\
& Y_{ik}\in\{0,1\}, \qquad i\in I_t,\ k=1,\dots,K_t \\
& Z_k(\omega)\in\mathbb{Z}^+, \qquad \omega\in\Omega_t,\ k=1,\dots,K_t.
\end{aligned} \tag{A.10}$$

Proof. Let us denote by $\mathcal{F}$ the feasible region of (4.5). Using the probability distribution P_t for the gas leak scenarios Ω_t, we can rewrite the objective function of (4.5) as $\sum_{\omega\in\Omega_t} P_t(\omega)F_t(Y, L(\omega))$.

Similarly, we can rewrite the second-stage recourse problem $F_t(Y, L(\omega))$ as an MIP:

$$
\begin{aligned}
\underset{V,Z}{\text{minimize}} \quad & V(\omega) \\
\text{subject to} \quad & d_L Z_k(\omega) + \sum_{i=1}^{n} d_i Y_{ik} \leq V(\omega), \qquad k = 1, \ldots, K_t, \\
& \sum_{k=1}^{K_t} Z_k(\omega) = L(\omega), \\
& Z_k(\omega) \in \mathbb{Z}^+, \qquad k = 1, \ldots, K_t.
\end{aligned}
$$

Therefore, (4.5) is equivalent to:

$$
\begin{aligned}
\underset{V,Y,Z}{\text{minimize}} \quad & \sum_{\omega\in\Omega_t} P_t(\omega)V(\omega) \\
\text{subject to} \quad & d_L Z_k(\omega) + \sum_{i=1}^{n} d_i Y_{ik} \leq V(\omega), \qquad \omega \in \Omega_t,\ k = 1, \ldots, K_t, \\
& \sum_{k=1}^{K_t} Z_k(\omega) = L(\omega), \qquad \omega \in \Omega_t, \\
& Z_k(\omega) \in \mathbb{Z}^+, \qquad \omega \in \Omega_t,\ k = 1, \ldots, K_t, \\
& Y \in \mathcal{F}.
\end{aligned}
$$

Since Ω_t is finite, there is a finite number of constraints, and this problem is a MIP. □

A.9 Optimal crew assignment for examples

In these examples, we will assign 15 standard jobs to 7 crews, under different probability distributions for the number of leaks. Table D.23 gives the durations of the standard jobs. Table D.24 show the seven different probability distributions used in our experiments. The optimal crew assignment solution is given in the Tables D.25–D.31.

A.10 Crew assignment with Algorithm Stoch-LPT

The algorithm for performing crew assignment under random occurrence of emergencies (Stoch-LPT) is described in the following Algorithm 4.

In several examples, we will assign using Algorithm Stoch-LPT standard jobs to crews, under different probability distributions for the number of leaks. Table D.23 gives the durations of the standard jobs. Table D.24 show the seven different probability distributions used in our experiments. The crew assignment resulting from Algorithm Stoch-LPT is given in the Tables D.32–D.38.

Algorithm 4 [Stoch-LPT] Crew assignment algorithm (stochastic variant of LPT)

Require: $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$, where $L(\omega_1) < L(\omega_2) < \cdots < L(\omega_m)$, and standard jobs sorted in decreasing job duration, i.e. $d_1 \geq d_2 \geq \cdots \geq d_n$

Ensure: Assignment of all standard jobs and gas leak jobs to crews under all leak scenarios

$B_k(\omega_{m+1}) \leftarrow 1$, for all $k \in K$
for $s = m$ **to** 1 **do**
  $B_k(\omega_s) \leftarrow 0$, for all $k \in K$
  for $l = 1$ **to** $L(\omega_2)$ **do**
    $\tilde{K} \leftarrow \arg\min_{k \in K} (B_k(\omega_s))$ {set of crews with smallest current load}
    $B_{k_0}(\omega_s) \leftarrow B_{k_0}(\omega_s) + d_L$, where $k_0 \in \tilde{K}$ such that $B_{k_0}(\omega_s) < B_{k_0}(\omega_{s+1})$
for $i = 1$ **to** n **do**
  for $k \in K$ **do**
    $\tilde{B}_k(\omega) \leftarrow B_k(\omega) + d_i$, for all $\omega \in \Omega$ {Load in scenario ω if job i is assigned to crew k}
    $A_k(\omega) \leftarrow \max\Big(B_1(\omega), \ldots, B_{k-1}(\omega), \tilde{B}_k(\omega), B_{k+1}(\omega), \ldots, B_K(\omega)\Big)$, for all $\omega \in \Omega$ {Makespan in scenario ω if job i is assigned to crew k}
  $K_i \leftarrow \arg\min_{k \in K} \{\sum_s P(\omega_s) A_k(\omega_s)\}$
  $k_0 \in \arg\min_{k \in K_i} \{\sum_s P(\omega_s) B_k(\omega_s)\}$
  $B_{k_0}(\omega) \leftarrow \tilde{B}_{k_0}(\omega)$, for all $\omega \in \Omega$

Appendix B

Proofs

B.1 Proof of Theorem 2.3.2

As a preliminary for the proof, let us first state a version of Bernstein's inequality (Bernstein, 1927):

Theorem B.1.1 (Bernstein's inequality)**.** *Let $X^1, X^2, \ldots, X^N$ be i.i.d. random variables such that $|X^1| \leq c$ almost surely, and $Var(X^1) = \sigma^2$. Then, for any $t > 0$,*

$$\Pr\left(\frac{1}{N}\sum_{i=1}^{N} X^i - E[X^1] \geq t\right) \leq \exp\left(\frac{-Nt^2}{2\sigma^2 + 2tc/3}\right).$$

For the proof of Theorem 2.3.2, we will require the following proposition.

Proposition B.1.2. *Suppose $\hat{Q}_N$ is the $\frac{b}{b+h}$ quantile of a random sample from D with size N. Then, for any $\gamma > 0$,*

$$\Pr\left(\partial_- C(\hat{Q}_N) \leq \gamma \text{ and } \partial_+ C(\hat{Q}_N) \geq -\gamma\right) \geq 1 - 2\exp\left(\frac{-3N\gamma^2}{6bh + 8\gamma(b+h)}\right).$$

Proof. Let $\bar{F}$ be the complementary cdf of D, i.e., $\bar{F}(q) = \Pr(D \geq q) = 1 - F(q) + \Pr(D = q)$. For a random sample $\{D^1, \ldots, D^N\}$ drawn from D, let $\hat{Q}_N$ be the $\frac{b}{b+h}$ sample quantile. Define

$$\hat{F}_N(q) \triangleq \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{[D^i \leq q]},$$

$$\hat{\bar{F}}_N(q) \triangleq \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{[D^i \geq q]}.$$

For simplicity, define $\alpha \triangleq \frac{\gamma}{b+h}$ and $\beta \triangleq \frac{b}{b+h}$. Define the events $B \triangleq [\partial_+ C(\hat{Q}_N) < -\gamma] = [F(\hat{Q}_N) < \beta - \alpha]$ and $L \triangleq [\partial_- C(\hat{Q}_N) > \gamma] = [\bar{F}(\hat{Q}_N) < 1 - \beta - \alpha]$. To prove Proposition B.1.2, we need to find an upper bound for $\Pr(B)$ and for $\Pr(L)$.

Define the quantile $q_1 \triangleq \inf\{q : F(q) \geq \beta - \alpha\}$. Since F is nondecreasing, we have that $B = [\hat{Q}_N < q_1]$. Consider a monotonically decreasing, nonnegative sequence $\{\tau^k\}_{k=1}^{\infty}$, where $\tau^k \downarrow 0$. Define the sequence of events $\{B_k\}_{k=1}^{\infty}$, where

$$B_k \triangleq [\hat{Q}_N \leq q_1 - \tau^k] = [\hat{F}_N(q_1 - \tau^k) \geq \beta].$$

Note that since $\hat{F}_N(q_1 - \tau^k) \leq \hat{F}_N(q_1 - \tau^{k+1})$, then it follows that $B_k \subseteq B_{k+1}$. Thus, we have that $B_k \uparrow \lim_{k\to\infty} B_k \triangleq \bar{B}$, which implies $\Pr(B_k) \uparrow \Pr(\bar{B})$. Note also that $B \subseteq \bar{B}$, thus $\Pr(B) \leq \Pr(\bar{B})$.

From the definition of q_1, observe that for every $k \geq 1$, there exists $\varepsilon_k > \alpha$ such that $F(q_1 - \tau^k) = \beta - \varepsilon_k < \beta - \alpha$. Note that

$$F(q_1 - \tau^k)\left(1 - F(q_1 - \tau^k)\right) < (\beta - \alpha)(1 - \beta + \varepsilon_k). \tag{B.1}$$

Thus, we have that

$$\begin{aligned}
\Pr(B_k) &= \Pr(\hat{F}_N(q_1 - \tau^k) \geq \beta), \\
&= \Pr(\hat{F}_N(q_1 - \tau^k) - F(q_1 - \tau^k) \geq \varepsilon_k), \\
&\leq \exp\left(\frac{-N\varepsilon_k^2/2}{F(q_1 - \tau^k)(1 - F(q_1 - \tau^k)) + \frac{\varepsilon_k}{3}}\right), && \text{(B.2)} \\
&\leq \exp\left(\frac{-N\varepsilon_k/2}{\frac{1}{\varepsilon_k}(\beta - \alpha)(1 - \beta) + \beta - \alpha + \frac{1}{3}}\right), && \text{(B.3)}
\end{aligned}$$

where (B.2) follows from Bernstein's inequality and (B.3) follows from inequality (B.1). Now, since $\varepsilon_k > \alpha$, for all $k \geq 1$, we have that

$$\begin{aligned}
\Pr(B_k) &\leq \exp\left(\frac{-N\alpha/2}{\frac{1}{\alpha}\beta(1-\beta) - \frac{2}{3} + 2\beta - \alpha}\right), \\
&\leq \exp\left(\frac{-N\alpha/2}{\frac{1}{\alpha}\beta(1-\beta) + \frac{4}{3} - 2\min(\beta, 1-\beta) - \alpha}\right), \\
&\leq \exp\left(\frac{-N\alpha/2}{\frac{1}{\alpha}\beta(1-\beta) + \frac{4}{3}}\right) = \exp\left(\frac{-3N\gamma^2}{6bh + 8\gamma(b+h)}\right) \triangleq \delta.
\end{aligned}$$

Thus, $\Pr(B) \leq \Pr(\bar{B}) \leq \delta$. In fact, by going through a similar argument, we can show that $\Pr(L) \leq \delta$. Thus, by the union bound, we have that

$$\Pr\left(\partial_- C(\hat{Q}_N) > \gamma \text{ or } \partial_+ C(\hat{Q}_N) < -\gamma\right) = \Pr(B \cup L) \leq \Pr(B) + \Pr(L) \leq 2\delta,$$

proving Proposition B.1.2. □

We can now proceed with the proof of Theorem 2.3.2. Note that S_ϵ^{LRS} consists of all q for which $\partial_- C(q) \leq \gamma$ and $\partial_+ C(q) \geq -\gamma$, with $\gamma = \frac{\epsilon}{3}\min(b, h)$. From Proposition B.1.2, the SAA solution from a random sample with size N lies in S_ϵ^{LRS} with probability at least

$$1-2\exp\left(\frac{-N\epsilon^2\min\{b,h\}^2}{18bh+8\epsilon(b+h)\min\{b,h\}}\right)=1-2\exp\left(\frac{-N\epsilon^2\min\{b,h\}}{18\max\{b,h\}+8\epsilon(b+h)}\right).$$
$$\geq 1-2\exp\left(\frac{-N\epsilon^2}{18+8\epsilon}\cdot\frac{\min\{b,h\}}{b+h}\right)\quad\square$$

B.2 Proof of Theorem 2.4.1

Since C is convex, $S_\epsilon^f\cap[q^*,\infty)$ can be equivalently expressed as $\{q : C'(q)\leq C'(\overline{q})$ and $q\geq q^*\}$. Note that,

$$\begin{aligned}C'(\overline{q})&=(b+h)(F(\overline{q})-F(q^*))=(b+h)\left[(\overline{q}-q^*)f(q^*)+O(\overline{q}-q^*)^2\right]\\&=\sqrt{2\epsilon bh\Delta(q^*)f(q^*)}+O(\epsilon),\end{aligned}\tag{B.4}$$

which follows from Taylor series approximation and from the definition of $\overline{q}$ in (2.7).

To prove Theorem 2.4.1, note that the event that $\tilde{Q}_N^\alpha\in S_\epsilon^f\cap[q^*,\infty)$, where $\alpha=C'(\overline{q})$, is equivalent to the intersection of events $[\tilde{Q}_N^\alpha\geq q^*]$ and $[C'(\tilde{Q}_N^\alpha)\leq\alpha]$. We will prove an upper bound on the probability of $[\tilde{Q}_N^\alpha<q^*]$ and on the probability of $[C'(\tilde{Q}_N^\alpha)>\alpha]$. It follows similar lines to the proof of Lemma 3.5 in Levi *et al.* (2007), except we will use Bernstein's inequality instead of Hoeffding's inequality.

Define $\beta\triangleq\frac{b}{b+h}$ and $\gamma\triangleq\frac{1}{2}\frac{\alpha}{b+h}$. First, let us bound the probability of $B\triangleq[\tilde{Q}_N^\alpha<q^*]$. For a real-valued sequence $\{\tau^k\}_{k=1}^\infty$ where $\tau^k\downarrow 0$, define

$$B_k\triangleq[\tilde{Q}_N^\alpha\leq q^*-\tau^k]=\left[-b+(b+h)\hat{F}_N(q^*-\tau^k)\geq\frac{\alpha}{2}\right]=[\hat{F}_N(q^*-\tau^k)\geq\beta+\gamma].$$

Note that since $\hat{F}_N$ is monotonically increasing, it follows that $B_k\subseteq B_{k+1}$. Thus, if $\bar{B}$ is the limiting event of the sequence of events $\{B_k\}_{k=1}^\infty$, then $B_k\uparrow\bar{B}$, implying that $\Pr(B_k)\uparrow\Pr(\bar{B})$. Note also that $B\subseteq\bar{B}$, thus $\Pr(B)\leq\Pr(\bar{B})$. Therefore, to bound $\Pr(B)$, we only need to find a uniform upper bound for $\Pr(B_k)$.

Note that for any $k\geq 1$, there exists $\varepsilon^k>0$ such that $F(q^*-\tau^k)=\beta-\varepsilon^k$. Thus,

$$F(q^*-\tau^k)\left(1-F(q^*-\tau^k)\right)=(\beta-\varepsilon^k)(1-\beta+\varepsilon^k)<\beta(1-\beta+\varepsilon^k).$$

From Bernstein's inequality, we have that

$$\begin{aligned}\Pr(B_k)&=\Pr\left(\hat{F}_N(q^*-\tau^k)\geq\beta+\gamma\right)=\Pr\left(\hat{F}_N(q^*-\tau^k)-F(q^*-\tau^k)\geq\gamma+\varepsilon^k\right)\\&\leq\exp\left(\frac{-N(\gamma+\varepsilon^k)^2}{2F(q^*-\tau^k)(1-F(q^*-\tau^k))+\frac{2}{3}(\gamma+\varepsilon^k)}\right)\\&=\exp\left(\frac{-N(\gamma+\varepsilon^k)}{\frac{2}{(\gamma+\varepsilon^k)}(\beta-\varepsilon^k)(1-\beta-\gamma)+2(\beta-\varepsilon^k)+\frac{2}{3}}\right)\\&\leq\exp\left(\frac{-N(\gamma+\varepsilon^k)}{\frac{2}{(\gamma+\varepsilon^k)}\beta(1-\beta-\gamma)+2\beta+\frac{2}{3}}\right)\leq\exp\left(\frac{-N\gamma}{\frac{2}{\gamma}\beta(1-\beta-\gamma)+2\beta+\frac{2}{3}}\right)\end{aligned}$$

where the inequality follows when $1-\beta-\gamma \geq 0$. Hence, for all $k \geq 1$,

$$\Pr(B_k) \leq \exp\left(\frac{-3N\alpha^2}{24bh + 4\alpha(b+h)}\right).$$

Since $\alpha = C'(\overline{q})$, from (B.4) we have that

$$\Pr(B_k) \leq \exp\left(-\frac{6N\epsilon bh\Delta(q^*)f(q^*) + O(\epsilon^{3/2})}{24bh + O(\epsilon^{1/2})}\right) \triangleq U(\epsilon). \tag{B.5}$$

Now, let us bound the probability of $L \triangleq [C'(\tilde{Q}_N^\alpha) > \alpha] = \left[\bar{F}(\tilde{Q}_N^\alpha) < \frac{h}{b+h} - \frac{\alpha}{b+h}\right]$. Define $q_0 \triangleq \sup\left\{q : \bar{F}(q) \geq \frac{h}{b+h} - \frac{\alpha}{b+h}\right\}$. Thus, $L = [\tilde{Q}_N^\alpha > q_0]$. Note that $\tilde{Q}_N^\alpha = \sup\{q : h - (b+h)\hat{\bar{F}}_N(q) \leq \frac{\alpha}{2}\}$. For a real-valued sequence $\{\tau^k\}_{k=1}^\infty$ where $\tau^k \downarrow 0$, define

$$\begin{aligned} L_k &\triangleq [\tilde{Q}_N^\alpha \geq q_0 + \tau^k] = \left[h - (b+h)\hat{\bar{F}}_N(q_0+\tau^k) \leq \frac{\alpha}{2}\right] \\ &= \left[\hat{\bar{F}}_N(q_0+\tau^k) \geq \frac{h}{b+h} - \frac{1}{2}\frac{\alpha}{b+h}\right] = \left[\hat{\bar{F}}_N(q_0+\tau^k) \geq 1-\beta-\gamma\right]. \end{aligned}$$

Since $\hat{\bar{F}}_N$ is nonincreasing, then it follows that $L_k \subseteq L_{k+1}$. Thus, if $\bar{L}$ is the limiting event of the sequence $\{L_k\}_{k=1}^\infty$, then $L_k \uparrow \bar{L}$, implying that $\Pr(L_k) \uparrow \Pr(\bar{L})$. Note also that $L \subseteq \bar{L}$, implying that $\Pr(L) \leq \Pr(\bar{L})$. Therefore, to prove a bound on $\Pr(L)$, it is sufficient to prove a uniform upper bound on $\Pr(L_k)$.

Note that for some $\epsilon^k > 0$, we have that $\bar{F}(q_0+\tau^k) = 1-\beta-2\gamma-\epsilon^k$. Thus, $L_k = [\hat{\bar{F}}_N(q_0+\tau^k) - \bar{F}(q_0+\tau^k) \geq \gamma + \epsilon^k]$. Finally, from Bernstein's inequality, we have that

$$\begin{aligned} \Pr(L_k) &\leq \exp\left(\frac{-N(\gamma+\epsilon^k)^2}{2\bar{F}(q_0+\tau^k)(1-\bar{F}(q_0+\tau^k)) + \frac{2}{3}(\gamma+\epsilon^k)}\right) \\ &= \exp\left(\frac{-N(\gamma+\epsilon^k)}{\frac{2}{\gamma+\epsilon^k}(1-\beta-2\gamma-\epsilon^k)(\beta+2\gamma+\epsilon^k) + \frac{2}{3}}\right) \\ &= \exp\left(\frac{-N(\gamma+\epsilon^k)}{\frac{2}{\gamma+\epsilon^k}(1-\beta-2\gamma-\epsilon^k)(\beta+\gamma) + 2(1-\beta-2\gamma-\epsilon^k) + \frac{2}{3}}\right) \\ &\leq \exp\left(\frac{-N(\gamma+\epsilon^k)}{\frac{2}{\gamma+\epsilon^k}(1-\beta-2\gamma)(\beta+\gamma) + 2(1-\beta-2\gamma) + \frac{2}{3}}\right) \\ &\leq \exp\left(\frac{-N\gamma}{\frac{2}{\gamma}(1-\beta-2\gamma)(\beta+\gamma) + 2(1-\beta-2\gamma) + \frac{2}{3}}\right) \\ &= \exp\left(\frac{-N\gamma}{\frac{2}{\gamma}\beta(1-\beta-2\gamma) + 4(1-\beta-2\gamma) + \frac{2}{3}}\right). \end{aligned}$$

Therefore, we have that for all $k \geq 1$,

$$\Pr(L_k) \leq \exp\left(\frac{-3N\alpha^2}{24bh + 4\alpha(7h-5b-6\alpha)}\right).$$

Since $\alpha = C'(\overline{q})$, we have from (B.4) that

$$\Pr(L_k) \le \exp\left(-\frac{6N\epsilon bh\Delta(q^*)f(q^*) + O(\epsilon^{3/2})}{24bh + O(\epsilon^{1/2})}\right) \triangleq U(\epsilon). \tag{B.6}$$

Summarizing from (B.5) and (B.6), we have that $\Pr(B) \le \Pr(\bar{B}) \le U(\epsilon)$ and that $\Pr(L) \le \Pr(\bar{L}) \le U(\epsilon)$. Thus,

$$\Pr\left\{\tilde{Q}_N^\alpha < q^* \text{ or } C'(\tilde{Q}_N^\alpha) > C'(\overline{q})\right\} = \Pr(B \cup L) \le \Pr(B) + \Pr(L)$$
$$\le 2U(\epsilon) \sim 2\exp\left(-\frac{1}{4}N\epsilon\Delta(q^*)f(q^*)\right), \text{ as } \epsilon \to 0. \quad \square$$

B.3 Proof of Lemma 2.5.1

Denote by $\partial_- g(x)$ (or $\partial_+ g(x)$) the left-side (or right-side) derivative of a function g at x. The failure rate and reverse hazard rate is given by $\bar{r}(x) = \frac{f(x)}{1-F(x)}$ and $r(x) = \frac{f(x)}{F(x)}$. Since f is a log-concave distribution, it has an increasing failure rate. This implies that $\log \bar{r}(x) = \log f(x) - \log(1 - F(x))$ is increasing, and $\partial_- \log \bar{r}(x) \ge 0$ for all x. Thus,

$$\gamma_1 + \gamma_0 \frac{b+h}{h} = \gamma_1 + \frac{f(q^*)}{1-F(q^*)} \ge \partial_- \log f(q^*) + \frac{f(q^*)}{1-F(q^*)} = \partial_- \log \bar{r}(q^*) \ge 0. \tag{B.7}$$

A log-concave distribution also has a decreasing reversed hazard rate. This implies that $\log r(x) = \log f(x) - \log F(x)$ is decreasing and $\partial_+ \log r(x) \le 0$ for all x. Thus,

$$\gamma_1 - \gamma_0 \frac{b+h}{b} = \gamma_1 - \frac{f(q^*)}{F(q^*)} \le \partial_+ \log f(q^*) - \frac{f(q^*)}{F(q^*)} = \partial_+ \log \bar{r}(q^*) \le 0. \tag{B.8}$$

Combining (B.7) and (B.8), we have that $-\frac{b+h}{h} \le \frac{\gamma_1}{\gamma_0} \le \frac{b+h}{b}$. $\square$

B.4 Proof of Lemma 2.5.2

Note that since $\log f$ is concave, then $\log f(x) \le \log \gamma_0 + \gamma_1(x - t)$, for all x such that $f(x) > 0$. Taking the exponent on both sides proves our result. $\square$

B.5 Proof of Lemma 2.5.3

Note that $\frac{d}{dx}F_1(x) \le \frac{d}{dx}F_2(x)$ by our assumption that $f_1(x) \le f_2(x)$. Moreover, since $F_1(t) = F_2(t)$, then $F_1(x) \ge F_2(x)$ for all $x \le t$ and $F_1(x) \le F_2(x)$ for all $x \ge t$. Note that

$$\begin{aligned} E(D_1 - t | D_1 > t) &= \int_0^\infty \Pr(D_1 > t + s | D_1 > t) ds, \\ &= \tfrac{1}{1-F_1(t)} \int_0^\infty (1 - F_1(t+s)) ds, \\ &\ge \tfrac{1}{1-F_2(t)} \int_0^\infty (1 - F_2(t+s)) ds, \\ &= E(D_2 - t | D_2 > t) \end{aligned}$$

With the same technique, we can also prove that $E(t - D_1 | D_1 \le t) \ge E(t - D_2 | D_2 \le t)$. Combining these results proves the lemma. $\square$

B.6 Proof of Lemma 2.5.5

We first introduce the following notation:

$$G(\alpha) \triangleq \left(\frac{1}{1-\beta}+\alpha\right)\log\left(1+\alpha(1-\beta)\right)+\left(\frac{1}{\beta}-\alpha\right)\log\left(1-\alpha\beta\right)-\min\{\beta,1-\beta\}\alpha^2,$$
$$U(\beta) \triangleq \frac{\beta}{1-\beta}\log\left(\frac{1}{\beta}\right)-\beta,$$
$$L(\beta) \triangleq \frac{1-\beta}{\beta}\log\left(\frac{1}{1-\beta}\right)-(1-\beta).$$

We need to prove that each of the three functions are nonnegative.

1. Let us prove the result for G. First, we prove the result for the case when $\beta \geq \frac{1}{2}$. Note that

$$G'(\alpha) = \log\left(\frac{1+\alpha(1-\beta)}{1-\alpha\beta}\right) - 2(1-\beta)\alpha.$$

The derivative is nonnegative if and only if $G_1(\alpha) \triangleq (1+\alpha(1-\beta))e^{-(1-\beta)\alpha} - (1-\alpha\beta)e^{(1-\beta)\alpha} \geq 0$. Note that for $\alpha \geq 0$,

$$\begin{aligned} G_1'(\alpha) &= -\alpha(1-\beta)^2 e^{-(1-\beta)\alpha} + \beta e^{(1-\beta)\alpha} - (1-\beta)(1-\alpha\beta)e^{(1-\beta)\alpha}, \\ &\geq -\alpha(1-\beta)^2 e^{-(1-\beta)\alpha} + \beta(1-\beta)\alpha e^{(1-\beta)\alpha}, \\ &\geq \alpha(1-\beta)^2\left(e^{(1-\beta)\alpha} - e^{-(1-\beta)\alpha}\right) \geq 0 \end{aligned}$$

Note that $G_1(0) = 0$, thus, $G_1(\alpha) \geq 0$ for all $\alpha \geq 0$. Now define $G_2(\alpha) \triangleq (1+\alpha(1-\beta))e^{-(1-\beta)\alpha} - (1-\alpha(1-\beta))e^{(1-\beta)\alpha}$. Note that $G_2(\alpha) \geq G_1(\alpha)$ if $\alpha \leq 0$. We have

$$G_2'(\alpha) = \alpha(1-\beta)^2\left(e^{(1-\beta)\alpha} - e^{-(1-\beta)\alpha}\right) \geq 0, \quad \text{for } \alpha \leq 0.$$

Note that $G_2(0) = 0$, thus, $G_1(\alpha) \leq G_2(\alpha) \leq 0$ for all $\alpha \leq 0$. Thus, $G(\alpha)$ is nondecreasing in $\alpha \geq 0$, and non-increasing in $\alpha \leq 0$. Since at $\alpha = 0$, this function is zero, then $G(\alpha) \geq 0$ for all α. Now we can also prove the result for $\beta \leq \frac{1}{2}$, if we define the function $\tilde{\beta} = 1-\beta \geq \frac{1}{2}$ and $\tilde{G}(\alpha) = G(-\alpha)$. □

2. Let us prove the result for U. The result is true if and only if $-\log\beta \geq 1-\beta$. Note that $-\log\beta$ is a convex function of β, thus the linear approximation at $\beta = 1$ (i.e., the function $1-\beta$) bounds it from below. □

3. Let us prove the result for L. Defining $\tilde{\beta} = 1-\beta$, note that $L(\beta) = U(\tilde{\beta}) \geq 0$, which follows from (2). □

B.7 Proof of Theorem 2.5.7

Recall that if $q \in S_\epsilon^f \cap [q^*, \infty)$, then $C(q) \leq (1+\epsilon)C(q^*)$. Also, $S_\epsilon^f \cap [q^*, \infty)$ can be equivalently expressed as $\{q : C'(q) \leq C'(\overline{q}) \text{ and } q \geq q^*\}$. Let $\tilde{Q}_N^\alpha$ be defined in (2.8), but with $\alpha = \sqrt{2\epsilon bh \frac{\min\{b,h\}}{b+h}} + O(\epsilon)$. Since,

$$\begin{aligned} C'(\overline{q}) &= (b+h)(F(\overline{q}) - F(q^*)) = (b+h)\left[(\overline{q} - q^*)f(q^*) + O(\overline{q} - q^*)^2\right] \\ &= \sqrt{2\epsilon bh\Delta(q^*)f(q^*)} + O(\epsilon), \end{aligned}$$

then it follows from Proposition 2.5.6 that $\alpha \leq C'(\overline{q})$ when the demand distribution is log-concave. This implies that

$$\left[\tilde{Q}_N^\alpha \geq q^*\right] \cap \left[C'(\tilde{Q}_N^\alpha) \leq \alpha\right] \subseteq \left[\tilde{Q}_N^\alpha \geq q^*\right] \cap \left[C'(\tilde{Q}_N^\alpha) \leq C'(\overline{q})\right].$$

Thus, we only need to derive a lower bound on the probability of the left-hand side event to prove Theorem 2.5.7. Modifying the proof of Theorem 2.4.1 by letting $\alpha = \sqrt{2\epsilon bh \frac{\min\{b,h\}}{b+h}} + O(\epsilon)$, we can prove that

$$\Pr\left(\tilde{Q}_N^\alpha < q^* \text{ or } C'(\tilde{Q}_N^\alpha) > \alpha\right) \leq 2U^*(\epsilon) \sim 2\exp\left(-\frac{1}{4}N\epsilon\frac{\min\{b,h\}}{b+h}\right), \text{ as } \epsilon \to 0. \quad \square$$

B.8 Proof of Proposition 3.4.1

Suppose $\mu - (1-\beta)\delta_L \geq 0$. Consider the two-point support distribution which puts a weight β on $\mu - (1-\beta)\delta_L$ and a weight $1-\beta$ on $\mu + \beta\delta_L$. This distribution is an element of $\mathcal{D}_{\mu,\delta_L,\delta_U,+}$, proving that $\mathcal{D}_{\mu,\delta_L,\delta_U,+}$ is nonempty. To prove the reverse implication, suppose that $\mathcal{D}_{\mu,\delta_L,\delta_U,+}$ is nonempty. Let F be a distribution in $\mathcal{D}_{\mu,\delta_L,\delta_U,+}$ where $F^{-1}(\beta) = w$ for some $w \geq 0$. From the arguments in the proof of Proposition 3.2.1, we have that $E(D|D \leq w) = \mu - (1-\beta)\Delta_F(F^{-1}(\beta))$. Since D is nonnegative, we have that $0 \leq \mu - (1-\beta)\Delta_F(F^{-1}(\beta)) \leq \mu - (1-\beta)\delta_L$. $\square$

B.9 Proof of Theorem 3.2.2

Consider the distribution set

$$\mathcal{D}_{\mu,\delta} = \left\{F \ : \ E_F(\mathbb{1}_{[-\infty,\infty]}(D)) = 1, E_F(D) = \mu \text{ and } \Delta(F^{-1}(\beta)) = \delta\right\}.$$

We also define the following constrained distribution set

$$\mathcal{D}_{w,\mu,\delta} = \left\{F \ : \ E_F(\mathbb{1}_{[-\infty,w]}(D)) = \beta, E_F(\mathbb{1}_{[w,\infty]}(D)) = 1-\beta, E_F(D) = \mu \text{ and } \Delta(w) = \delta\right\}.$$

We note that $\mathcal{D}_{w,\mu,\delta}$ includes all distributions $F \in \mathcal{D}_{\mu,\delta}$ with $F^{-1}(\beta) \leq w$ and $F^{-1}(\beta+\epsilon) \geq w$ for any $\epsilon > 0$.

Before proceeding with the proof, we require the following lemma which provides an interval of valid values for w.

Lemma B.9.1. *Suppose a distribution F has mean μ and AMS (at the β quantile) δ. Then,*

$$F^{-1}(\beta) \in [\mu - (1-\beta)\delta, \mu + \beta\delta].$$

Proof. Suppose that $F^{-1}(\beta) = w$. To prove the lower bound, note that $w \geq E(D|D \leq w) = \mu - (1-\beta)\delta$, where the latter equality is established (B.2) in the proof of Proposition 3.2.1. Now to prove the upper bound, we have that

$$\begin{aligned}\mu - w &= E(D-w)^+ - E(w-D)^+ \\ &= (1-\beta)E(D-w|D \geq w) - \beta E(w-D|D \leq w) \\ &= -\beta\left\{E(D|D \geq w) - E(D|D \leq w)\right\} + E(D-w|D \geq w) \\ &= -\beta\delta + E(D-w|D \geq w) \geq -\beta\delta,\end{aligned}$$

which establishes that $w \leq \mu + \beta\delta$. □

By changing the order of maximization, we can rewrite the maximum regret under $\mathcal{D}_{w,\mu,\delta}$ as

$$\rho(y) \triangleq \sup_{F \in \mathcal{D}_{w,\mu,\delta}} \left(\max_{z \geq 0} \Pi_F(z) - \Pi_F(y) \right) = \max_{z \geq 0} G(z;y)$$

where

$$G(z;y) \triangleq \sup_{F \in \mathcal{D}_{w,\mu,\delta}} \int_0^{\infty} (\min\{x,z\} - \min\{x,y\})\, dF(x) + (1-\beta)(y-z). \tag{B.9}$$

To find a closed form expression for the minimax regret ρ^*, we first need to solve the inner moment problem $G(z;y)$ for fixed (z,y). Although $G(z;y)$ is not necessarily concave on $\mathbb{R}$, Perakis & Roels (2008) prove that it is concave on $z \in (-\infty, y]$ and on $z \in [y, \infty)$. Therefore, $G^-(y) \triangleq \max_{z \in [0,y]} G(z;y)$ and $G^+(y) \triangleq \max_{z \in [y,\infty)}$ can be efficiently solved. Thus,

$$\rho(y) = \max\left\{G^-(y), G^+(y)\right\}.$$

Thus, to prove Theorem 3.2.2, we need to solve

$$\begin{aligned}(P) \quad \sup_f \quad & \int_{-\infty}^{\infty} (\min\{x,z\} - \min\{x,y\})\, f(x)dx, \\ \text{s.t.} \quad & \int_{-\infty}^{w} f(x)dx = \beta, \quad \int_{w}^{\infty} f(x)dx = 1-\beta, \\ & \int_{-\infty}^{\infty} xf(x)dx = \mu, \\ & \frac{1}{1-\beta}\int_{w}^{\infty} xf(x)dx - \frac{1}{\beta}\int_{-\infty}^{w} xf(x)dx = \delta, \\ & f(x) \geq, \ \forall x \in \mathbb{R}.\end{aligned}$$

The dual of the above moment problem is

$$
\begin{aligned}
(D) \qquad \min_{\alpha_0^L, \alpha_0^U, \alpha_1, \alpha_2} \quad & \alpha_0^L \beta + \alpha_0^U (1-\beta) + \alpha_1 \mu + \alpha_2 \delta, \\
\text{s.t.} \quad & \alpha_0^L + \left(\alpha_1 - \frac{1}{\beta}\alpha_2 \right) x \geq \min\{x, z\} - \min\{x, y\}, \quad \forall x \in (-\infty, w], \\
& \alpha_0^U + \left(\alpha_1 + \frac{1}{1-\beta}\alpha_2 \right) x \geq \min\{x, z\} - \min\{x, y\}, \quad \forall x \in [w, \infty).
\end{aligned}
$$

By weak duality, we have that the optimal primal cost is always less than or equal to the optimal dual cost. In fact, we prove that the primal and dual optimal costs are equal (i.e, there is no duality gap). We do this by constructing primal and dual feasible solutions that have equal primal and dual objective costs. Let $S(w; z, y)$ be the optimal primal and dual cost. Tables D.8–D.13 construct these optimal solutions for different cases of y, z, w.

Note that

$$
G(z; y) = \max_{w \in [\mu - (1-\beta)\delta, \mu + \beta\delta]} S(w; z, y) + (1-\beta)(y - z),
$$

where from Lemma B.9.1, we need only consider w that belong in the range $[\mu - (1-\beta)\delta, \mu + \beta\delta]$. Table D.14 summarizes the closed form expression for $G(z; y)$ based on the primal optimal cost. From this, we have that

$$
G^-(y) = \begin{cases} 0, & \text{for } y \in (-\infty, \mu - (1-\beta)\delta], \\ (1-\beta)(y - \mu + (1-\beta)\delta), & \text{for } y \in [\mu - (1-\beta)\delta, \infty), \end{cases}
$$

$$
G^+(y) = \begin{cases} \beta(\mu + \beta\delta - y), & \text{for } y \in (-\infty, \mu + \beta\delta], \\ 0, & \text{for } y \in [\mu + \beta\delta, \infty). \end{cases}
$$

Finally, we have that the maximum regret is

$$
\rho(y) = \max\{G^-(y), G^+(y)\} = \begin{cases} \beta(\mu + \beta\delta - y), & \text{for } y \in (-\infty, \mu + (2\beta - 1)\delta], \\ (1-\beta)(y - \mu + (1-\beta)\delta), & \text{for } y \in (\mu + (2\beta - 1)\delta, \infty), \end{cases}
$$

which is minimized at $y^* = \mu + (2\beta - 1)\delta$ with a minimax regret $\rho^* = \beta(1-\beta)\delta$. □

B.10 Proofs for Theorems 3.2.3 and A.3.1

Since Theorem 3.2.3 is a special case of Theorem A.3.1 (by letting $\delta = \delta_L = \delta_U$), we only prove the latter theorem. Consider the distribution set

$$
\mathcal{D}_{\mu, \delta_L, \delta_U, +} = \left\{ F \ : \ E_F(\mathbb{1}_{[0,\infty]}(D)) = 1, E_F(D) = \mu \text{ and } \Delta(F^{-1}(\beta)) \in [\delta_L, \delta_U] \right\}.
$$

Consider the constrained distribution set

$$
\mathcal{D}_{w, \mu, \delta_L, \delta_U, +} = \left\{ F \ : \ E_F(\mathbb{1}_{[0,w]}(D)) = \beta, E_F(\mathbb{1}_{[w,\infty]}(D)) = 1 - \beta, E_F(D) = \mu \text{ and } \Delta(w) \in [\delta_L, \delta_U] \right\},
$$

which includes all distributions $F \in \mathcal{D}_{\mu,\delta_L,\delta_U,+}$ such that $F^{-1}(\beta) \leq w$ and $F^{-1}(\beta+\epsilon) \geq w$ for any $\epsilon > 0$. The following lemma provides an interval of valid values for w.

Lemma B.10.1. *Suppose a nonnegative distribution F has mean μ and AMS (at the β quantile) in the range $[\delta_L, \delta_U]$. Then,*

$$F^{-1}(\beta) \in \left[\max\{0, \mu - (1-\beta)\delta_U\}, \min\left\{\mu + \beta\delta_U, \frac{\mu}{1-\beta}\right\}\right].$$

Proof. Proof. Let $F \in \mathcal{D}_{\mu,\delta_L,\delta_U,+}$ such that $F^{-1}(\beta) = w$ for some $w \geq 0$. The proof that $w \geq \mu - (1-\beta)\delta_U$ and $w \leq \mu + \beta\delta_U$ follows in a manner similar to the proof of Lemma B.9.1. Moreover, from nonnegativity of D it follows that $w \geq 0$. We are left to prove $w \leq \frac{\mu}{1-\beta}$. Due to nonnegativity, we have that

$$\mu = \int_0^\infty t dF(t) \geq \int_w^\infty t dF(t) \geq w \int_w^\infty dF(t) = w(1-\beta).$$

□

By changing the order of maximization, we can rewrite the maximum regret under $\mathcal{D}_{\mu,\delta_L,\delta_U,+}$ as

$$\rho(y) \triangleq \sup_{F \in \mathcal{D}_{\mu,\delta_L,\delta_U,+}} \left(\max_{z \geq 0} \Pi_F(z) - \Pi_F(y) \right) = \max_{z \geq 0} G(z; y)$$

where

$$G(z; y) \triangleq \sup_{F \in \mathcal{D}_{\mu,\delta_L,\delta_U,+}} \int_0^\infty (\min\{x, z\} - \min\{x, y\})\, dF(x) + (1-\beta)(y - z). \tag{B.10}$$

To find a closed form expression for the minimax regret ρ^*, we first need to solve the inner moment problem $G(z; y)$ for fixed (z, y). Although $G(z; y)$ is not necessarily concave on $\mathbb{R}$, Perakis & Roels (2008) prove that it is concave on $z \in (-\infty, y]$ and on $z \in [y, \infty)$. Therefore, $G^-(y) \triangleq \max_{z \in [0,y]} G(z; y)$ and $G^+(y) \triangleq \max_{z \in [y,\infty)}$ can be efficiently solved. Thus,

$$\rho(y) = \max\left\{G^-(y), G^+(y)\right\}.$$

Consider the following moment problem

$$
\begin{aligned}
(P) \quad \sup_{f} \quad & \int_0^\infty \left(\min\{x,z\} - \min\{x,y\}\right) f(x)dx \\
\text{s.t.} \quad & \int_0^w f(x)dx = \beta, \quad \int_w^\infty f(x)dx = 1-\beta, \\
& \delta_L \le \frac{1}{1-\beta}\int_w^\infty xf(x)dx - \frac{1}{\beta}\int_0^w xf(x)dx \le \delta_U, \\
& \int_0^\infty xf(x)dx = \mu, \quad f(x) \ge 0 \quad \forall x \ge 0.
\end{aligned}
$$

The dual of the above moment problem is:

$$
\begin{aligned}
(D) \quad \min_{\alpha_0^L,\alpha_0^U,\alpha_1,\alpha_2^L,\alpha_2^U} \quad & \alpha_0^L\beta + \alpha_0^U(1-\beta) + \alpha_1\mu + \alpha_2^L\delta^L + \alpha_2^U\delta^U \\
\text{s.t.} \quad & \alpha_0^L + \left(\alpha_1 - \frac{1}{\beta}(\alpha_2^L + \alpha_2^U)\right)x \ge \min\{x,z\} - \min\{x,y\}, \quad \forall x \in [0,w), \\
& \alpha_0^U + \left(\alpha_1 + \frac{1}{1-\beta}(\alpha_2^L + \alpha_2^U)\right)x \ge \min\{x,z\} - \min\{x,y\}, \quad \forall x \in [w,\infty), \\
& \alpha_2^L \le 0, \ \alpha_2^U \ge 0.
\end{aligned}
$$

If we prove that there is a primal feasible distribution and dual feasible solution that both achieve the same cost, then by weak duality, they are primal and dual optimal, respectively. Tables D.15–D.20 construct these optimal solutions for different cases of w, z, y. We denote by $S(w; z, y)$ the optimal primal and dual cost.

Note that $G(z;y) = \max_{w\in\mathcal{W}} S(w;z,y) + (1-\beta)(y-z)$, where $\mathcal{W} \triangleq [\mu-(1-\beta)\delta_U, \mu+\beta\delta_U]$ if $\mu \ge (1-\beta)\delta_U$ and $\mathcal{W} \triangleq [0, \frac{\mu}{1-\beta}]$, otherwise. It is straightforward to compute this function due to the existence of a closed form for $S(w;z,y)$. Tables D.21 and D.22 summarize the values for $G(z;y)$ under these cases. We have that when $\mu - (1-\beta)\delta_U \le 0$,

$$
G^-(y) = (1-\beta)y, \quad \text{for } y \in [0,\infty)
$$

$$
G^+(y) = \begin{cases}
\frac{\beta(\mu-(1-\beta)\delta_L)(\mu+\beta\delta_L-y)}{\mu+\beta\delta_L}, & \text{for } y \in \left[0, \frac{(1-\beta)}{\mu}(\mu+\beta\delta_L)^2\right], \\
(\sqrt{\mu} - \sqrt{(1-\beta)y})^2, & \text{for } y \in \left[\frac{(1-\beta)}{\mu}(\mu+\beta\delta_L)^2, \frac{\mu}{1-\beta}\right], \\
0, & \text{for } y \in \left[\frac{\mu}{1-\beta}, \infty\right),
\end{cases}
$$

and when $\mu - (1-\beta)\delta_U \ge 0$,

$$
G^-(y) = \begin{cases}
0, & \text{for } y \in [0, \mu-(1-\beta)\delta_U], \\
(1-\beta)(y-\mu+(1-\beta)\delta_U), & \text{for } y \in [\mu-(1-\beta)\delta_U, \infty),
\end{cases}
$$

$$
G^+(y) = \begin{cases}
\frac{\beta(\mu-(1-\beta)\delta_L)(\mu+\beta\delta_L-y)}{\mu+\beta\delta_L}, & \text{for } y \in \left[0, \frac{(1-\beta)}{\mu}(\mu+\beta\delta_L)^2\right], \\
(\sqrt{\mu} - \sqrt{(1-\beta)y})^2, & \text{for } y \in \left[\frac{(1-\beta)}{\mu}(\mu+\beta\delta_L)^2, \frac{(1-\beta)}{\mu}(\mu+\beta\delta_U)^2\right], \\
\frac{\beta(\mu-(1-\beta)\delta_U)(\mu+\beta\delta_U-y)}{\mu+\beta\delta_U}, & \text{for } y \in \left[\frac{(1-\beta)}{\mu}(\mu+\beta\delta_U)^2, \mu+\beta\delta_U\right] \\
0, & \text{for } y \in [\mu+\beta\delta_U, \infty).
\end{cases}
$$

Finally, solving for $\rho(y) = \max\{G^-(y), G^+(y)\}$ and y^* gives us Theorem A.3.1.

Now we can proceed with a proof of Theorem 3.2.3. When the spread is exactly equal to δ, i.e. $\delta = \delta_L = \delta_U$, then functions G^- and G^+ reduce to

$$G^-(y) = \begin{cases} 0, & \text{for } y \in [0, \mu - (1-\beta)\delta], \\ (1-\beta)(y - \mu + (1-\beta)\delta), & \text{for } y \in [\mu - (1-\beta)\delta, \infty), \end{cases}$$

$$G^+(y) = \begin{cases} \frac{\beta(\mu-(1-\beta)\delta)(\mu+\beta\delta-y)}{\mu+\beta\delta}, & \text{for } y \in [0, \mu + \beta\delta] \\ 0, & \text{for } y \in [\mu + \beta\delta, \infty) \,. \end{cases}$$

For any $y \geq 0$, the maximum regret is the convex function $\rho(y) = \max\{G^-(y), G^+(y)\}$, i.e.,

$$\rho(y) = \begin{cases} \frac{\beta(\mu-(1-\beta)\delta)(\mu+\beta\delta-y)}{\mu+\beta\delta}, & \text{for } y \in \left[0, \frac{1}{\mu}(\mu - (1-\beta)\delta)(\mu+\beta\delta)\right] \\ (1-\beta)(y - \mu + (1-\beta)\delta), & \text{for } y \in \left[\frac{1}{\mu}(\mu - (1-\beta)\delta)(\mu+\beta\delta), \infty\right) \end{cases}$$

The quantity y^* that minimizes ρ occurs at the breakpoint of the piecewise linear function at $\frac{1}{\mu}(\mu - (1-\beta)\delta)(\mu + \beta\delta)$. □

B.11 Proof of Theorem 4.4.2

Let X^{LP} and C^{LP} be the optimal solution for the LP relaxation (A.7) of the job scheduling problem. To prove Theorem 4.4.2, we first require proving the following proposition.

Proposition B.11.1. *Let X^{LP} be the optimal solution to the LP relaxation* (A.7). *Define the randomized rounding $\tilde{X}$ such that for each $i = 1, \ldots, n$, randomly round exactly one of the indices $\{1, 2, \ldots, T\}$ to 1, with index t chosen with probability X_{it}^{LP}. Then with* positive probability*,*

$$\max_{t=1,\ldots,T} \frac{1}{K_t} \left(d_L E_t[L(\omega)] + \sum_{i=1}^{n} d_i \tilde{X}_{it} \right) \leq C^{LP} \left(1 + H\left(C^{LP}, \frac{1}{e\delta} \right) \right),$$

where

$$H(w,p) \triangleq \frac{1}{w} \left(\min_{s=1,\ldots,T} K_s \right)^{-1} \sqrt{\frac{1}{2} \left(\sum_{i=1}^{n} d_i^2 \right) \ln\left(\frac{1}{p}\right)}. \tag{B.11}$$

Proof. For a given t, define $\tilde{X}_t = (\tilde{X}_{1t}, \tilde{X}_{2t}, \ldots, \tilde{X}_{nt})$. Moreover, define the function $f_t : [0,1]^n \mapsto \mathbb{R}$ as

$$f_t(x_1, x_2, \ldots, x_n) \triangleq \frac{1}{K_t} \left(d_L E_t[L(\omega)] + \sum_{i=1}^{n} d_i x_i \right).$$

That is $f_t(\tilde{X}_t)$ is a function of a random variable which represents the ratio of expected hours scheduled on day t to the number of crews on day t under the randomly rounded solution.

Define the "bad" event B_t as the event that the random schedule results in a ratio of scheduled hours to number of crews exceeding the bound in Proposition B.11.1, i.e.,

$$B_t \triangleq \left[f_t(\tilde{X}_t) > C^{LP} \left(1 + H \left(C^{LP}, \frac{1}{e\delta} \right) \right) \right].$$

Therefore, proving Proposition B.11.1 is equivalent to proving

$$0 < \Pr \left(\max_{t=1,\dots,T} f_t(\tilde{X}_t) \leq C^{LP} \left(1 + H \left(C^{LP}, \frac{1}{e\delta} \right) \right) \right) = \Pr \left(\bigcap_{t=1}^{T} \overline{B_t} \right). \tag{B.12}$$

Since there is limited dependency among the "bad events" (i.e., each event B_t is mutually dependent on at most $\delta - 1$ other events), then we can use Lovász's Local Lemma to prove (B.12).

Lemma B.11.2 (Lovász's Local Lemma)**.** *Let $B_1, \dots, B_m$ be a set of events with $\Pr(B_i) \leq p < 1$ and each event B_i is mutually of all but at most s of the other B_j. If $e \cdot p(s+1) \leq 1$, then* $\Pr \left(\bigcap_{i=1}^{m} \overline{B_i} \right) > 0$.

Thus, to use Lovász's Local Lemma, we need to find a bound p such that $\Pr(B_t) \leq p$ and $e \cdot p(s+1) \leq 1$. Since f_t a function of bounded differences, and $\tilde{X}_t$ are independent random variables, we use a large deviations bound (McDiarmid's inequality) to derive a bound on $\Pr(B_t)$.

Lemma B.11.3 (McDiarmid's inequality)**.** *Let $X_1, X_2, \dots, X_m$ be independent random variables all taking values in the set $\mathcal{X}$. Further, let $f : \mathcal{X}^m \mapsto \mathbb{R}$ be a function of $X_1, \dots, X_m$ that satisfies $\forall i, \forall x_1, \dots, x_m, x_i' \in \mathcal{X}$,*

$$|f(x_1, \dots, x_i, \dots, x_m) - f(x_1, \dots, x_{i-1}, \hat{x}_i, x_{i+1}, \dots, x_m)| \leq c_i. \tag{B.13}$$

Then for any $\epsilon > 0$, $\Pr\left(f - E[f] \geq \epsilon\right) \leq \exp\left(\frac{-2\epsilon^2}{\sum_{i=1}^{m} c_i^2} \right)$.

It is easy to verify that f_t satisfies condition (B.13) in McDiarmid's inequality, with $c_i = d_i / K_t$. Note that we can bound $E[f_t(\tilde{X}_t)]$, since

$$E[f_t(\tilde{X}_t)] = \frac{1}{K_t} \left(d_L E_t[L(\omega)] + \sum_{i=1}^{n} d_i E[\tilde{X}_{it}] \right) = \frac{1}{K_t} \left(d_L E_t[L(\omega)] + \sum_{i=1}^{n} d_i X_{it}^{LP} \right) \leq C^{LP},$$

where the last inequality follows since C^{LP} and X^{LP} are feasible for the LP relaxation (A.7).

Therefore, we have that

$$\Pr\left\{f_t(\tilde{X}_t) \geq C^{LP}\left(1 + H\left(C^{LP}, \frac{1}{e\delta}\right)\right)\right\} \leq \Pr\left\{f_t(\tilde{X}_t) - E[f_t(\tilde{X}_t)] \geq C^{LP} H\left(C^{LP}, \frac{1}{e\delta}\right)\right\} \tag{B.14}$$

$$\leq \exp\left(\frac{-2\left(K_t C^{LP} H\left(C^{LP}, \frac{1}{e\delta}\right)\right)^2}{\sum_i d_i^2}\right), \tag{B.15}$$

$$= \exp\left(-\ln(e\delta)\left(\frac{K_t}{\min_s K_s}\right)^2\right), \tag{B.16}$$

$$\leq \exp(-\ln(e\delta)) = \frac{1}{e\delta} \tag{B.17}$$

where inequality (B.14) follows from $E[f_t] \leq C^{LP}$, and inequality (B.15) follows from McDiarmid's inequality with $\epsilon = C^{LP} H\left(C^{LP}, \frac{1}{e\delta}\right)$.

Therefore, since $\Pr(B_t) \leq \frac{1}{e\delta}$, and each event B_t is mutually dependent on at most $\delta - 1$ other events, the conditions of Lovász's Local Lemma are met, proving (B.12) and Proposition B.11.1. □

Note that all realizations of $\tilde{X}$ are all the roundings of the LP solution X^{LP} into a feasible job schedule. Since out of all roundings, X^H produced by Algorithm LP-schedule has the smallest value for the maximum ratio of scheduled hours to number of crews

$$\max_{t=1,\ldots,T} \frac{1}{K_t}\left(d_L E_t[L(\omega)] + \sum_{i=1}^{n} d_i X_{it}^H\right),$$

then by Proposition B.11.1, we have found a deterministic rounding X^H for which the maximum threshold for worst-case scheduled hours per crew is at most $C^{LP}\left(1 + H\left(C^{LP}, \frac{1}{e\cdot\delta}\right)\right) \leq C^{OPT}\left(1 + H\left(C^{LP}, \frac{1}{e\cdot\delta}\right)\right)$. □

B.12 Proof of Proposition 4.5.1

We will use the following lemma to prove Proposition 4.5.1.

Lemma B.12.1. *Let $L(\omega_1) < L(\omega_2)$ for some $\omega_1, \omega_2 \in \Omega$. Then for any optimal solution $(Y^*, Z^*(\omega), \omega \in \Omega)$ to the stochastic assignment problem, we have that*

$$\max_{k=1,\ldots,K}\left\{d_L Z_k^*(\omega_1) + \sum_{i\in I} d_i Y_{ik}^*\right\} \leq \max_{k=1,\ldots,K}\left\{d_L Z_k^*(\omega_2) + \sum_{i\in I} d_i Y_{ik}^*\right\}. \tag{B.18}$$

Proof. We will prove the lemma by contradiction. Suppose the converse is true, that is:

$$\max_{k=1,\ldots,K}\left\{d_L Z_k^*(\omega_1) + \sum_{i\in I} d_i Y_{ik}^*\right\} > \max_{k=1,\ldots,K}\left\{d_L Z_k^*(\omega_2) + \sum_{i\in I} d_i Y_{ik}^*\right\}. \tag{B.19}$$

We will show that we can define a new leak assignment for $\tilde{Z}$, which achieves a strictly smaller maximum work hours. Let $\tilde{Z}(\omega) = Z^*(\omega)$ for all $\omega \neq \omega_1$. Now, choose an arbitrary crew $k_0 \in \{1, \ldots, K\}$. Let $\tilde{Z}_k(\omega_1) = Z_k^*(\omega_2)$ for any $k \neq k_0$, and $\tilde{Z}_{k_0}(\omega_1) = Z_{k_0}^*(\omega_2) - (L(\omega_2) - L(\omega_1))$. Since $\sum_{k=1}^K \tilde{Z}_k(\omega) = L(\omega)$ for all $\omega \in \Omega$, then $\tilde{Z}$ is a feasible leak assignment. Moreover, by construction,

$$\max_{k=1,\ldots,K} \left\{ d_L \tilde{Z}_k(\omega_1) + \sum_{i \in I} d_i Y_{ik}^* \right\} \leq \max_{k=1,\ldots,K} \left\{ d_L Z_k^*(\omega_2) + \sum_{i \in I} d_i Y_{ik}^* \right\}.$$

And by (B.19), $(Y^*, \tilde{Z}(\omega), \omega \in \Omega)$ has a strictly smaller maximum work hours than $(Y^*, Z^*(\omega), \omega \in \Omega)$, violating the optimality of $(Y^*, Z^*(\omega), \omega \in \Omega)$. □

Now let us prove Proposition 4.5.1. Suppose that $Z_{k_0}^*(\omega_1) > Z_{k_0}^*(\omega_2)$ for some $k_0 \in \{1, \ldots, K\}$. We will construct a gas leak assignment $\tilde{Z}(\omega_2)$ for scenario ω_2 which has maximum hours (makespan) no greater than that of $Z^*(\omega_2)$. First, note that since there are less gas leak jobs in scenario ω_1, inequality (B.18) holds due to Lemma B.12.1.

Define $\tilde{Z}(\omega_2)$, a new gas leak assignment for ω_2, by letting $\tilde{Z}_{k_0}(\omega_2) = Z_{k_0}^*(\omega_2) + 1$, $\tilde{Z}_{k_1}(\omega_2) = Z_{k_1}^*(\omega_2) - 1$ (where k_1 is some crew in $\{1, \ldots, K\}$ with $Z_{k_1}^*(\omega_2) > 0$), and $\tilde{Z}_k(\omega_2) = Z_k^*(\omega_2)$ for all $k \in \{1, \ldots, K\} \setminus \{k_0, k_1\}$. Note that the assigned work hours (load) of crew k_1 is strictly smaller under this new assignment. Now all that is left to prove is that the load of crew k_0 is smaller than the maximum load in assignment $Z^*(\omega_2)$. Note that since $\tilde{Z}_{k_0}(\omega_2) \leq Z_{k_0}^*(\omega_1)$, the load of crew k_0 in assignment $\tilde{Z}(\omega_2)$ under scenario ω_2 is no greater than its load in assignment $Z^*(\omega_1)$ under scenario ω_1. Inequality (B.18) implies that the load of k_0 under both scenarios is no greater than the maximum load of the assignment $Z^*(\omega_2)$ under scenario ω_2. Therefore, the load of crew k_0 does not increase the maximum load beyond the makespan of assignment $Z^*(\omega_2)$. □

B.13 Proof of Proposition 4.5.2

Without loss of generality, let $L(\omega_j) = j$ for $j = 1, \ldots, m$. Label the crews using the following procedure. Let A be the set of labeled crews, which is initialized to be $\emptyset$. Starting with $j = 1$, scan the solution $Z(\omega_j)$ for leak assignments with $Z_k(\omega_j) > 0$. If $k \notin A$, let $k_j = k$. If all crews with positive leak assignments are in A, move on to the next leak scenario, $j = 2$, scanning for leak assignments with $Z_k(\omega_j) > 0$. Perform the labeling procedure for each scenario, until all the scenarios are exhausted. If at the end of the procedure, the number of labeled crews is less than the total number of crews, label the rest of the crews arbitrarily with the remaining labels. This labeling procedure results with labels $k_1, k_2, \ldots, k_K$ with $Z_{k_{j-1}}(\omega) \geq Z_{k_j}(\omega)$ for all $\omega \in \Omega$. □

B.14 Proof of Corollary 4.5.3

From Proposition 4.5.1, there exists an optimal solution that satisfies the condition for Proposition 4.5.2. Hence, from Proposition 4.5.2, there exists a ranking of the crews such that $Z^*_{k_{j-1}}(\omega) \geq Z^*_{k_j}(\omega)$ for all $\omega \in \Omega$. Next, we prove that with this labeling, $\sum_{i\in I} d_i Y^*_{ik_{j-1}} \leq \sum_{i\in I} d_i Y^*_{ik_j}$. We prove this contradiction. Suppose that for a pair of crews k_{j-1}, k_j, we have that $\sum_{i\in I} d_i Y^*_{ik_{j-1}} > \sum_{i\in I} d_i Y^*_{ik_j}$. Then, we can find an assignment $\tilde{Y}$ which has a makespan no greater than Y^*, by letting $\tilde{Y}_{ik_{j-1}} = Y^*_{ik_j}$ and $\tilde{Y}_{ik_j} = Y^*_{ik_{j-1}}$. This violates the optimality of Y^*. □

B.15 Proof of Proposition 4.5.6

Define $D^k = kd$ as the cumulative duration of the first k jobs in Algorithm Stoch-LPT. Denote by y^k the total standard job hours assigned by the algorithm to crew B at the end of the kth iteration. At the kth iteration of the algorithm, the cost function $F^k(y) = p\max(D^k - y, y) + (1-p)\max(d_L + D^k - y, y)$ for $y \in \{y^{k-1}, y^{k-1} + d\}$. It chooses y^k to be the quantity which gives the smaller value for F^k.

Let us prove the first statement of Proposition 4.5.6. Define $m = \lfloor \frac{d_L}{d} \rfloor$. It is easy to verify that for $k \leq m$, F^k is a decreasing function. Therefore, Algorithm Stoch-LPT sets $y^k = kd$ for $k \leq m$. That is, Stoch-LPT assigns the first m standard jobs to crew B.

Now let us consider $k > m$. For easy reference later, note that

$$F^k(y) = \begin{cases} kd - y + (1-p)md, & \text{if } y \in [0, kd/2], \\ (2p-1)y + (1-p)(m+k)d, & \text{if } y \in (kd/2, (m+k)d/2], \\ y, & \text{if } y \in [(m+k)d/2, kd]. \end{cases}$$

We would like to prove the following lemma, which if we are able to prove, is equivalent to saying that Stoch-LPT assigns the remaining jobs alternatively between crew A and crew B.

Lemma B.15.1. *For $p \leq \frac{1}{2}$, the Stoch-LPT algorithm produces a series of crew B assignments such that*

$$\begin{cases} y^{m+2k} = (m+k)d, & \text{for } k = 0, 1, 2, \dots \\ y^{m+2k+1} = (m+k)d, & \text{for } k = 0, 1, 2, \dots \end{cases} \tag{B.20}$$

Proof. Let us prove this by induction. First, we check it for $k = 0$. Note that $y^m = md$. Next, we need to check y^{m+1}. It is easy to verify that if $m \geq 1$, then $F^{m+1}(y^m) = (2p-1)y^m + (1-p)(2m+1)d = md + (1-p)d$. Otherwise, $F^{m+1}(y^m) = d + (1-p)md$. Since $F^{m+1}(y^m + d) = (m+1)d$, then $F^{m+1}(y^m) \leq F^{m+1}(y^m + d)$. Therefore, Algorithm Stoch-LPT will choose $y^{m+1} = md$. Therefore, the statement is true for $k = 0$.

Now suppose the statement is true for $k = 0, 1, \dots, s-1$. Let us show that it is also true for $k = s$. Let us check that $y^{m+2s} = y^{m+2s-1} + d = (m+s)d$. It is easy to verify that F^{m+2s} is decreasing for $y \leq (m+s)d$. Therefore, $F^{m+2s}(y^{m+2s-1}) \geq F^{m+2s}(y^{m+2s-1} + d)$, implying that $y^{m+2s} = (m+s)d$. Now let us check that $y^{m+2s+1} = (m+s)d$. It can be verified that, if $m \geq 1$, then $F^{m+2s+1}(y^{m+2s}) = (m+s)d + (1-p)d$. Otherwise, $F^{m+2s+1}(y^{m+2s}) = (s+1)d + (1-p)md$. Moreover, $F^{m+2s+1}(y^{m+2s} + d) = (m+s+1)d$.

Since $F^{m+2s+1}(y^{m+2s}) \leq F^{m+2s+1}(y^{m+2s} + d)$, then $y^{m+2s+1} = y^{m+2s} = (m + s)d$.

Thus, we prove the lemma by induction. $\square$

Now let us prove the second statement of Proposition 4.5.6. We first prove the following lemma, which if we are able to prove, is equivalent to saying that Stoch-LPT assigns the jobs alternatively between crew A and crew B.

Lemma B.15.2. *For $p > \frac{1}{2}$, the Stoch-LPT algorithm produces a series of crew B assignments such that*

$$\begin{cases} y^{2k-1} = kd, & \textit{for } k = 1, 2, 3, \ldots \\ y^{2k} = kd, & \textit{for } k = 1, 2, 3, \ldots \end{cases}$$

Proof. We prove the lemma by induction. First, let us check the condition when $k = 1$. Note that $F^1(0) = d+(1-p)md$, and that $F^1(d) = d$ if $m < 1$ or $F^1(d) = (2p-1)d+(1-p)(m+1)d$ if $m \geq 1$. In both cases, we have that $F^1(d) \leq F^1(0)$, implying that $y^1 = d$. Now let us check the condition for y^2. Note that F^2 is increasing in $y \geq y_1 = d$. Therefore, $y^2 = d$.

Now suppose that the statement is true for $k = 1, 2, \ldots, s-1$. We will show that it is true for s. We have to check that $F^{2s-1}(y)$ for $y = y^{2(s-1)}$ and $y = y^{2(s-1)} + d$. Note that $F^{2s-1}((s-1)d) = sd + (1-p)md$. Moreover, for $m \geq 1$, we have that $F^{2s-1}(sd) = sd + (1-p)md - (1-p)d$. If $m < 1$, we have that $F^{2s-1}(sd) = sd$. Therefore, $y^{2s-1} = sd$. Now let us check the condition for y^{2s}. Since F^{2s} is increasing for $y \geq sd$, then $y^{2s} = sd$. $\square$

Now we will prove the third statement of Proposition 4.5.6. Let us consider the case when $p \leq \frac{1}{2}$. Recall (B.20) which gives the sequence of crew B assignments under Stoch-LPT. Assume without loss of generality that the number of jobs is more than m. Note that at the $m+2k$ iteration, Stoch-LPT evaluates F^{m+2k} at $y = (m+k-1)d$ and $y = (m+k)d$. The unconstrained minimizer of F^{m+2k} is $y = (m+k)d$, which is between these two values. Therefore, there is no multiple of d which achieves a smaller value for F^{m+2k} than $y = y^{m+2k}$.

Similarly, at the $m+2k-1$ iteration, Stoch-LPT evaluates F^{m+2k+1} at $y = (m+k)d$ and $y = (m+k+1)d$. The unconstrained minimizer of F^{m+2k+1} is $y = (m+k+1/2)d$, which is between these two values. Therefore, there is no multiple of d which achieves a smaller value for F^{m+2k+1} than $y = y^{m+2k+1}$.

The proof for $p > \frac{1}{2}$ follows a similar line of argument. $\square$

Appendix C

Figures

Figure C-1: Upper bound for a log-concave distribution with $\frac{b}{b+h}$ quantile q^*.

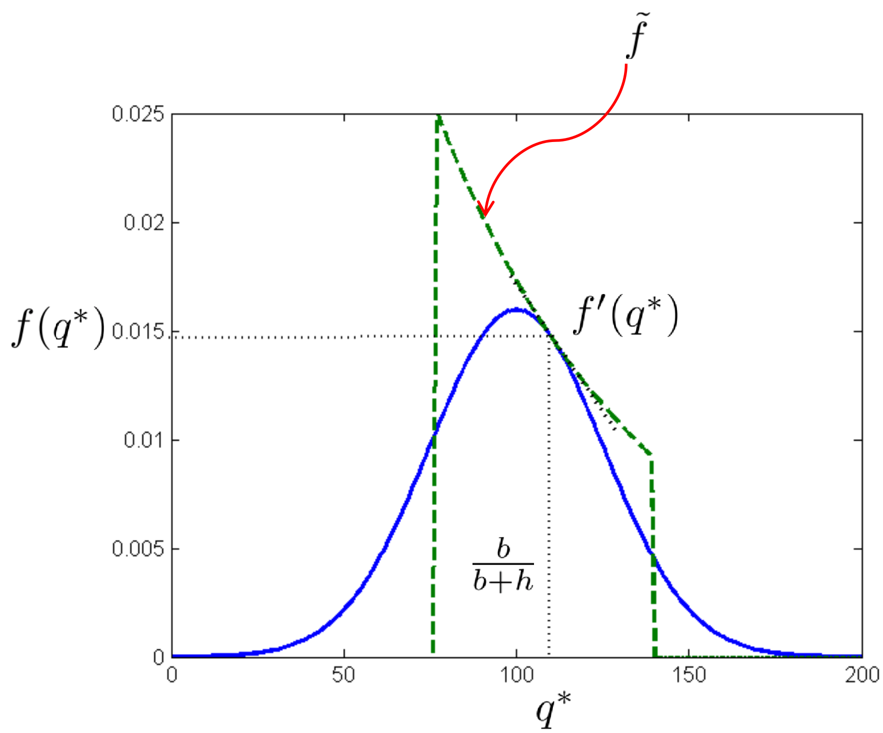


Figure C-2: Regression analysis to estimate relationship between bias of Δ_n and AMS value $\Delta_F(\beta)$.

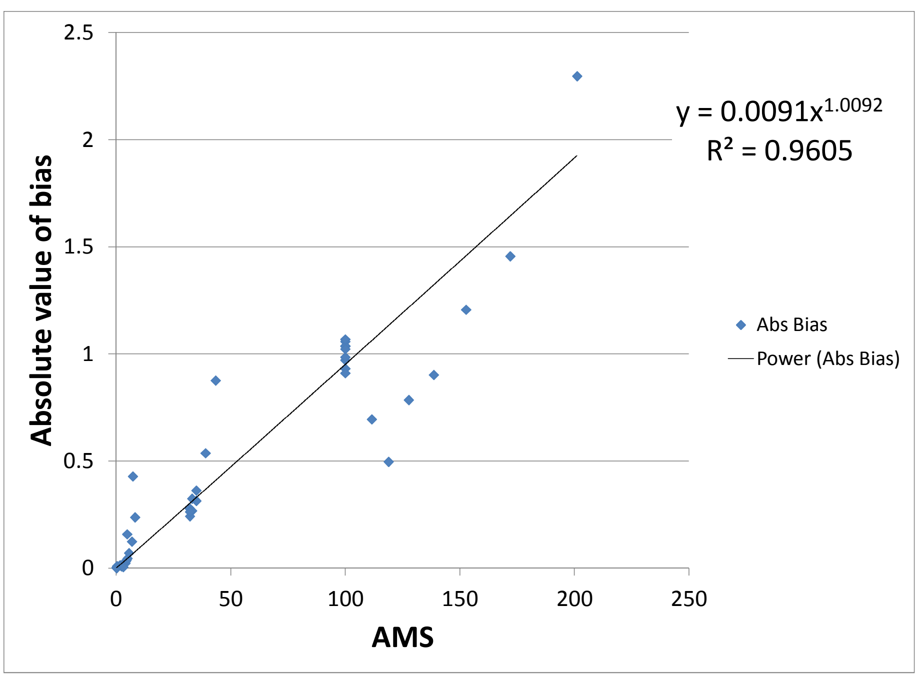


Figure C-3: Probability density functions of demand distributions used in experiments.

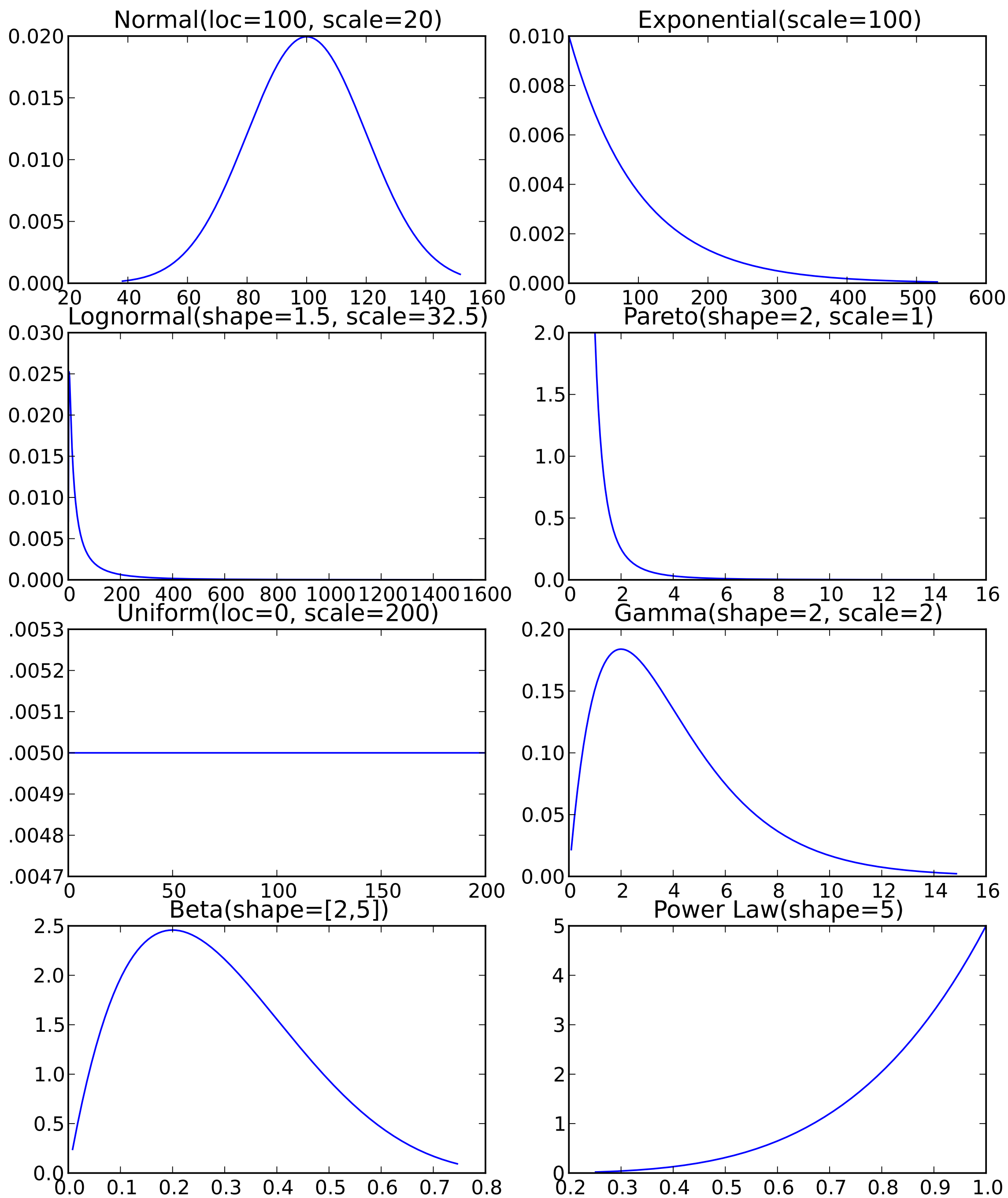
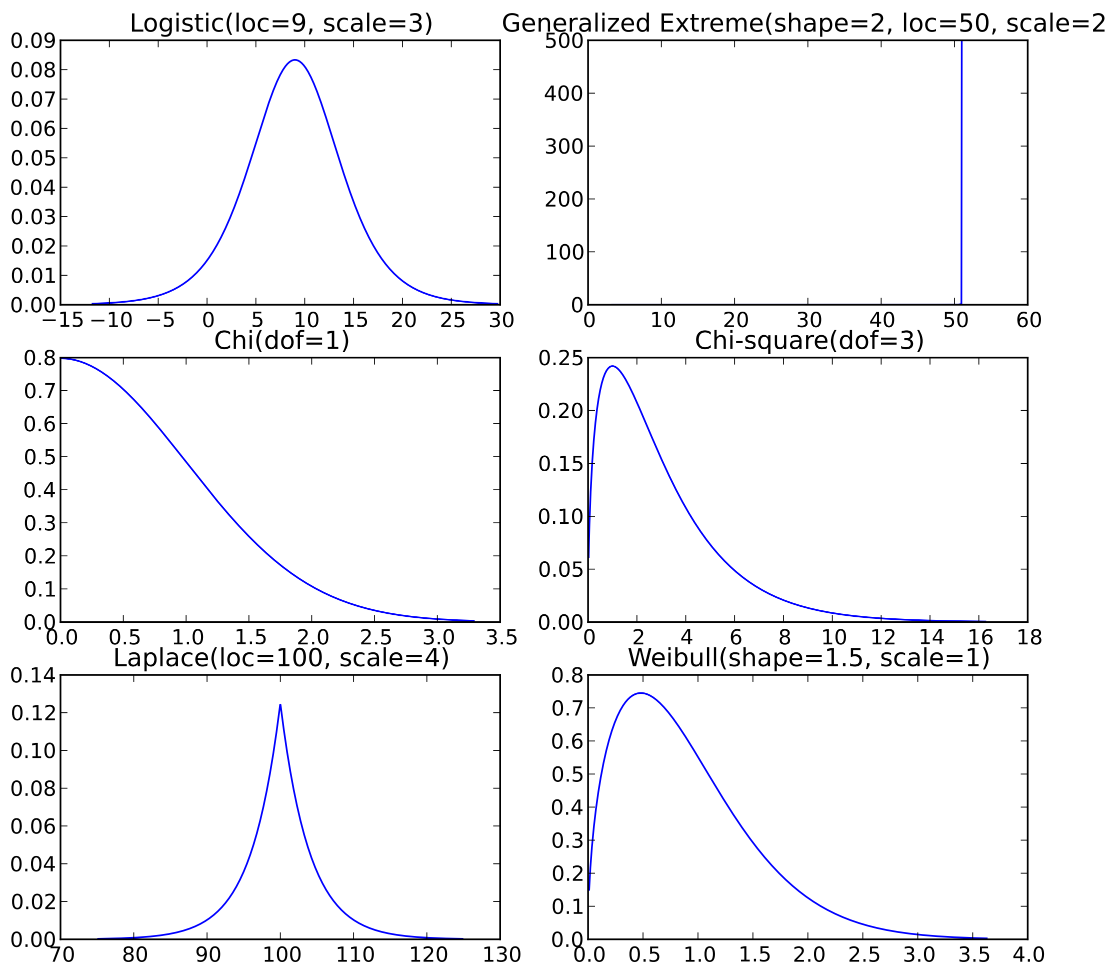


Figure C-4: Probability density functions of demand distributions used in experiments.



Appendix D

Tables

Table D.1: Range of critical fractile values where Assumption 2.4.1 holds.

Distribution	*When is Assumption 1 satisfied?*	*Notes*
Normal(μ, σ)	$\frac{b}{b+h} \geq \frac{1}{2}$	
Exponential(λ)	$\frac{b}{b+h} \geq 0$	
Lognormal(μ, σ)	$\frac{b}{b+h} \geq \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(-\frac{\sigma}{2}\right)$	erf: error function
Pareto(x_m, α)	$\frac{b}{b+h} \geq 0$	
Uniform(A, B)	$\frac{b}{b+h} \geq 0$	
Gamma(α, β)	$\frac{b}{b+h} \geq \frac{1}{\Gamma(\alpha)}\gamma(\alpha, \alpha - 1)$	Γ: gamma function, γ: incomplete gamma function
Beta(α, β)	$\frac{b}{b+h} \geq \frac{B\left(\frac{\alpha-1}{\alpha+\beta-2};\alpha,\beta\right)}{B(\alpha,\beta)}$	B: beta function
Power Law(α)	$\frac{b}{b+h} \geq 0$	
Logistic(μ, s)	$\frac{b}{b+h} \geq \frac{1}{2}$	
GEV(μ, σ, ξ)	$\frac{b}{b+h} \geq e^{-1-\xi}$	for $\xi \geq 0$
Chi(k)	$\frac{b}{b+h} \geq P\left(\frac{k}{2}, \frac{k-1}{2}\right)$	for $k \geq 1$; P: regularized gamma function
Chi-squared(k)	$\frac{b}{b+h} \geq \begin{cases} \frac{1}{\Gamma\left(\frac{k}{2}\right)}\gamma\left(\frac{k}{2}, \frac{k-2}{2}\right), & \text{if } k \geq 2 \\ 0, & \text{if } k < 2 \end{cases}$	
Laplace(μ, β)	$\frac{b}{b+h} \geq \frac{1}{2}$	
Weibull(λ, k)	$\frac{b}{b+h} \geq \begin{cases} 1 - e^{-\frac{k-1}{k}}, & \text{if } k \geq 1 \\ 0 & \text{if } k < 1 \end{cases}$	

Table D.2: Average errors (%) with samples from an exponential distribution.

(a) Sample average approximation

	Critical quantile										
Sample size	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
25	2.39	1.83	2.08	2.24	2.62	3.22	4.05	4.67	7.65	10.87	33.76
50	0.77	0.73	0.81	0.87	1.35	1.49	1.93	2.38	3.10	7.33	16.89
100	0.54	0.34	0.48	0.60	0.70	0.91	0.96	1.50	2.03	3.24	8.56
200	0.27	0.23	0.27	0.29	0.34	0.40	0.49	0.64	1.22	2.22	4.36

(b) Distribution fitting

	Critical quantile										
Sample size	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
25	1.88	1.54	1.54	1.69	2.03	2.60	3.37	4.26	5.81	9.64	40.06
50	0.65	0.64	0.69	0.80	0.99	1.23	1.53	1.90	2.72	4.88	22.93
100	0.36	0.34	0.39	0.46	0.57	0.73	0.96	1.33	1.91	2.62	9.03
200	0.21	0.20	0.21	0.24	0.28	0.34	0.43	0.59	0.94	1.64	7.25

Table D.3: Average errors (%) with samples from a normal distribution.

(a) Sample average approximation

	Critical quantile										
Sample size	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
25	6.03	3.84	3.81	3.11	2.60	2.95	3.50	4.91	6.23	8.71	42.85
50	2.31	1.69	1.62	1.58	1.41	1.60	1.59	2.06	3.26	4.57	13.76
100	1.63	1.15	0.92	0.86	0.83	0.75	0.92	1.08	1.56	2.18	5.94
200	0.81	0.45	0.38	0.36	0.30	0.29	0.38	0.47	0.81	1.41	3.65

(b) Distribution fitting

	Critical quantile										
Sample size	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
25	4.65	3.53	3.07	2.73	2.62	2.77	3.16	3.74	5.48	12.83	75.12
50	1.91	1.43	1.27	1.20	1.24	1.38	1.60	1.87	2.53	4.41	18.77
100	1.13	0.90	0.78	0.71	0.68	0.69	0.76	0.89	1.17	1.75	6.59
200	0.47	0.36	0.28	0.25	0.25	0.27	0.33	0.42	0.63	1.03	3.92

Table D.4: Average errors (%) with samples from a Pareto distribution.

(a) Sample average approximation

	Critical quantile										
Sample size	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
25	0.88	0.69	0.83	0.93	1.14	1.59	2.18	3.12	6.70	28.33	34.35
50	0.28	0.28	0.31	0.37	0.60	0.73	1.02	1.39	2.28	6.12	33.39
100	0.19	0.13	0.19	0.24	0.29	0.39	0.45	0.83	1.54	3.28	39.74
200	0.09	0.08	0.10	0.11	0.14	0.17	0.24	0.35	0.80	1.97	6.86

(b) Distribution fitting

	Critical quantile										
Sample size	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
25	0.70	0.61	0.69	0.79	0.96	1.24	1.68	2.47	4.83	9.69	40.31
50	0.25	0.25	0.28	0.32	0.39	0.53	0.76	1.15	2.25	4.52	18.92
100	0.15	0.14	0.16	0.20	0.24	0.31	0.41	0.62	1.34	2.85	11.71
200	0.08	0.08	0.09	0.10	0.12	0.15	0.19	0.30	0.72	1.65	6.97

Table D.5: Average errors (%) with samples from a Beta distribution.

(a) Sample average approximation

	Critical quantile										
Sample size	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
25	5.15	4.80	4.07	3.07	3.06	2.63	2.92	2.90	4.30	4.27	14.99
50	2.69	2.28	2.15	1.99	1.63	1.41	1.26	1.25	1.34	1.88	2.47
100	1.86	1.17	0.94	0.82	0.88	0.85	0.73	0.77	0.79	0.89	0.78
200	1.11	0.59	0.40	0.35	0.36	0.35	0.32	0.31	0.32	0.30	0.41

(b) Distribution fitting

	Critical quantile										
Sample size	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
25	5.62	4.42	3.41	2.70	2.38	2.32	2.34	2.39	3.40	7.13	35.94
50	2.90	2.24	1.77	1.50	1.43	1.49	1.61	1.65	1.68	2.88	9.40
100	1.43	1.06	0.83	0.69	0.62	0.61	0.62	0.63	0.64	0.99	4.45
200	0.70	0.43	0.30	0.26	0.25	0.25	0.26	0.26	0.24	0.37	2.29

Table D.6: Average errors (%) with samples from a mixed normal distribution.

(a) Sample average approximation

	Critical quantile										
Sample size	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
25	3.99	2.64	1.89	2.14	6.17	3.98	5.52	11.72	5.78	3.84	4.41
50	1.29	0.86	0.61	0.39	0.38	0.35	0.53	0.79	1.81	1.62	4.26
100	0.74	0.43	0.35	0.27	0.39	0.45	0.33	0.39	0.55	0.71	2.51
200	0.37	0.21	0.16	0.13	0.08	0.08	0.12	0.19	0.22	0.59	1.47

(b) Distribution fitting

	Critical quantile										
Sample size	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
25	2.96	1.67	1.86	3.32	3.09	2.36	10.46	16.65	6.55	13.74	32.75
50	1.56	1.14	0.52	0.47	0.33	0.54	0.49	2.18	0.24	1.84	25.18
100	0.90	0.40	0.35	1.08	0.81	0.18	0.16	1.85	0.33	1.19	7.76
200	0.69	0.38	0.15	0.59	0.42	0.48	0.57	0.37	0.94	4.61	2.97

Table D.7: Regression analysis to estimate relationship between bias of Δ_n and sample size n.

	$\beta = 0.8$			$\beta = 0.9$			$\beta = 0.95$		
	$\hat{C}_1$	$\hat{k}_1$	R^2	$\hat{C}_1$	$\hat{k}_1$	R^2	$\hat{C}_1$	$\hat{k}_1$	R^2
Lognormal	-149.4	-0.84	0.87	-554.7	-0.88	0.97	-3132.6	-1.05	0.97
Normal	-35.7	-1.01	0.99	-71.7	-1.06	0.99	-63.8	-0.92	1.00
Exponential	-319.7	-1.05	0.97	-528.1	-1.04	0.98	-651.8	-0.92	0.99
Pareto	-2.0	-0.91	0.74	-34.6	-1.36	0.88	-19.1	-1.04	0.88
Uniform	-76.5	-0.95	0.99	-130.0	-1.07	0.99	-80.8	-0.96	0.99
Gamma	-9.1	-1.08	0.97	-11.6	-0.98	0.99	-19.5	-0.97	1.00
Beta	-0.4	-1.04	0.99	-0.4	-0.96	0.99	-1.2	-1.07	0.98
Power Law	-0.6	-1.38	0.90	-0.6	-1.40	0.92	-0.1	-1.02	0.97

Table D.8: Theorem 3.2.2: Optimal primal and dual solutions when $z \le y \le w$

$$S(w; z, y) = \begin{cases} (1-\beta)(z-y) + \beta \frac{y-z}{w-z} \left(z - \mu + (1-\beta)\delta\right), & \text{if } z < \mu - (1-\beta)\delta, \\ (1-\beta)(z-y), & \text{if } z \ge \mu - (1-\beta)\delta, \end{cases}$$

Primal and dual solutions

For $S(w; z, y) = (1-\beta)(z-y) + \beta \frac{y-z}{w-z} \left(z - \mu + (1-\beta)\delta\right)$,

$$D = \begin{cases} z, & \text{w.p. } \beta \left(\frac{w-\mu+(1-\beta)\delta}{w-z}\right), \\ w, & \text{w.p. } \beta \left(\frac{\mu-(1-\beta)\delta-z}{w-z}\right), \\ \mu + \beta\delta, & \text{w.p. } 1-\beta \end{cases} \qquad \begin{aligned} &\alpha_0^L = \left(\frac{y-z}{w-z}\right) z, \alpha_0^U = z - y, \\ &\alpha_1 = -\frac{\beta(y-z)}{w-z}, \\ &\alpha_2 = \frac{\beta(1-\beta)(y-z)}{w-z} \end{aligned}$$

For $S(w; z, y) = (1-\beta)(z-y)$

$$D = \begin{cases} \mu - (1-\beta)\delta, & \text{w.p. } \beta, \\ \mu + \beta\delta, & \text{w.p. } 1-\beta \end{cases} \qquad \begin{aligned} &\alpha_0^L = 0, \alpha_0^U = -(y-z), \\ &\alpha_1 = \alpha_2 = 0 \end{aligned}$$

Table D.9: Theorem 3.2.2: Optimal primal and dual solutions when $z \le w \le y$

$$S(w; z, y) = \begin{cases} (1-\beta)(z-w) + \beta(z-\mu) + \beta(1-\beta)\delta, & \text{if } z < \mu - (1-\beta)\delta, \\ (1-\beta)(z-w), & \text{if } z \ge \mu - (1-\beta)\delta, \end{cases}$$

Primal and dual solutions

For $S(w; z, y) = (1-\beta)(z-w) + \beta(z-\mu) + \beta(1-\beta)\delta$

$$D = \begin{cases} \mu - (1-\beta)\delta, & \text{w.p. } \beta, \\ w, & \text{w.p. } (1-\beta)p, \\ x, & \text{w.p. } (1-\beta)(1-p) \end{cases} \qquad \begin{aligned} &\alpha_0^L = z, \alpha_0^U = z - w, \\ &\alpha_1 = -\beta, \alpha_2 = \beta(1-\beta) \end{aligned}$$

where $x = \frac{\mu+\beta\delta-pw}{1-p}$ and $p \to 1$

For $S(w; z, y) = (1-\beta)(z-w)$,

$$D = \begin{cases} \mu - (1-\beta)\delta, & \text{w.p. } \beta, \\ w, & \text{w.p. } (1-\beta)p, \\ x, & \text{w.p. } (1-\beta)(1-p) \end{cases} \qquad \begin{aligned} &\alpha_0^L = 0, \alpha_0^U = z - w, \\ &\alpha_1 = \alpha_2 = 0 \end{aligned}$$

where $x = \frac{\mu+\beta\delta-pw}{1-p}$ and $p \to 1$

Table D.10: Theorem 3.2.2: Optimal primal and dual solutions when $y \leq z \leq w$

$S(w; z, y) = z - y$

Primal and dual solutions

$\alpha_0^L = \alpha_0^U = z - y, \alpha_1 = \alpha_2 = 0$

When $z \leq \mu - (1-\beta)\delta$

$$D = \begin{cases} \mu - (1-\beta)\delta, & \text{w.p. } \beta, \\ \mu + \beta\delta, & \text{w.p. } 1-\beta, \end{cases}$$

When $z > \mu - (1-\beta)\delta$,

$$D = \begin{cases} x, & \text{w.p. } \beta p, \\ z, & \text{w.p. } \beta(1-p), \\ \mu + \beta\delta, & \text{w.p. } 1-\beta, \end{cases} \quad \text{where } x = \frac{\mu-(1-\beta)\delta - z(1-p)}{p} \text{ and } p \to 0$$

Table D.11: Theorem 3.2.2: Optimal primal and dual solutions when $y \leq w \leq z$

$$S(w; z, y) = \begin{cases} (1-\beta)(z-y) + \beta(w-y), & \text{if } z \leq \mu + \beta\delta, \\ S(w; z, y) = (1-\beta)(\mu + \beta\delta - y) + \beta(w-y), & \text{if } z > \mu + \beta\delta, \end{cases}$$

Primal and dual solutions

For $S(w; z, y) = (1-\beta)(z-y) + \beta(w-y)$

$$D = \begin{cases} x, & \text{w.p. } \beta p, \\ w, & \text{w.p. } \beta(1-p), \\ \mu + \beta\delta, & \text{w.p. } 1-\beta \end{cases} \qquad \begin{array}{l} \alpha_0^L = w - y, \alpha_0^U = z - y, \\ \alpha_1 = \alpha_2 = 0 \end{array}$$

where $x = \frac{\mu-(1-\beta)\delta - w(1-p)}{p}$ and $p \to 0$

For $S(w; z, y) = (1-\beta)(\mu + \beta\delta - y) + \beta(w-y)$

$$D = \begin{cases} x, & \text{w.p. } \beta p, \\ w, & \text{w.p. } \beta(1-p), \\ \mu + \beta\delta, & \text{w.p. } 1-\beta \end{cases} \qquad \begin{array}{l} \alpha_0^L = w - y, \alpha_0^U = -y, \\ \alpha_1 = 1-\beta, \alpha_2 = \beta(1-\beta) \end{array}$$

where $x = \frac{\mu-(1-\beta)\delta - w(1-p)}{p}$ and $p \to 0$

Table D.12: Theorem 3.2.2: Optimal primal and dual solutions when $w \leq z \leq y$

$S(w; z, y) = 0$

Primal and dual solutions

$\alpha_0^L = \alpha_0^U = \alpha_1 = \alpha_2 = 0$

When $z \leq \mu + \beta\delta$

$$D = \begin{cases} \mu - (1-\beta)\delta, & \text{w.p. } \beta, \\ z, & \text{w.p. } (1-\beta)p, \\ x, & \text{w.p. } (1-\beta)(1-p), \end{cases} \quad \text{where } x = \frac{\mu+\beta\delta-pz}{1-p} \text{ and } p \to 0$$

When $z > \mu + \beta\delta$

$$D = \begin{cases} \mu - (1-\beta)\delta, & \text{w.p. } \beta, \\ \mu + \beta\delta, & \text{w.p. } 1-\beta, \end{cases}$$

Table D.13: Theorem 3.2.2: Optimal primal and dual solutions when $w \le y \le z$

$$S(w;z,y) = \begin{cases} (1-\beta)(z-y), & \text{if } z \le \mu+\beta\delta, \\ \frac{(1-\beta)(z-y)}{z-w}\left(\mu+\beta\delta-w\right), & \text{if } z > \mu+\beta\delta, \end{cases}$$

Primal and dual solutions

For $S(w;z,y) = (1-\beta)(z-y)$

$$D = \begin{cases} \mu-(1-\beta)\delta, & \text{w.p. } \beta, \\ \mu+\beta\delta, & \text{w.p. } 1-\beta, \end{cases} \qquad \alpha_0^L = 0, \alpha_0^U = z-y, \quad \alpha_1 = \alpha_2 = 0$$

For $S(w;z,y) = \frac{(1-\beta)(z-y)}{z-w}\left(\mu+\beta\delta-w\right)$

$$D = \begin{cases} \mu-(1-\beta)\delta, & \text{w.p. } \beta, \\ w, & \text{w.p. } (1-\beta)\left(\frac{z-\mu-\beta\delta}{z-w}\right), \\ z, & \text{w.p. } (1-\beta)\left(\frac{\mu+\beta\delta-w}{z-w}\right) \end{cases} \qquad \alpha_0^L = 0, \alpha_0^U = -\frac{w(z-y)}{z-w}, \quad \alpha_1 = \frac{(1-\beta)(z-y)}{z-w}, \alpha_2 = \frac{\beta(1-\beta)(z-y)}{z-w}$$

Table D.14: Theorem 3.2.2: $G(z;y)$

$G(z;y)$ for $z \in (-\infty, y]$

If $y \le \mu-(1-\beta)\delta$

$$G(z;y) = -\frac{\beta(y-z)(\mu-(1-\beta)\delta-z)}{\mu+\beta\delta-z}, \quad \text{for } z \in (-\infty, y]$$

If $\mu-(1-\beta)\delta < y \le \mu-(1-\beta)\delta+\beta\delta$

$$G(z;y) = \begin{cases} -\frac{\beta(y-z)(\mu-(1-\beta)\delta-z)}{\mu+\beta\delta-z}, & \text{for } z \in \left(-\infty, \frac{(\mu-(1-\beta)\delta)(\mu+\beta\delta)-\mu y}{\mu-(1-\beta)\delta+\beta\delta-y}\right], \\ (1-\beta)y+\beta z-\mu+(1-\beta)\delta, & \text{for } z \in \left(\frac{(\mu-(1-\beta)\delta)(\mu+\beta\delta)-\mu y}{\mu-(1-\beta)\delta+\beta\delta-y}, \mu-(1-\beta)\delta\right], \\ (1-\beta)(y-z), & \text{for } z \in (\mu-(1-\beta)\delta, y] \end{cases}$$

If $y > \mu-(1-\beta)\delta+\beta\delta$

$$G(z;y) = \begin{cases} (1-\beta)y+\beta z-\mu+(1-\beta)\delta, & \text{for } z \in (-\infty, \mu-(1-\beta)\delta], \\ (1-\beta)(y-z), & \text{for } z \in (\mu-(1-\beta)\delta, y] \end{cases}$$

$G(z;y)$ for $z \in [y, \infty)$

If $y \le \mu-(1-\beta)\delta+\beta\delta$

$$G(z;y) = \begin{cases} \beta(z-y), & \text{for } z \in [y, \mu+\beta\delta], \\ -\beta y-(1-\beta)z+\mu+\beta\delta, & \text{for } z \in (\mu+\beta\delta, \infty) \end{cases}$$

If $\mu-(1-\beta)\delta+\beta\delta < y \le \mu+\beta\delta$

$$G(z;y) = \begin{cases} \beta(z-y), & \text{for } z \in [y, \mu+\beta\delta], \\ -\beta y-(1-\beta)z+\mu+\beta\delta, & \text{for } z \in \left(\mu+\beta\delta, \frac{\mu y-(\mu-(1-\beta)\delta)(\mu+\beta\delta)}{y-\mu+(1-\beta)\delta-\beta\delta}\right], \\ -\frac{(1-\beta)(z-y)(z-\mu-\beta\delta)}{z-\mu+(1-\beta)\delta}, & \text{for } z \in \left(\frac{\mu y-(\mu-(1-\beta)\delta)(\mu+\beta\delta)}{y-\mu+(1-\beta)\delta-\beta\delta}, \infty\right) \end{cases}$$

If $y > \mu+\beta\delta$

$$G(z;y) = -\frac{(1-\beta)(z-y)(z-\mu-\beta\delta)}{z-\mu+(1-\beta)\delta}, \quad \text{for } z \in [y, \infty)$$

Table D.15: Theorem A.3.1: Optimal primal and dual solutions when $z \le y \le w$

If $\mu - (1-\beta)\delta_U < 0$,

$S(w; z, y) = (1-\beta)(z-y)$,

If $\mu - (1-\beta)\delta_U \ge 0$,

$$S(w; z, y) = \begin{cases} (1-\beta)(z-y) + \beta \frac{y-z}{w-z} \left(z - \mu + (1-\beta)\delta_U\right), & \text{if } z \le \mu - (1-\beta)\delta_U, \\ (1-\beta)(z-y), & \text{if } z > \mu - (1-\beta)\delta_U, \end{cases}$$

Primal and Dual Solutions

For $S(w; z, y) = (1-\beta)(z-y) + \beta \frac{y-z}{w-z} \left(z - \mu + (1-\beta)\delta_U\right)$

$$D = \begin{cases} z, & \text{w.p. } \beta \left(\frac{w-\mu+(1-\beta)\delta_U}{w-z}\right), \\ w, & \text{w.p. } \beta \left(\frac{\mu-(1-\beta)\delta_U - z}{w-z}\right), \\ \mu + \beta\delta_U, & \text{w.p. } 1-\beta \end{cases} \qquad \begin{array}{l} \alpha_0^L = \left(\frac{y-z}{w-z}\right) z, \alpha_0^U = z - y, \\ \alpha_1 = -\frac{\beta(y-z)}{w-z}, \\ \alpha_2^L = 0, \alpha_2^U = \frac{\beta(1-\beta)(y-z)}{w-z} \end{array}$$

For $S(w; z, y) = (1-\beta)(z-y)$

$$D = \begin{cases} \mu - (1-\beta)\delta_U, & \text{w.p. } \beta, \\ \mu + \beta\delta_U, & \text{w.p. } 1-\beta \end{cases} \qquad \begin{array}{l} \alpha_0^L = 0, \alpha_0^U = -(y-z), \\ \alpha_1 = \alpha_2^L = \alpha_2^U = 0 \end{array}$$

Table D.16: Theorem A.3.1: Optimal primal and dual solutions when $z \le w \le y$

If $\mu - (1-\beta)\delta_U < 0$,

$S(w; z, y) = (1-\beta)(z-w)$,

If $\mu - (1-\beta)\delta_U \ge 0$,

$$S(w; z, y) = \begin{cases} (1-\beta)(z-w) + \beta(z-\mu) + \beta(1-\beta)\delta_U, & \text{if } z \le \mu - (1-\beta)\delta_U, \\ (1-\beta)(z-w), & \text{if } z > \mu - (1-\beta)\delta_U, \end{cases}$$

Primal and dual solutions

For $S(w; z, y) = (1-\beta)(z-w) + \beta(z-\mu) + \beta(1-\beta)\delta_U$

$$D = \begin{cases} \mu - (1-\beta)\delta_U, & \text{w.p. } \beta, \\ w, & \text{w.p. } (1-\beta)p, \\ x, & \text{w.p. } (1-\beta)(1-p), \end{cases} \qquad \begin{array}{l} \alpha_0^L = z, \alpha_0^U = z - w, \\ \alpha_1 = -\beta, \alpha_2^L = 0, \alpha_2^U = \beta(1-\beta) \end{array}$$

where $x = \frac{\mu+\beta\delta_U - wp}{1-p}$ as $p \to 1$

For $S(w; z, y) = (1-\beta)(z-w)$ when $\mu - (1-\beta)\delta_U < 0$,

$$D = \begin{cases} 0, & \text{w.p. } \beta, \\ w, & \text{w.p. } (1-\beta)p, \\ x, & \text{w.p. } (1-\beta)(1-p), \end{cases} \qquad \begin{array}{l} \alpha_0^L = 0, \alpha_0^U = z - w, \\ \alpha_1 = \alpha_2^L = \alpha_2^U = 0 \end{array}$$

where $x = \frac{1}{1-p}\left(\frac{\mu}{1-\beta} - pw\right)$ as $p \to 1$

For $S(w; z, y) = (1-\beta)(z-w)$ when $\mu - (1-\beta)\delta_U \ge 0$,

$$D = \begin{cases} \mu - (1-\beta)\delta_U, & \text{w.p. } \beta, \\ w, & \text{w.p. } (1-\beta)p, \\ x, & \text{w.p. } (1-\beta)(1-p), \end{cases} \qquad \begin{array}{l} \alpha_0^L = 0, \alpha_0^U = z - w, \\ \alpha_1 = \alpha_2^L = \alpha_2^U = 0 \end{array}$$

where $x = \frac{\mu+\beta\delta_U - wp}{1-p}$ as $p \to 1$

Table D.17: Theorem A.3.1: Optimal primal and dual solutions when $y \leq z \leq w$

If $w < \mu + \beta\delta_L$,

$$S(w;z,y) = \begin{cases} z-y, & \text{if } z \leq \mu-(1-\beta)\delta_L, \\ (1-\beta)(z-y) + \frac{\beta(z-y)}{z}\left(\mu-(1-\beta)\delta_L\right), & \text{if } z > \mu-(1-\beta)\delta_L, \end{cases}$$

If $w \geq \mu + \beta\delta_L$,

$$S(w;z,y) = \begin{cases} z-y, & \text{if } w \leq \frac{\mu-\beta z}{1-\beta}, \\ (1-\beta)(z-y) + \frac{z-y}{z}\left(\mu-(1-\beta)w\right), & \text{if } w > \frac{\mu-\beta z}{1-\beta}, \end{cases}$$

Primal and dual solutions

For $S(w;z,y) = z-y$ when $w < \mu-(1-\beta)\delta_L$

$$D = \begin{cases} w, & \text{w.p. } \beta, \\ \frac{\mu-\beta w}{1-\beta}, & \text{w.p. } 1-\beta, \end{cases} \qquad \begin{array}{l} \alpha_0^L = \alpha_0^U = z-y, \\ \alpha_1 = \alpha_2^L = \alpha_2^U = 0 \end{array}$$

For $S(w;z,y) = z-y$ when $\mu-(1-\beta)\delta_L \leq w < \mu+\beta\delta_L$

$$D = \begin{cases} \mu-(1-\beta)\delta_L, & \text{w.p. } \beta, \\ \mu+\beta\delta_L, & \text{w.p. } 1-\beta, \end{cases} \qquad \begin{array}{l} \alpha_0^L = \alpha_0^U = z-y, \\ \alpha_1 = \alpha_2^L = \alpha_2^U = 0 \end{array}$$

For $S(w;z,y) = z-y$ when $w \geq \mu+\beta\delta_L$

$$D = \begin{cases} \frac{\mu-(1-\beta)w}{\beta}, & \text{w.p. } \beta, \\ w, & \text{w.p. } 1-\beta, \end{cases} \qquad \begin{array}{l} \alpha_0^L = \alpha_0^U = z-y, \\ \alpha_1 = \alpha_2^L = \alpha_2^U = 0 \end{array}$$

For $S(w;z,y) = (1-\beta)(z-y) + \frac{\beta(z-y)}{z}\left(\mu-(1-\beta)\delta_L\right)$

$$D = \begin{cases} 0, & \text{w.p. } \beta\left(1-\frac{\mu-(1-\beta)\delta_L}{z}\right), \\ z, & \text{w.p. } \beta\left(\frac{\mu-(1-\beta)\delta_L}{z}\right), \\ \mu+\beta\delta_L, & \text{w.p. } 1-\beta \end{cases} \qquad \begin{array}{l} \alpha_0^L = 0, \alpha_0^U = z-y, \\ \alpha_1 = \frac{\beta(z-y)}{z}, \alpha_2^L = -\beta(1-\beta)\frac{z-y}{z}, \alpha_2^U = 0 \end{array}$$

For $S(w;z,y) = (1-\beta)(z-y) + \frac{z-y}{z}\left(\mu-(1-\beta)w\right)$

$$D = \begin{cases} 0, & \text{w.p. } \beta-\frac{\mu-(1-\beta)w}{z}, \\ z, & \text{w.p. } \frac{\mu-(1-\beta)w}{z}, \\ w, & \text{w.p. } 1-\beta \end{cases} \qquad \begin{array}{l} \alpha_0^L = 0, \alpha_0^U = \frac{z-y}{z}(z-w), \\ \alpha_1 = \frac{z-y}{z}, \alpha_2^L = \alpha_2^U = 0 \end{array}$$

Table D.18: Theorem A.3.1: Optimal primal and dual solutions when $y \le w \le z$

If $\mu-(1-\beta)\delta_U < 0$ and $w \le \mu-(1-\beta)\delta_L$,

$$S(w;z,y) = \begin{cases} (1-\beta)(z-y)+\beta(w-y), & \text{if } z \le \frac{\mu-\beta w}{1-\beta}, \\ \mu-(1-\beta)y - \frac{y}{w}\left(\mu-(1-\beta)z\right), & \text{if } \frac{\mu-\beta w}{1-\beta} < z \le \frac{\mu}{1-\beta}, \\ \mu-(1-\beta)y, & \text{if } z > \frac{\mu}{1-\beta}, \end{cases}$$

If $\mu-(1-\beta)\delta_U < 0$ and $w > \mu-(1-\beta)\delta_L$,

$$S(w;z,y) = \begin{cases} (1-\beta)(z-y)+\frac{\beta(w-y)}{w}\left(\mu-(1-\beta)\delta_L\right), & \text{if } z \le \mu+\beta\delta_L, \\ \mu-(1-\beta)y - \frac{y}{w}\left(\mu-(1-\beta)z\right), & \text{if } \mu+\beta\delta_L < z \le \frac{\mu}{1-\beta}, \\ \mu-(1-\beta)y, & \text{if } z > \frac{\mu}{1-\beta}, \end{cases}$$

If $\mu-(1-\beta)\delta_U \ge 0$ and $w \le \mu-(1-\beta)\delta_L$,

$$S(w;z,y) = \begin{cases} (1-\beta)(z-y)+\beta(w-y), & \text{if } z \le \frac{\mu-\beta w}{1-\beta}, \\ \mu-(1-\beta)y - \frac{y}{w}\left(\mu-(1-\beta)z\right), & \text{if } \frac{\mu-\beta w}{1-\beta} < z \le \mu+\beta\delta_U, \\ \mu-(1-\beta)y - \frac{y}{w}\left(\beta\mu-\beta(1-\beta)\delta_U\right), & \text{if } z > \mu+\beta\delta_U, \end{cases}$$

If $\mu-(1-\beta)\delta_U \ge 0$ and $w > \mu-(1-\beta)\delta_L$,

$$S(w;z,y) = \begin{cases} (1-\beta)(z-y)+\frac{\beta(w-y)}{w}\left(\mu-(1-\beta)\delta_L\right), & \text{if } z \le \mu+\beta\delta_L, \\ \mu-(1-\beta)y - \frac{y}{w}\left(\mu-(1-\beta)z\right), & \text{if } \mu+\beta\delta_L < z \le \mu+\beta\delta_U, \\ \mu-(1-\beta)y - \frac{y}{w}\left(\beta\mu-\beta(1-\beta)\delta_U\right), & \text{if } z > \mu+\beta\delta_U, \end{cases}$$

Primal and dual solutions

For $S(w;z,y) = (1-\beta)(z-y)+\beta(w-y)$

$$D = \begin{cases} w, & \text{w.p. } \beta, \\ \frac{\mu-\beta w}{1-\beta}, & \text{w.p. } 1-\beta, \end{cases} \qquad \begin{array}{l} \alpha_0^L = w-y, \alpha_0^U = z-y, \\ \alpha_1 = \alpha_2^L = \alpha_2^U = 0 \end{array}$$

For $S(w;z,y) = (1-\beta)(z-y)+\beta\frac{w-y}{w}\left(\mu-(1-\beta)\delta_L\right)$

$$D = \begin{cases} 0, & \text{w.p. } \beta\left(1-\frac{\mu-(1-\beta)\delta_L}{w}\right), \\ w, & \text{w.p. } \beta\left(\frac{\mu-(1-\beta)\delta_L}{w}\right), \\ \mu+\beta\delta_L, & \text{w.p. } 1-\beta, \end{cases} \qquad \begin{array}{l} \alpha_0^L = 0, \alpha_0^U = z-y, \\ \alpha_1 = \beta\frac{w-y}{w}, \alpha_2^L = -\beta(1-\beta)\frac{w-y}{w}, \alpha_2^U = 0 \end{array}$$

For $S(w;z,y) = \mu-(1-\beta)y-\frac{y}{w}\left(\mu-(1-\beta)z\right)$

$$D = \begin{cases} 0, & \text{w.p. } \beta-\frac{\mu-z(1-\beta)}{w}, \\ w, & \text{w.p. } \frac{\mu-z(1-\beta)}{w}, \\ z, & \text{w.p. } 1-\beta, \end{cases} \qquad \begin{array}{l} \alpha_0^L = 0, \alpha_0^U = \frac{y(z-w)}{w}, \\ \alpha_1 = \frac{w-y}{w}, \alpha_2^L = \alpha_2^U = 0 \end{array}$$

For $S(w;z,y) = \mu-(1-\beta)y-\frac{y}{w}\left(\mu\beta-\beta(1-\beta)\delta_U\right)$

$$D = \begin{cases} 0, & \text{w.p. } \beta\left(1-\frac{\mu-(1-\beta)\delta_U}{w}\right), \\ w, & \text{w.p. } \beta\left(\frac{\mu-(1-\beta)\delta_U}{w}\right), \\ \mu+\beta\delta_U, & \text{w.p. } 1-\beta, \end{cases} \qquad \begin{array}{l} \alpha_0^L = 0, \alpha_0^U = -y, \\ \alpha_1 = \frac{w-\beta y}{w}, \alpha_2^L = 0, \alpha_2^U = \frac{\beta(1-\beta)y}{w} \end{array}$$

For $S(w;z,y) = \mu-(1-\beta)y$

$$D = \begin{cases} 0, & \text{w.p. } \beta, \\ \frac{\mu}{1-\beta}, & \text{w.p. } 1-\beta, \end{cases} \qquad \begin{array}{l} \alpha_0^L = 0, \alpha_0^U = -y, \\ \alpha_1 = 1, \alpha_2^L = \alpha_2^U = 0 \end{array}$$

Table D.19: Theorem A.3.1: Optimal primal and dual solutions when $w \leq z \leq y$

$S(w; z, y) = 0$

Primal and dual solutions

$\alpha_0^L = \alpha_0^U = \alpha_1 = \alpha_2 = 0$

If $\mu - (1-\beta)\delta_U < 0$ and $z < \frac{\mu}{1-\beta}$,

$$D = \begin{cases} 0, & \text{w.p. } \beta, \\ z, & \text{w.p. } (1-\beta)p, \\ x, & \text{w.p. } (1-\beta)(1-p), \end{cases} \qquad \text{where } x = \frac{1}{1-p}\left(\frac{\mu}{1-\beta} - pz\right) \text{ and } p \to 1,$$

If $\mu - (1-\beta)\delta_U < 0$ and $z \geq \frac{\mu}{1-\beta}$,

$$D = \begin{cases} 0, & \text{w.p. } \beta, \\ \frac{\mu}{1-\beta}, & \text{w.p. } 1-\beta, \end{cases}$$

If $\mu - (1-\beta)\delta_U \geq 0$ and $z < \mu + \beta\delta_U$,

$$D = \begin{cases} \mu - (1-\beta)\delta_U, & \text{w.p. } \beta, \\ z, & \text{w.p. } (1-\beta)p, \\ x, & \text{w.p. } (1-\beta)(1-p), \end{cases} \qquad \text{where } x = \frac{\mu+\beta\delta_U - pz}{1-p} \text{ and } p \to 1,$$

If $\mu - (1-\beta)\delta_U \geq 0$ and $z \geq \mu + \beta\delta_U$,

$$D = \begin{cases} \mu - (1-\beta)\delta_U, & \text{w.p. } \beta, \\ \mu + \beta\delta_U, & \text{w.p. } 1-\beta, \end{cases}$$

Table D.20: Theorem A.3.1: Primal and dual solutions when $w \leq y \leq z$

If $\mu - (1-\beta)\delta_U < 0$,

$$S(w; z, y) = \begin{cases} (1-\beta)(z-y), & \text{if } z \leq \frac{\mu}{1-\beta}, \\ \frac{z-y}{z-w}\left(\mu - (1-\beta)w\right), & \text{if } z > \frac{\mu}{1-\beta}, \end{cases}$$

If $\mu - (1-\beta)\delta_U \geq 0$,

$$S(w; z, y) = \begin{cases} (1-\beta)(z-y), & \text{if } z \leq \mu + \beta\delta_U, \\ \frac{(1-\beta)(z-y)}{z-w}\left(\mu + \beta\delta_U - w\right), & \text{if } z > \mu + \beta\delta_U, \end{cases}$$

Primal and dual solutions

For $S(w; z, y) = (1-\beta)(z-y)$ and $\mu - (1-\beta)\delta_U < 0$,

$$D = \begin{cases} 0, & \text{w.p. } \beta, \\ \frac{\mu}{1-\beta}, & \text{w.p. } 1-\beta, \end{cases} \qquad \begin{aligned} &\alpha_0^L = 0, \alpha_0^U = z - y, \\ &\alpha_1 = \alpha_2^L = \alpha_2^U = 0 \end{aligned}$$

For $S(w; z, y) = (1-\beta)(z-y)$ and $\mu - (1-\beta)\delta_U \geq 0$,

$$D = \begin{cases} \mu - (1-\beta)\delta_U, & \text{w.p. } \beta, \\ \mu + \beta\delta_U, & \text{w.p. } 1-\beta, \end{cases} \qquad \begin{aligned} &\alpha_0^L = 0, \alpha_0^U = z - y, \\ &\alpha_1 = \alpha_2^L = \alpha_2^U = 0 \end{aligned}$$

For $S(w; z, y) = \frac{(1-\beta)(z-y)}{z-w}\left(\mu - w + \beta\delta\right)$

$$D = \begin{cases} \mu - (1-\beta)\delta_U, & \text{w.p. } \beta, \\ w, & \text{w.p. } (1-\beta)\left(\frac{z-\mu-\beta\delta_U}{z-w}\right), \\ z, & \text{w.p. } (1-\beta)\left(\frac{\mu-w+\beta\delta_U}{z-w}\right) \end{cases} \qquad \begin{aligned} &\alpha_0^L = \alpha_2^L = 0, \alpha_0^U = -\frac{w(z-y)}{z-w}, \\ &\alpha_1 = \frac{(1-\beta)(z-y)}{z-w}, \alpha_2^U = \frac{\beta(1-\beta)(z-y)}{z-w} \end{aligned}$$

For $S(w; z, y) = \frac{(z-y)}{z-w}\left(\mu - (1-\beta)w\right)$

$$D = \begin{cases} 0, & \text{w.p. } \beta, \\ w, & \text{w.p. } \frac{(1-\beta)z-\mu}{z-w}, \\ z, & \text{w.p. } \frac{\mu-(1-\beta)w}{z-w}, \end{cases} \qquad \begin{aligned} &\alpha_0^L = 0, \alpha_0^U = -\frac{w(z-y)}{z-w}, \\ &\alpha_1 = \frac{z-y}{z-w}, \alpha_2^L = \alpha_2^U = 0 \end{aligned}$$

Table D.21: Theorem A.3.1: $G(z;y)$ when $\mu-(1-\beta)\delta_U<0$

$G(z;y)$ for $z\in[0,y]$

$G(z;y)=(1-\beta)(y-z), \quad \text{for } z\in[0,y]$

$G(z;y)$ for $z\in[y,\infty)$

If $0\le y\le \mu-(1-\beta)\delta_L$,

$$G(z;y)=\begin{cases}\beta(z-y), & \text{for } z\in[y,\mu-(1-\beta)\delta_L],\\ \beta\frac{(z-y)}{z}(\mu-(1-\beta)\delta_L), & \text{for } z\in[\mu-(1-\beta)\delta_L,\mu+\beta\delta_L],\\ \frac{z-y}{z}\left(\mu-(1-\beta)z\right), & \text{for } z\in[\mu+\beta\delta_L,\infty),\end{cases}$$

If $\mu-(1-\beta)\delta_L\le y\le \mu+\beta\delta_L$,

$$G(z;y)=\begin{cases}\beta\frac{(z-y)}{z}(\mu-(1-\beta)\delta_L), & \text{for } z\in[y,\mu+\beta\delta_L],\\ \frac{z-y}{z}\left(\mu-(1-\beta)z\right), & \text{for } z\in[\mu+\beta\delta_L,\infty),\end{cases}$$

If $y>\mu+\beta\delta_L$,

$$G(z;y)=\frac{z-y}{z}\left(\mu-(1-\beta)z\right), \quad \text{for } z\in[y,\infty)$$

Table D.22: Theorem A.3.1: $G(z;y)$ when $\mu-(1-\beta)\delta_U \geq 0$

$G(z;y)$ for $z \in [0,y]$

If $0 \leq y \leq \mu-(1-\beta)\delta_U$

$$G(z;y) = -\frac{\beta(y-z)(\mu-(1-\beta)\delta_U-z)}{\mu+\beta\delta_U-z}, \quad \text{for } z \in [0,y]$$

If $\mu-(1-\beta)\delta_U \leq y \leq \frac{1}{\mu}\left(\mu-(1-\beta)\delta_U\right)\left(\mu+\beta\delta_U\right)$

$$G(z;y) = \begin{cases} -\frac{\beta(y-z)(\mu-(1-\beta)\delta_U-z)}{\mu+\beta\delta_U-z}, & \text{for } z \in \left[0, \frac{(\mu-(1-\beta)\delta_U)(\mu+\beta\delta_U)-\mu y}{\mu-(1-\beta)\delta_U+\beta\delta_U-y}\right], \\ (1-\beta)y-\mu+\beta z+(1-\beta)\delta_U, & \text{for } z \in \left[\frac{(\mu-(1-\beta)\delta_U)(\mu+\beta\delta_U)-\mu y}{\mu-(1-\beta)\delta_U+\beta\delta_U-y}, \mu-(1-\beta)\delta_U\right], \\ (1-\beta)(y-z), & \text{for } z \in [\mu-(1-\beta)\delta_U, y] \end{cases}$$

If $y \geq \frac{1}{\mu}\left(\mu-(1-\beta)\delta_U\right)\left(\mu+\beta\delta_U\right)$

$$G(z;y) = \begin{cases} (1-\beta)y-\mu+\beta z+(1-\beta)\delta_U, & \text{for } z \in \left[0, \mu-(1-\beta)\delta_U\right], \\ (1-\beta)(y-z), & \text{for } z \in [\mu-(1-\beta)\delta_U, y] \end{cases}$$

$G(z;y)$ for $z \in [y,\infty)$

If $0 \leq y \leq \mu-(1-\beta)\delta_U$

$$G(z;y) = \begin{cases} \beta(z-y), & \text{for } z \in [y, \mu-(1-\beta)\delta_L], \\ \beta\frac{(z-y)}{z}(\mu-(1-\beta)\delta_L), & \text{for } z \in [\mu-(1-\beta)\delta_L, \mu+\beta\delta_L], \\ \frac{z-y}{z}\left(\mu-(1-\beta)z\right), & \text{for } z \in [\mu+\beta\delta_L, \mu+\beta\delta_U], \\ \mu-(1-\beta)z-\frac{\beta y(\mu-(1-\beta)\delta_U)}{\mu+\beta\delta_U}, & \text{for } z \in [\mu+\beta\delta_U, \infty) \end{cases}$$

If $\mu-(1-\beta)\delta_U \leq y \leq \mu-(1-\beta)\delta_L$

$$G(z;y) = \begin{cases} \beta(z-y), & \text{for } z \in [y, \mu-(1-\beta)\delta_L], \\ \beta\frac{(z-y)}{z}(\mu-(1-\beta)\delta_L), & \text{for } z \in [\mu-(1-\beta)\delta_L, \mu+\beta\delta_L], \\ \frac{z-y}{z}\left(\mu-(1-\beta)z\right), & \text{for } z \in [\mu+\beta\delta_L, \mu+\beta\delta_U], \\ \text{decreasing function in } z, & \text{for } z \in [\mu+\beta\delta_U, \infty) \end{cases}$$

If $\mu-(1-\beta)\delta_L \leq y \leq \mu+\beta\delta_L$

$$G(z;y) = \begin{cases} \beta\frac{(z-y)}{z}(\mu-(1-\beta)\delta_L), & \text{for } z \in [y, \mu+\beta\delta_L], \\ \frac{z-y}{z}\left(\mu-(1-\beta)z\right), & \text{for } z \in [\mu+\beta\delta_L, \mu+\beta\delta_U], \\ \text{decreasing function in } z, & \text{for } z \in [\mu+\beta\delta_U, \infty) \end{cases}$$

If $\mu+\beta\delta_L \leq y \leq \mu+\beta\delta_U$

$$G(z;y) = \begin{cases} \frac{z-y}{z}\left(\mu-(1-\beta)z\right), & \text{for } z \in [y, \mu+\beta\delta_U], \\ \text{decreasing function in } z, & \text{for } z \in \left[\mu+\beta\delta_U, \infty\right), \end{cases}$$

If $y \geq \mu+\beta\delta_U$

$$G(z;y) = -\frac{(1-\beta)(z-y)(z-\mu-\beta\delta_U)}{z-\mu+(1-\beta)\delta_U}, \quad \text{for } z \in [y,\infty)$$

Table D.23: Durations of standard jobs.

Job no.	Duration	Job no.	Duration	Job no.	Duration
1	6.58	6	5.36	11	3.48
2	6.41	7	4.96	12	2.66
3	5.63	8	4.85	13	2.61
4	5.49	9	4.25	14	2.26
5	5.47	10	3.83	15	1.51

Table D.24: Probability distributions of number of gas leaks.

	Leak scenario					
	0 leaks	1 leak	2 leaks	3 leaks	E[no. leaks]	Stdev[no. leaks]
Leak distribution 1	0.0	1.0	0.0	0.0	1.0	0.00
Leak distribution 2	0.1	0.8	0.1	0.0	1.0	0.45
Leak distribution 3	0.2	0.6	0.2	0.0	1.0	0.63
Leak distribution 4	0.4	0.2	0.4	0.0	1.0	0.89
Leak distribution 5	0.3	0.5	0.1	0.1	1.0	0.89
Leak distribution 6	0.4	0.3	0.2	0.1	1.0	1.00
Leak distribution 7	0.5	0.2	0.1	0.2	1.0	1.18

Table D.25: Optimal crew assignment with leak distribution 1.

		Crew 1	Crew 2	Crew 3	Crew 4	Crew 5	Crew 6	Crew 7
Standard jobs	Total # jobs	2	2	2	1	2	3	3
	Total hours	10.3	10.7	10.4	2.6	10.5	10.3	10.6
Gas leak jobs	Total hours *(1 leak, $p = 1$)*	0	0	0	8	0	0	0

[a] If a cell is highlighted, then the crew (column) is assigned a leak during that scenario (row).

Table D.26: Optimal crew assignment with leak distribution 2.

		Crew 1	Crew 2	Crew 3	Crew 4	Crew 5	Crew 6	Crew 7
Standard jobs	Total # jobs	3	2	2	1	3	2	2
	Total hours	10.4	10.4	10.2	2.6	10.6	10.4	10.6
Gas leak jobs	Total hours *(0 leak, $p = 0.1$)*	0	0	0	0	0	0	0
	Total hours *(1 leak, $p = 0.8$)*	0	0	0	8	0	0	0
	Total hours *(2 leaks, $p = 0.1$)*	0	0	8	8	0	0	0

[a] If a cell is highlighted, then the crew (column) is assigned a leak during that scenario (row).

Table D.27: Optimal crew assignment with leak distribution 3.

		Crew 1	Crew 2	Crew 3	Crew 4	Crew 5	Crew 6	Crew 7
Standard jobs	Total # jobs	2	3	2	1	2	2	3
	Total hours	3.8	11.6	11.8	3.8	11.5	11.1	11.8
Gas leak jobs	Total hours *(0 leak, $p = 0.2$)*	0	0	0	0	0	0	0
	Total hours *(1 leak, $p = 0.6$)*	8	0	0	0	0	0	0
	Total hours *(2 leaks, $p = 0.2$)*	8	0	0	8	0	0	0

[a] If a cell is highlighted, then the crew (column) is assigned a leak during that scenario (row).

Table D.28: Optimal crew assignment with leak distribution 4.

		Crew 1	Crew 2	Crew 3	Crew 4	Crew 5	Crew 6	Crew 7
Standard jobs	Total # jobs	2	1	2	2	3	2	3
	Total hours	3.8	3.8	11.5	11.1	11.6	11.8	11.7
Gas leak jobs	Total hours *(0 leak, $p = 0.4$)*	0	0	0	0	0	0	0
	Total hours *(1 leak, $p = 0.2$)*	8	0	0	0	0	0	0
	Total hours *(2 leaks, $p = 0.4$)*	8	8	0	0	0	0	0

[a] If a cell is highlighted, then the crew (column) is assigned a leak during that scenario (row).

Table D.29: Optimal crew assignment with leak distribution 5.

		Crew 1	Crew 2	Crew 3	Crew 4	Crew 5	Crew 6	Crew 7
Standard jobs	Total # jobs	3	1	2	2	2	3	2
	Total hours	10.6	2.7	10.2	10.4	10.5	10.3	10.7
Gas leak jobs	Total hours *(0 leak, $p = 0.3$)*	0	0	0	0	0	0	0
	Total hours *(1 leak, $p = 0.5$)*	0	8	0	0	0	0	0
	Total hours *(2 leaks, $p = 0.1$)*	0	8	8	0	0	0	0
	Total hours *(3 leaks, $p = 0.1$)*	0	8	8	0	0	8	0

[a] If a cell is highlighted, then the crew (column) is assigned a leak during that scenario (row).

Table D.30: Optimal crew assignment with leak distribution 6.

		Crew 1	Crew 2	Crew 3	Crew 4	Crew 5	Crew 6	Crew 7
Standard jobs	Total # jobs	1	2	3	2	2	3	2
	Total hours	3.8	11.0	11.8	3.8	11.5	11.7	11.8
Gas leak jobs	Total hours *(0 leak, $p = 0.4$)*	0	0	0	0	0	0	0
	Total hours *(1 leak, $p = 0.3$)*	0	0	0	8	0	0	0
	Total hours *(2 leaks, $p = 0.2$)*	8	0	0	8	0	0	0
	Total hours *(3 leaks, $p = 0.1$)*	8	8	0	8	0	0	0

[a] If a cell is highlighted, then the crew (column) is assigned a leak during that scenario (row).

Table D.31: Optimal crew assignment with leak distribution 7.

		Crew 1	Crew 2	Crew 3	Crew 4	Crew 5	Crew 6	Crew 7
Standard jobs	Total # jobs	3	2	1	2	2	3	2
	Total hours	11.4	8.1	3.5	11.4	8.1	11.4	11.4
Gas leak jobs	Total hours *(0 leak, $p = 0.5$)*	0	0	0	0	0	0	0
	Total hours *(1 leak, $p = 0.2$)*	0	0	8	0	0	0	0
	Total hours *(2 leaks, $p = 0.1$)*	0	0	8	0	8	0	0
	Total hours *(3 leaks, $p = 0.2$)*	0	8	8	0	8	0	0

[a] If a cell is highlighted, then the crew (column) is assigned a leak during that scenario (row).

Table D.32: Crew assignment using Algorithm Stoch-LPT with leak distribution 1.

		Crew 1	Crew 2	Crew 3	Crew 4	Crew 5	Crew 6	Crew 7
Standard jobs	Total # jobs	1	3	3	2	2	2	2
	Total hours	2.6	11.5	11.4	9.5	9.7	10.3	10.3
Gas leak jobs	Total hours *(1 leak, $p = 1$)*	8	0	0	0	0	0	0

[a] If a cell is highlighted, then the crew (column) is assigned a leak during that scenario (row).

Table D.33: Crew assignment using Algorithm Stoch-LPT with leak distribution 2.

		Crew 1	Crew 2	Crew 3	Crew 4	Crew 5	Crew 6	Crew 7
Standard jobs	Total # jobs	1	3	2	2	3	2	2
	Total hours	2.7	10.2	10.1	10.2	11.4	10.3	10.4
Gas leak jobs	Total hours *(0 leak,* $p = 0.1$*)*	0	0	0	0	0	0	0
	Total hours *(1 leak,* $p = 0.8$*)*	8	0	0	0	0	0	0
	Total hours *(2 leaks,* $p = 0.1$*)*	8	8	0	0	0	0	0

[a] If a cell is highlighted, then the crew (column) is assigned a leak during that scenario (row).

Table D.34: Crew assignment using Algorithm Stoch-LPT with leak distribution 3.

		Crew 1	Crew 2	Crew 3	Crew 4	Crew 5	Crew 6	Crew 7
Standard jobs	Total # jobs	1	3	2	2	3	2	2
	Total hours	2.7	10.2	10.1	10.2	11.4	10.3	10.4
Gas leak jobs	Total hours *(0 leak, $p = 0.2$)*	0	0	0	0	0	0	0
	Total hours *(1 leak, $p = 0.6$)*	8	0	0	0	0	0	0
	Total hours *(2 leaks, $p = 0.2$)*	8	8	0	0	0	0	0

[a] If a cell is highlighted, then the crew (column) is assigned a leak during that scenario (row).

Table D.35: Crew assignment using Algorithm Stoch-LPT with leak distribution 4.

		Crew 1	Crew 2	Crew 3	Crew 4	Crew 5	Crew 6	Crew 7
Standard jobs	Total # jobs	1	1	3	2	3	3	2
	Total hours	5.0	5.4	11.9	9.9	11.7	11.2	10.3
Gas leak jobs	Total hours *(0 leak, $p = 0.4$)*	0	0	0	0	0	0	0
	Total hours *(1 leak, $p = 0.2$)*	8	0	0	0	0	0	0
	Total hours *(2 leaks, $p = 0.4$)*	8	8	0	0	0	0	0

[a] If a cell is highlighted, then the crew (column) is assigned a leak during that scenario (row).

Table D.36: Crew assignment using Algorithm Stoch-LPT with leak distribution 5.

		Crew 1	Crew 2	Crew 3	Crew 4	Crew 5	Crew 6	Crew 7
Standard jobs	Total # jobs	1	3	3	2	2	2	2
	Total hours	2.7	10.2	10.4	10.4	10.7	10.5	10.5
Gas leak jobs	Total hours *(0 leak, $p = 0.3$)*	0	0	0	0	0	0	0
	Total hours *(1 leak, $p = 0.5$)*	8	0	0	0	0	0	0
	Total hours *(2 leaks, $p = 0.1$)*	8	8	0	0	0	0	0
	Total hours *(2 leaks, $p = 0.1$)*	8	8	8	0	0	0	0

[a] If a cell is highlighted, then the crew (column) is assigned a leak during that scenario (row).

Table D.37: Crew assignment using Algorithm Stoch-LPT with leak distribution 6.

		Crew 1	Crew 2	Crew 3	Crew 4	Crew 5	Crew 6	Crew 7
Standard jobs	Total # jobs	1	2	3	2	2	3	2
	Total hours	5.0	7.6	10.7	10.1	10.2	11.4	10.3
Gas leak jobs	Total hours *(0 leak, $p = 0.4$)*	0	0	0	0	0	0	0
	Total hours *(1 leak, $p = 0.3$)*	8	0	0	0	0	0	0
	Total hours *(2 leaks, $p = 0.2$)*	8	8	0	0	0	0	0
	Total hours *(2 leaks, $p = 0.1$)*	8	8	8	0	0	0	0

[a] If a cell is highlighted, then the crew (column) is assigned a leak during that scenario (row).

Table D.38: Crew assignment using Algorithm Stoch-LPT with leak distribution 7.

		Crew 1	Crew 2	Crew 3	Crew 4	Crew 5	Crew 6	Crew 7
Standard jobs	Total # jobs	2	2	3	2	2	2	2
	Total hours	7.2	8.0	9.6	10.1	10.2	9.9	10.3
Gas leak jobs	Total hours *(0 leak, $p = 0.5$)*	0	0	0	0	0	0	0
	Total hours *(1 leak, $p = 0.2$)*	8	0	0	0	0	0	0
	Total hours *(2 leaks, $p = 0.1$)*	8	8	0	0	0	0	0
	Total hours *(2 leaks, $p = 0.2$)*	8	8	8	0	0	0	0

[a] If a cell is highlighted, then the crew (column) is assigned a leak during that scenario (row).

Table D.39: Data for job types used for simulations.

Job Type	Average job duration (hours)	Probability of daily quota				Average daily quota (no. jobs)	Stdev daily quota (no. jobs)
		0	1	2	3		
LEAK2A	6.24	0.93	0.07	0	0	0.07	0.25
CMP	5.85	0.45	0.4	0.12	0.03	0.71	0.79
PVIP	2.50	0.45	0.4	0.12	0.03	0.71	0.79
LKTPDP	5.45	0.73	0.28	0	0	0.28	0.45
LEAK2	6.76	0.45	0.4	0.12	0.03	0.71	0.79
LEAK3	7.80	0.46	0.54	0	0	0.54	0.5
CUSTREQ	6.24	0.05	0.1	0.1	0.75	2.86	0.86
SRP	6.34	0.45	0.4	0.12	0.03	0.71	0.79
MPP	3.87	0.89	0.11	0	0	0.11	0.31
LKEMER	8.79	—	—	—	—	—	—

[a] LKEMER refers to an emergency job.

Bibliography

Ahmed, S. 2010. *Two-Stage Stochastic Integer Programming: A Brief Introduction*. Wiley Encyclopedia of Operations Research and Management Science, vol. 8. John Wiley & Sons, Inc.

Akcay, A., Biller, B., & Tayur, S. 2009. *Improved Inventory Targets in the Presence of Limited Historical Demand Data*. Tepper School of Business Working Paper, Carnegie Mellon University, Pittsburgh, PA.

An, M.Y. 1995. *Log-concave probability distributions: theory and statistical testing*. Technical Report, Economics Department, Duke University, Durham, NC.

Asmussen, S., & Glynn, P. 2007. *Stochastic simulation: algorithms and analysis*. New York: Springer. Chapter 4, pages 77–80.

Badinelli, R. D. 1990. The inventory costs of common misspecification of demand forecasting models. *International Journal of Production Research*, **28**(12), 2321–2340.

Ball, M., & Queyranne, M. 2009. Toward robust revenue management: competitive analysis of online booking. *Operations Research*, **57**, 950–963.

Barnhart, C., Cohn, A. M., Johnson, E. L., Klabjan, D., Nemhauser, G. L., & Vance, P. H. 2003. Airline Crew Scheduling. *Pages 517–560 of:* Hall, Randolph W. (ed), *Handbook of Transportation Science*, second edn. Norwell, MA: Kluwer Academic Publishers.

Bernstein, S. N. 1927. *Theory of Probability*.

Bertsimas, D., & Popescu, I. 2004. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal of Optimization*, **15**(3), 780–804.

Birge, J. R. 1997. Stochastic programming computation and applications. *INFORMS Journal on Computing*, **9**(2), 111–133.

Bookbinder, J. H., & Lordahl, A. E. 1989. Estimation of Inventory Reorder Level using the bootstrap statistical procedure. *IEE Transactions*, **21**, 302–312.

Burke, Garance. 2010. *Aging gas pipe at risk of explosion nationwide. Associated Press.* Online; accessed 17-July-2012.

Carøe, C.C., & Schultz, R. 1999. Dual decomposition in stochastic integer programming. *Operations Research Letters*, **24**, 37–45.

Chandra, N. K., & Roy, D. 2001. Some results on reversed hazard rate. *Probability in the Engineering and Informational Sciences*, **15**, 95–102.

Erdős, P., & Lovász, L. 1975. Problems and results on 3-chromatic hypergraphs and some related questions. *Volume 2, pages 609–628 of:* Hajnal, A., Rado, R., & Sós, V. T. (eds), *Infinite and Finite Sets (to Paul Erdős on his 60th birthday).*

Eren, S., & Maglaras, C. 2006. *Revenue Management heuristics under limited market information: A maximum entropy approach.* Presented at The 6th Annual INFORMS Revenue Management Conference, Jun 5-6, Columbia University.

Flaxman, A. D., Kalai, A. T., & McMahan, H. B. 2005. Online convex optimization in the bandit setting: Gradient descent without a gradient. *Pages 385–394 of: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms.* SODA '05.

Gallego, G., & Moon, I. 1993. The distribution free newsboy problem: Review and extensions. *Journal of the Operational Research Society*, **44**(8), 825–834.

Glass, C. A., & Kellerer, H. 2007. Parallel machine scheduling with job assignment restrictions. *Naval Research Logistics*, **54**(3), 250–257.

Glasserman, P., & Ho, Y. C. 1991. *Gradient Estimation via perturbation analysis.* Kluwer Academic Publishers.

Goberna, M. A., & Lopez, M. A. 1998. *Linear semi-infinite optimization.* New York: Wiley.

Godfrey, G., & Powell, W. B. 2002. An adaptive, dynamic programming algorithm for stochastic resource allocation problems, I: single period travel times. *Transportation Science*, **36**(1), 21–39.

Godfrey, G. A., & Powell, W. B. 2001. An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution. *Management Science*, **47**, 1101–1112.

Govindarajulu, Z., & Mason, D. M. 1983. A strong representation for linear combinations of order statistics with application to fixed-width confidence intervals for location and scale parameters. *Scandinavian Journal of Statistics*, **10**, 97–115.

Hoeffding, W. 1963. Probability Inequalities for sums of Bounded Random Variables. *Journal of the American Statistical Association*, **58**, 13–30.

Homem-De-Mello, T. 2000. Monte carlo methods for discrete stochastic optimization. *Pages 95–117 of:* Uryasev, S., & Parados, P. M. (eds), *Stochastic Optimization: Algorithms and Applications.* Norwell, MA: Kluwer Academic Publishers.

Huh, W. T., & Rusmevichientong, P. 2009. A nonparametric asymptotic analysis of inventory planning with censored data. *Mathematics of Operations Research*, **34**, 103–123.

Huh, W. T., Levi, R., Rusmevichientong, P., & Orlin, J. 2008. *Adaptive data-driven inventory control policies based on Kaplan-Meier estimator.* Working paper.

Huh, W. T., Janakiraman, G., Muckstadt, J. A., & Rusmevichientong, P. 2009. Asymptotic Optimality of Order-up-to Policies in Lost Sales Inventory Systems. *Management Science*, **55**, 404–420.

Huh, W.T, Liu, N., & Truong, V.-A. 2013. Multiresource Allocation Scheduling in Dynamic Environments. *Manufacturing & Service Operations Management.*

Hwang, H.-C., Chang, S. Y., & Lee, K. 2004. Parallel machine scheduling under a grade of service provision. *Computers and Operations Research*, **31**(12), 2055–2061.

Kafura, D. G., & Shen, V. Y. 1977. Task scheduling on a multiprocessor system with independent memories. *SIAM Journal on Computing*, **6**(1), 167–187.

Kleywegt, A. J., Shapiro, A., & Homem-De-Mello, T. 2001. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, **12**, 479–502.

Lamiri, M., Xie, X., Dolgui, A., & Grimaud, F. 2008. A stochastic model for operating room planning with elective and emergency demand for surgery. *European Journal of Operational Research*, **185**(3), 1026–1037.

Laporte, G., & Louveaux, F. V. 1993. The integer L-shaped method for stochastic integer programs with complete recourse. *Operations Research Letters*, **13**(3), 133–142.

Lenstra, J. K., Shmoys, D. B., & Tardos, E. 1990. Approximation algorithms for scheduling unrelated parallel machines. *Mathematical Programming*, **46**, 259–271.

Leung, J. Y.-T., & Li, C.-L. 2008. Scheduling with processing set restrictions: A survey. *International Journal of Production Economics*, **116**, 251–262.

Levi, R., Roundy, R., & Shmoys, D. B. 2007. Provably Near-Optimal Sampling-Based Policies for Stochastic Inventory Control Models. *Mathematics of Operations Research*, **32**(4), 821–839.

Levi, R., Perakis, G., & Uichanco, J. 2013a. *The data-driven newsvendor problem: new bounds and insights.* Massachusetts Institute of Technology. Working Paper.

Levi, R., Perakis, G., & Uichanco, J. 2013b. *Regret optimization for stochastic inventory models with spread information.* Massachusetts Institute of Technology. Working paper.

Liyanage, L. H., & Shanthikumar, J. G. 2005. A practical inventory control policy using operational statistics. *Operations Research Letters*, **33**, 341–348.

McDiarmid, C. 1989. On the method of bounded differences. *Pages 148–188 of: Surveys in Combinatorics.* Cambridge: Cambridge University Press.

Ou, J., Leung, J. Y.-T., & Li, C.-L. 2008. Scheduling parallel machines with inclusive processing set restrictions. *Naval Research Logistics*, **55**(4), 328–338.

Perakis, G., & Roels, G. 2008. Regret in the Newsvendor Model with Partial Information. *Operations Research*, **56**, 188–203.

Pinedo, M., & Weiss, G. 1979. Scheduling stochastic tasks on two parallel processors. *Naval Research Logistics Quarterly*, **26**, 527–536.

Pinedo, M. L. 2002. *Scheduling: Theory, Algorithms, and Systems.* Second edn. Upper Saddle River, NJ: Prentice-Hall.

Pipeline & Hazardous Materials Safety Administration. 2011a. *Facts & Stats – Pacific Gas & Electric Pipeline Rupture in San Bruno, CA.* Online; accessed 16-July-2012.

Pipeline & Hazardous Materials Safety Administration. 2011b. *Reports & Research – Cast iron pipeline R&D projects.* Online; accessed 18-December-2012.

Raghavan, R., & Thompson, C. D. 1987. Randomized rounding: a technique for provably good algorithms and algorithmic proofs. *Combinatorica*, **7**, 365–374.

Rockafellar, R. T. 1972. *Convex Analysis.* Princeton, NJ: Princeton University Press.

Savage, L. J. 1951. The theory of statistical decisions. *Journal of the American Statistical Association*, **46**(253), 55–67.

Scarf, H. 1959. Bayes Solution to the statistical inventory problem. *Annals of Mathematical Statistics*, **30**(2), 490–508.

Scarf, H. E. 1958. A min-max solution to an inventory problem. *Pages 201–209 of:* K. J. Arrow and S. Karlin and H. E. Scarf (ed), *Studies in Mathematical Theory of Inventory and Production*. Stanford, CA: Stanford Univerity Press.

Sengupta, D., & Nanda, A. 1997. Log-concave and concave distributions in reliability. *Naval Research Logistics*, **46**, 419–433.

Sengupta, D., & Paul, D. 2005. *Some tests for log-concavity of life distributions.* Preprint available at `http://anson.ucdavis.edu/~debashis/techrep/logconca.pdf`.

Shapiro, A. 2008. Stochastic programming approach to optimization under uncertainty. *Mathematical Programming*, **112**, 183–220.

Sherali, H. D., & Fraticelli, B. M. P. 2002. A modification of Benders' decomposition algorithm for discrete subproblems: An approach for stochastic programs with integer recourse. *Journal of Global Optimization*, **22**, 319–342.

Smith, S.A., & Agrawal, N. 2000. Management of multi-item retail inventory systems with demand substitution. *Operations Research*, **48**(1), 50–64.

Srinivasan, A. 1996. An Extension of the Lovász Local Lemma and its applications to integer programming. *Pages 6–15 of: Proc. of the seventh annual ACM-SIAM symposium on discrete algorithms.*

Stigler, S. M. 1974. Linear Functions of Order Statistics with Smooth Weight Functions. *The Annals of Statistics*, **2**(4), 676–693.

Swamy, C., & Shmoys, D. B. 2005. Sampling-based approximation algorithms for multi-stage stochastic optimization. *In: Proceedings of the 46th Annual IEEE Symposium on the Foundations of Computer Science.*

Weber, R. R. 1982. Scheduling jobs with stochastic processing requirements on parallel machines to minimize makespan or flow time. *Journal of Applied Probability*, **19**, 167–182.

Yue, J., Chen, B., & Wang, M. C. 2006. Expected value of distribution information for the newsvendor problem. *Operations Research*, **54**(6), 1128–1136.

Zipkin, P. H. 2000. *Foundations of Inventory Management.* New York: The McGraw-Hill Companies, Inc.