PERFORMANCE OF THE SMALLEST-VARIANCE-FIRST RULE IN APPOINTMENT SEQUENCING

MADELON A. DE KEMP, MICHEL MANDJES, NEIL OLVER

ABSTRACT. A classical problem in appointment scheduling, with applications in health care, concerns the determination of the patients' arrival times that minimize a cost function that is a weighted sum of mean waiting times and mean idle times. Part of this problem is the sequencing problem, which focuses on ordering the patients. We assess the performance of the *smallest-variance-first* (SVF) rule, which sequences patients in order of increasing variance of their service durations. While it was known that SVF is not always optimal, many papers have found that it performs well in practice and simulation. We give theoretical justification for these observations by proving quantitative *worst-case* bounds on the ratio between the cost incurred by the SVF rule and the minimum attainable cost, in a number of settings. We also show that under quite general conditions, this ratio approaches 1 as the number of patients grows large, showing that the SVF rule is asymptotically optimal. While this viewpoint in terms of approximation ratio is a standard approach in many algorithmic settings, our results appear to be the first of this type in the appointment scheduling literature.

SUBJECT CLASSIFICATION. Health care: appointment scheduling. Scheduling: stochastic appointment sequencing. Optimization: approximation algorithm.
AREA OF REVIEW. Stochastic Models.

AFFILIATIONS. M. de Kemp and M. Mandjes are with Korteweg-de Vries Institute, University of Amsterdam. N. Olver is with the Department of Econometrics & Operations Research, Vrije Universiteit Amsterdam, and is also affiliated with CWI, Amsterdam. The research for this paper is partly funded by the NWO Gravitation Programme NETWORKS, Grant Number 024.002.003 (de Kemp, Mandjes) and NWO Vidi grant 016.Vidi.189.087 (Olver).

Date: December 5, 2018.

1. Introduction

Setting up appointment schedules plays an important role in health care and various other domains. The main challenge lies in efficiently running the system, but at the same time providing the customers an acceptable level of service. The service level can be expressed in terms of the waiting times the customers are facing, and the system efficiency in terms of the service provider's idle time. The problem of generating an optimal schedule is generally formulated as minimizing a cost function (or simply "cost") that is a weighted average of the expected idle time and the expected waiting times. As most literature on this topic focuses on applications in health care, we refer throughout this paper to customers as patients, and to the server as the doctor.

The problem of scheduling appointments can be split into two parts: one needs to determine the amount of time scheduled for each appointment, and one needs to determine in which order the patients should arrive. These problems are usually referred to as the *scheduling* problem and *sequencing* problem, respectively. This paper will focus on the sequencing problem (and later, the combined sequencing and scheduling problem), in a context with a single doctor seeing a sequence of patients. We impose the common assumptions that the service times of the patients form a sequence of independent random variables, while they arrive punctually at the scheduled times (which we will refer to as *epochs*). In this setting, a variety of techniques is available that determines for a given order the optimal arrival epochs; see, e.g., [2, 24] and references therein. However, much less is known about the efficient computation of "good" sequences. Already for a relatively modest number of patients, the number of possible sequences is huge, thus seriously complicating the search for an optimal order. An appointment scheduling review paper from 2017 [1] states that the optimal sequencing problem is one of the main open problems in the area:

> "[...] one of the biggest challenges for future research is to find optimal (or near-optimal) solutions to more realistic appointment sequencing problems."

A number of papers consider the sequencing (or combined sequencing and scheduling) problem and develop various stochastic programming models for it [3, 9, 26, 28]. However, the resulting optimization problems are very difficult to solve. Variants of the problem have been shown to be NP-hard [23, 28], indicating that this difficulty is inherent.

A more popular approach has been to consider simple heuristics for the sequencing problem. The most frequently used heuristic is to order the patients by the variance of their service times, from smallest to largest. Throughout this paper we refer to this sequence as the SVF (smallest-variance-first) sequence. The intuition for using the SVF sequence is that an unusually long service time in the beginning could cause many later patients to have to wait, and the SVF sequence aims to reduce the risk of this occurring. This is a very simple and appealing rule, and only requires the evaluation of the variances of the service times. It has been observed by simulation that the SVF rule typically performs very well, often even optimally. It was proven that it is optimal for two patients under some distributional assumptions [13, 36]. Recently, however, Kong et al. [23] provided instances showing it need not be optimal, even for simple cases with uniform or lognormal service times and a substantial number of patients.

Despite the SVF sequence appearing promising in simulations, little is known about its theoretical performance, or of any other simple heuristic for that matter. In this paper, we propose a new direction of research for the sequencing problem: *finding sequences that provably perform well.* Instead of finding an optimal sequence, such research aims at finding performance bounds on easily-computed sequences. Considering previous research, the SVF sequence is the obvious candidate for such an easily-computed and well-performing sequence, and will therefore be our focus. The precise quantity of interest to us will be the *ratio* of the cost of the schedule coming from the SVF sequence, and the cost of the schedule coming from the optimal sequence.

Our main goal in this paper is to prove upper bounds on this ratio – known as the *approximation ratio* – in various settings. This direction of study is very standard in the algorithmic community when considering intractable (NP-hard) problems, for example in machine scheduling (see [15, 16, 34] and references therein). However, it has not been studied in the appointment sequencing context. Note that for *typical* problem instances the SVF sequence could perform significantly better than suggested by an upper bound on the approximation ratio, as the bound must also hold for *worst-case* instances.

1.1. **Main contributions.** We first concentrate exclusively on the effect of the sequence, using the simplest choice of schedule: each patient is assigned a slot of length equal to its mean service time. In other words, the arrival time of any patient is set equal to the sum of the mean service times of all preceding patients. This is certainly not the optimal solution to the scheduling problem, but it has the advantage of being very simple and easily applicable, and also completely independent of the choice of tradeoff in the cost function between doctor idle time and patient waiting time. As was stated in, e.g., the survey paper [1] and in [11], this "mean-based" type of schedule is a commonly used approach in practice.

Under the mean-based scheduling rule, we prove a number of results. Under an assumption (namely that the service-time distributions are comparable according to a certain ordering), we prove in Section 3 that the approximation ratio of SVF is at most 2 for symmetric service-time distributions, and at most 4 in general. In other words, we show that *for all instances* (i.e., for all numbers of patients and all service-time distributions satisfying the assumption imposed) the SVF cost is at most four times the optimal cost. We also consider two special cases:

- Service times are evidently nonnegative, but one could consider the situation that normal distributions are used as an approximation of the actual distributions of service times. In Section 3.2, we prove that then the approximation ratio is at most $4(\sqrt{2}-1) \approx 1.6569$. While we do not believe that our result here is sharp, it indicates that the performance of SVF for well-behaved service-time distributions is most likely substantially better than suggested by the bounds 2 and 4 mentioned above.
- In Section 3.3 we bridge the gap between the upper bound of 2 for symmetric distributions and the general upper bound of 4, by developing a method that isolates the effect of asymmetry. For the lognormal distributions fitted to real data in Çayırlı et al. [6], this method results in an approximation ratio of at most 3.43.

In Section 4, we consider the combined sequencing and scheduling problem. Here, we wish to compare a heuristic for this combined problem to the overall optimal schedule, over all possible sequences and schedules. Observe that the simple mean-based scheduling rule may lead to high cost, because waiting times could easily propagate. We therefore consider a simple alternative scheduling rule, suggested by Charnetski [8]: the slot assigned to a patient is equal to its mean service time, plus some multiple α of the standard deviation of its service time (where this α is optimized). Again under some assumptions, we show that this scheduling rule, combined with the SVF sequencing rule, yields a cost that is (relative to the optimal cost) off by at most a constant factor.

We also consider the special case of lognormally distributed service times, as these are often seen in practice [6, 21]. Using a slightly different scheduling heuristic (the interarrival time being a multiple of the mean service time), we find an upper bound on the approximation ratio. Applying this result to the data in Çayırlı et al. [6], we find an upper bound of 2.90 in the case that in the cost function the waiting and idle times are equally important.

In Section 5, we return to the mean-based setting of Section 3. We show that as the number of patients grows large, the approximation ratio tends to 1. This result requires only a very weak assumption on the service-time distributions. The important practical implication of this result is that SVF is close to optimal in settings where the number of patients is substantial.

Finally, in Section 6, we give an example that demonstrate that the assumptions made in previous sections are necessary: without any restrictions on the service-time distributions, no bound on the approximation ratio of SVF is possible. This holds true also when optimal rather than mean-based schedules are used. This example involves only two patients. Then we give a (still relatively straightforward) example that shows that it is impossible to obtain a better bound than 1.28 on the approximation ratio in the mean-based setting of Section 3 (i.e., the setting for which we found the upper bound of 4).

1.2. **Further related work.** Here we will mention some of the most relevant literature for this paper. For more extensive reviews on the appointment scheduling and sequencing literature, we refer the reader to, e.g., Ahmadi-Javid et al. [1], Çayırlı and Veral [5], and Gupta and Denton [14].

As already noted, Kong et al. [23] showed that SVF is not in general optimal. In some very specific cases, optimality of SVF has been demonstrated. For only two patients, the SVF sequence is optimal when the service times are both exponentially distributed or both uniformly distributed [36], or more generally, when the two service times are comparable according to a certain convex ordering [13]. For three patients, Kong et al. [23] find sufficient conditions for the SVF sequence to be optimal, when the time scheduled for each appointment is equal to the mean service time. (We have verified that this result can be extended to four patients using the same methods.)

Kemper et al. [18] analyze a sequential optimization approach, meaning that the arrival time of a patient is optimized without taking into account its impact on later patients. They show that under this rather different notion of optimality, and if the service times come from the same scale-family, then SVF does provide the best ordering.

One line of research focuses on comparing various sequencing heuristics (including SVF) through simulation. Denton, Viapiano and Vogl [10] consider a model similar to ours, and discuss the effectiveness of a number of simple sequencing heuristics using simulation, based on real surgery data. The SVF heuristic performed best of all the heuristics they considered. Mak, Rong and Zhang [26] consider a model where waiting time costs may be different for different patients; by studying some more tractable approximations, they also find that SVF performs well.

Klassen and Rohleder [21] and Rohleder and Klassen [32] consider an appointment scheduling model where not all patient information is known in advance; rather, patients must be scheduled as they call in to make an appointment (and so without information about patients who call later). Once again, it was empirically found that it worked best to put patients with low-variance service times early in the schedule.

A number of works model variants of the combined sequencing and scheduling problem as stochastic integer or linear programming problems. Solving these programs is very challenging however, and generally exact results were only obtained for small instances. Works along these lines include Denton and Gupta [9], Mancilla and Storer [28] and Berg et al. [3]. For larger instances, it was necessary to resort to heuristics such as SVF for the sequencing problem. We mention Vanden Bosch and Dietz [35] who propose instead a local search heuristic to iteratively improve the sequence by finding pairs of patients who can be swapped to improve the solution.

There are also a number of papers which take a robust optimization approach [22, 29, 27]. Here, instead of working with explicitly given service-time distributions, the goal is to find a schedule minimizing the worst-case expected cost given only that the distributions meet certain constraints (such as certain given moments). Most relevant to us, Mak et al. [27] discuss one such robust model, and are able to prove that under mild assumptions SVF is optimal in this context. In their model, the joint distribution of the service times could be any distribution matching known moments for individual service times (e.g., the means and variances). However, the worst-case distributions corresponding to the optimal schedule are typically highly correlated; these results do not carry over to a model where independence is assumed. Mittal et al. [29] discuss another robust model, in which each service time can take any value in a certain interval. They find a $(2+\epsilon)$-approximation algorithm for the combined scheduling and sequencing problem.

Finally, we would like to point out the relation with machine scheduling (see the book by Pinedo [30] for more background). The main difference between machine scheduling and appointment scheduling is that in the former the arrival times of jobs/patients are given, while in the latter these are decision variables. The machine scheduling problem most closely related to our problem can be found in Guda et al. [12]. In this paper, the due dates and sequence of jobs need to be minimized, in order to minimize a weighted average of expected earliness and tardiness around the due dates. The SVF rule is optimal in the model of Guda et al., under some assumption on the service times of jobs. However, in their model all jobs are present from the start and so that there is no idle time. Compared to our model, this greatly simplifies the expression for the cost function, which facilitates finding an optimal solution.

2. Model and preliminaries

Consider a problem instance with n patients, numbered 1 up to n. We denote the service time of patient i in this problem instance by B_i, which has mean μ_i and variance σ_i^2. As pointed out in the introduction, one should distinguish between the scheduling problem and the sequencing problem. The sequencing problem, on which we primarily focus, is to decide which patient is assigned which appointment slot. The sequence is denoted by a permutation $\tau \in \mathsf{S}_n$ (where S_n denotes the set of all permutations on $\{1, \ldots, n\}$). The value $\tau(i)$ will denote the index of the patient that is assigned to appointment slot i. The scheduling problem is to decide the interarrival times between patients, given the sequence in which they arrive. We use x_j to denote the interarrival time between patient j and the next patient, i.e., the length of the appointment slot reserved for patient j. The vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ will be referred to as the schedule.

Let W_i denote the waiting time of the patient in appointment slot i. Let I_i be the idle time before the start of appointment slot i after the previous patient has been served. Given a sequence τ and interarrival times x_i, the waiting times and idle times can be computed using the Lindley recursions [25], which read

$$W_{i+1} = (W_i + B_{\tau(i)} - x_{\tau(i)})^+, \quad I_{i+1} = (W_i + B_{\tau(i)} - x_{\tau(i)})^-, \tag{1}$$

using the notation $x^+ := \max\{0, x\}$ and $x^- := \max\{0, -x\}$.

We use a parameter $\omega \in (0, 1)$ to indicate the relative importance of idle time and waiting time. As a cost function, we seek to minimize

$$C_\omega(\tau, \boldsymbol{x}) := \omega \sum_{i=1}^{n} \mathbb{E} I_i + (1 - \omega) \sum_{i=1}^{n} \mathbb{E} W_i, \tag{2}$$

a weighted average of the expected total idle time and expected total waiting time. Observe that this cost function still depends on the sequence τ, on the schedule $\boldsymbol{x}$, and on the patient service-time distributions $\boldsymbol{B} = (B_1, \ldots, B_n)$. We generally suppress the dependence on $\boldsymbol{B}$, but we may write $C_\omega(\boldsymbol{B}, \tau, \boldsymbol{x})$ if we wish to be explicit. As an aside, we mention that an approach to estimate ω in a practical context can be found in [31].

Throughout this paper, we assume the patients are indexed such that $\sigma_1^2 \leqslant \sigma_2^2 \leqslant \ldots \leqslant \sigma_n^2$. The SVF sequence is then the sequence given by the identity permutation id given by $\mathrm{id}(i) = i$. The waiting times and idle times under this sequence are denoted as W_i^{SVF} and I_i^{SVF} respectively. We compare this sequence with the sequence that minimizes (2). The waiting times and idle times under this optimal sequence are denoted by W_i^{OPT} and I_i^{OPT} respectively.

To compare these sequences, we study the ratio between the cost functions under the SVF sequence and the optimal sequence. If this ratio is small, then this is evidence that the SVF sequence performs well. We do so in two settings. In one setting, the schedule is restricted to be the mean-based schedule given by $\boldsymbol{x} = \boldsymbol{\mu}$. We then consider the approximation ratio

$$\varrho_\omega(\boldsymbol{B}) := \frac{C_\omega(\boldsymbol{B}, \mathrm{id}, \boldsymbol{\mu})}{\min\{C_\omega(\boldsymbol{B}, \tau, \boldsymbol{\mu}) : \tau \in \mathsf{S}_n\}} = \frac{\omega \sum_{i=1}^n \mathbb{E}I_i^{\mathrm{SVF}} + (1-\omega)\sum_{i=1}^n \mathbb{E}W_i^{\mathrm{SVF}}}{\omega \sum_{i=1}^n \mathbb{E}I_i^{\mathrm{OPT}} + (1-\omega)\sum_{i=1}^n \mathbb{E}W_i^{\mathrm{OPT}}}.$$

We will write just ϱ_ω when the service-time distributions under consideration are unambiguous.

In the second setting, we compare the SVF sequence along with a given schedule $\boldsymbol{x}$ with the optimal combination of sequence and schedule. We then use the notation W_i^{SVF} and I_i^{SVF} for the combination of the SVF sequence and the given schedule $\boldsymbol{x}$, and we use the notation W_i^{OPT} and I_i^{OPT} for the combination of sequence and schedule that minimizes (2). We then consider the approximation ratio

$$r_\omega(\boldsymbol{B}, \boldsymbol{x}) := \frac{C_\omega(\boldsymbol{B}, \mathrm{id}, \boldsymbol{x})}{\min\{C_\omega(\boldsymbol{B}, \tau, \boldsymbol{y}) : \tau \in \mathsf{S}_n, \boldsymbol{y} \in \mathbb{R}_+^n\}} = \frac{\omega \sum_{i=1}^n \mathbb{E}I_i^{\mathrm{SVF}} + (1-\omega)\sum_{i=1}^n \mathbb{E}W_i^{\mathrm{SVF}}}{\omega \sum_{i=1}^n \mathbb{E}I_i^{\mathrm{OPT}} + (1-\omega)\sum_{i=1}^n \mathbb{E}W_i^{\mathrm{OPT}}}.$$

Once again, we will omit $\boldsymbol{x}$ and $\boldsymbol{B}$ when their choice is unambiguous.

In this paper, we prove, under some assumptions, upper bounds on ϱ_ω and r_ω. Such an upper bound then guarantees that the SVF sequence always has a cost function of at most such an upper bound times the optimal cost function. We also show, under some condition, that $\varrho_\omega(\boldsymbol{B})$ converges to 1 as the number of patients tends to infinity, thus proving that the SVF sequence is asymptotically optimal when mean-based schedules are used.

Remark 2.1. Service times are inherently nonnegative, but our framework (based on the Lindley recursions (1)) carries over to situations where the B_i are allowed to take negative values. This might be useful if the true distributions of service times can be approximated using distributions that can take negative values (with some small probability), for example normal distributions. If such distributions that can take negative values form a good fit to the data in some application, the theoretical performance of the SVF rule for these distributions gives some indication for the performance of the SVF rule in this application.

2.1. **Preliminaries.** We need the following well-known results concerning the waiting and idle times. It follows by iterating the Lindley recursion (1) that W_{k+1} is the maximum of a random walk with steps $B_{\tau(k)} - x_{\tau(k)}, B_{\tau(k-1)} - x_{\tau(k-1)}, \ldots, B_{\tau(1)} - x_{\tau(1)}$, that is,

$$W_{k+1} = \max\left\{0, \max\left\{\sum_{i=j}^{k} B_{\tau(i)} - x_{\tau(i)} : j \in \{1, \ldots, k\}\right\}\right\}. \tag{3}$$

In the setting of mean-based schedules $\boldsymbol{x} = \boldsymbol{\mu}$, we introduce the notation $X_i := B_i - \mu_i$, and the random walk

$$S_j := \sum_{i=1}^{j} X_{\tau(k-i+1)}.$$

We then find for the mean-based schedule that

$$W_{k+1} = \max\{0, S_1, \dots, S_k\}. \tag{4}$$

Computing the total time until all patients have been served in two separate ways, we find the identity

$$\sum_{i=1}^{n} I_i + \sum_{i=1}^{n} B_i = \sum_{i=1}^{n-1} x_{\tau(i)} + W_n + B_{\tau(n)}. \tag{5}$$

For a given schedule, this relation can be used to express the expected total idle time in the expected waiting time of the last patient. Therefore, we can focus on the waiting times, and derive results for the idle time from (5).

We also need the concept of a convex ordering on random variables. More information on the convex ordering and related concepts can be found in Shaked and Shanthikumar [33].

Definition 2.2. The random variable A is said to be smaller in the convex order than the random variable B if $\mathbb{E}\phi(A) \leqslant \mathbb{E}\phi(B)$ for all convex functions $\phi : \mathbb{R} \to \mathbb{R}$ for which the expectations exist. This will be denoted by $A \leqslant_{\text{cx}} B$. If $A - \mathbb{E}A \leqslant_{\text{cx}} B - \mathbb{E}B$, then A is said to be smaller than B in the dilation order, denoted as $A \leqslant_{\text{dil}} B$.

Note that $A \leqslant_{\text{cx}} B$ implies $A \leqslant_{\text{dil}} B$, and $A \leqslant_{\text{dil}} B$ implies $\text{Var}A \leqslant \text{Var}B$. The following lemma [33] is useful when checking whether given random variables satisfy a convex order.

Lemma 2.3. *The random variables A and B satisfy $A \leqslant_{\text{cx}} B$ if and only if there exists a coupling $\widehat{A} \overset{\text{d}}{=} A$ and $\widehat{B} \overset{\text{d}}{=} B$ such that* $\mathbb{E}\big[\widehat{B}\big|\widehat{A}\big] = \widehat{A}$.

3. Bounds on performance under mean-based schedules

In this section, we provide bounds for ϱ_ω, the approximation ratio under the mean-based schedule given by $\boldsymbol{x} = \boldsymbol{\mu}$. This amounts to giving an upper bound on the cost function when using the SVF sequence, and a lower bound on the cost function that is valid for any sequence, hence also for the optimal sequence.

This section is structured as follows. In Section 3.1 we prove the main results: Theorem 3.3 and Theorem 3.4. These theorems give bounds on the approximation ratio ϱ_ω, when we assume that the service times are symmetrically distributed and follow a dilation order

(Theorem 3.3), and when we only assume that they follow a dilation order (Theorem 3.4). In Section 3.2 we consider the special case of normally distributed service times. Theorem 3.8 gives an improved bound on ϱ_ω in this case. In Section 3.3 we discuss a method for improving numerically upon the bound of Theorem 3.4; informally, the more symmetric the service-time distributions, the closer the resulting bound is to the value stated in Theorem 3.3.

3.1. **Main results.** We impose the following assumption.

Assumption 3.1 **(ordering).** *We have* $B_1 \leqslant_{\text{dil}} B_2 \leqslant_{\text{dil}} \dots \leqslant_{\text{dil}} B_n$.

We remark that this is the condition under which Gupta [13] proves optimality of SVF for two patients. Note also that this assumption implies $\text{Var}(B_1) \leqslant \dots \leqslant \text{Var}(B_n)$. Examples of instances satisfying this assumption include all B_i having exponential distributions (by Theorem 3.A.18 in [33]), and all B_i having lognormal distributions such that both $\mathbb{E}[\ln B_1] \leqslant \dots \leqslant \mathbb{E}[\ln B_n]$ and $\text{Var}(\ln B_1) \leqslant \dots \leqslant \text{Var}(\ln B_n)$, as proved in Appendix D. In Section 6 it will be shown that this assumption is necessary.

For one of the bounds we prove on ϱ_ω we also make the following assumption.

Assumption 3.2 **(symmetry).** *The* B_i *have symmetric distributions around their mean.*

Examples of instances satisfying both the ordering and symmetry assumption include all B_i having normal distributions, all B_i having uniform distributions and all B_i having Laplace distributions. For all three examples, the ordering assumption follows from Theorem 3.A.18 in [33].

In this section, we prove the following theorems.

Theorem 3.3. *Under the ordering and symmetry assumptions, we have* $\varrho_\omega \leqslant 2$.

Theorem 3.4. *Under the ordering assumption, we have* $\varrho_\omega \leqslant 4$.

A first key point is that to prove these theorems, it suffices to prove bounds on $\mathbb{E}W_{k+1}$ for given k when the first k slots are constrained to contain patients $1, \dots, k$. This is made explicit in the next lemma, proved in Appendix A.

Lemma 3.5. *Let* $\mathbb{E}W_{k+1}^{\text{OPT}'}$ *denote the expected waiting time of the patient in appointment slot* $k+1$*, under the sequence that minimizes this expected waiting time, subject to the constraint that* $\tau(i) \leqslant k$ *for all* $i = 1, \dots, k$*, i.e. the first* k *patients are assigned to the first* k *slots. Suppose* $\mathbb{E}W_{k+1}^{\text{SVF}}/\mathbb{E}W_{k+1}^{\text{OPT}'} \leqslant \varrho'$ *for all* k*. Then, under the ordering assumption,* $\varrho_\omega \leqslant \varrho'$.

The following lemma is another key ingredient.

Lemma 3.6. *Under the symmetry assumption, the random variable W_{k+1} is stochastically dominated by $|S_k|$, and thus*

$$\mathbb{E}W_{k+1} \leqslant \mathbb{E}|S_k| = \mathbb{E}|X_{\tau(1)} + X_{\tau(2)} + \cdots + X_{\tau(k)}|.$$

Proof. Recall that we have $W_{k+1} = \max\{0, S_1, \ldots, S_k\}$ from (4). Under the symmetry assumption, the steps X_i of the random walk S have a symmetric distribution around zero, and hence the same is true for the S_i.

Let $T(a) = \inf\{i : S_i \geqslant a\}$, and note that $\mathbb{P}(W_{k+1} \geqslant a) = \mathbb{P}(T(a) \leqslant k)$. To bound this probability, we look at the random walk reflected in a after $T(a)$. This reflected process $\widehat{S}_i$ is defined by

$$\widehat{S}_i = \begin{cases} S_i & \text{if } i < T(a) \\ 2a - S_i & \text{if } i \geqslant T(a). \end{cases} \tag{6}$$

We have $S_{T(a)} \geqslant a$, so $\widehat{S}_{T(a)} = 2a - S_{T(a)} \leqslant a \leqslant S_{T(a)}$. As the X_i have symmetric distributions, the increments of S_i and $\widehat{S}_i$ for $i \geqslant T(a)$ have the same distribution. Therefore, we see that $\widehat{S}_i$ is stochastically dominated by S_i, for every i. We conclude that $\mathbb{P}(\widehat{S}_k > a) \leqslant \mathbb{P}(S_k > a)$ for all a.

Now note that $W_{k+1} \geqslant a$ implies that either $S_k \geqslant a$ or $\widehat{S}_k = 2a - S_k > a$. As these are disjoint events we now have

$$\mathbb{P}(W_{k+1} \geqslant a) = \mathbb{P}(S_k \geqslant a) + \mathbb{P}(\widehat{S}_k > a) \leqslant \mathbb{P}(S_k \geqslant a) + \mathbb{P}(S_k > a) \leqslant \mathbb{P}(|S_k| \geqslant a).$$

This holds for any $a \geqslant 0$, so W_{k+1} is stochastically dominated by $|S_k|$, as was claimed. □

Proof of Theorem 3.3. As $W_{k+1} = \max\{0, S_1, \ldots, S_k\}$, we have

$$W_{k+1} \geqslant S_k^+ = (X_{\tau(1)} + \cdots + X_{\tau(k)})^+.$$

Note that $\tau(i) \leqslant k$ for all $i \leqslant k$ when we consider $\mathbb{E}W_{k+1}^{\text{OPT}'}$, so now

$$\mathbb{E}W_{k+1}^{\text{OPT}'} \geqslant \mathbb{E}(X_1 + \cdots + X_k)^+. \tag{7}$$

On the other hand, by Lemma 3.6,

$$\mathbb{E}W_{k+1}^{\text{SVF}} \leqslant \mathbb{E}|X_1 + \cdots + X_k| = 2\mathbb{E}(X_1 + \cdots + X_k)^+ \leqslant 2\mathbb{E}W_{k+1}^{\text{OPT}'}.$$

As $\mathbb{E}W_{k+1}^{\text{SVF}}/\mathbb{E}W_{k+1}^{\text{OPT}'}$ is now bounded by 2, Theorem 3.3 follows from Lemma 3.5. □

Proof of Theorem 3.4. Note that Lemma 3.5 and the lower bound (7) are valid without the symmetry assumption being needed. We therefore only need an upper bound on $\mathbb{E}W_{k+1}^{\text{SVF}}$.

Let $X'_1, X'_2, \dots, X'_n$ have the same distributions as respectively $X_1, X_2, \dots, X_n$ such that all these random variables are independent. Let W'_{k+1} be the maximum of the random walk with steps $X_k - X'_k, X_{k-1} - X'_{k-1}, \dots, X_1 - X'_1$. As

$$\mathbb{E}[X_i - X'_i | X_i] = X_i,$$

we see using Lemma 2.3 that $X_i \leqslant_{\text{cx}} X_i - X'_i$. Note that $W_{k+1} = \max\{0, S_1, \dots, S_k\}$ is a convex function in X_i, as it is the maximum of functions linear in X_i. Therefore, each time we replace a step X_i with a step $X_i - X'_i$ the expected maximum of the random walk will increase, so $\mathbb{E}W^{\text{SVF}}_{k+1} \leqslant \mathbb{E}W'_{k+1}$.

Now note that the steps $X_i - X'_i$ all have a symmetric distribution, so we can apply Lemma 3.6 to find

$$\begin{aligned}\mathbb{E}W^{\text{SVF}}_{k+1} &\leqslant \mathbb{E}W'_{k+1} \leqslant \mathbb{E}|X_1 + \cdots + X_k - (X'_1 + \cdots + X'_k)| \\ &\leqslant 2\mathbb{E}|X_1 + \cdots + X_k| = 4\mathbb{E}(X_1 + \cdots + X_k)^+ \leqslant 4\mathbb{E}W^{\text{OPT}'}_{k+1}.\end{aligned}$$

As $\mathbb{E}W^{\text{SVF}}_{k+1}/\mathbb{E}W^{\text{OPT}'}_{k+1}$ is now bounded by 4, the result follows from Lemma 3.5. □

Remark 3.7. In case the scheduled session end time equals the expected total service time, the overtime reads

$$W_{n+1} = \left(W_n + X_{\tau(n)}\right)^+,$$

which can also be included in the cost function. As such, overtime is handled similarly to waiting time, and consequently the results of Theorems 3.3 and 3.4 remain valid when some extra term $c\mathbb{E}W_{n+1}$ with $c > 0$ is added to the cost function.

3.2. **Normally distributed service times.** The results of Theorems 3.3 and 3.4 can be strengthened for specific service-time distributions. One such result is the following.

Theorem 3.8. *When the B_i are all normally distributed we have $\varrho_\omega \leqslant 4(\sqrt{2} - 1)$.*

In order to prove Theorem 3.8, we need the following two lemmas, giving stronger bounds on $\mathbb{E}W^{\text{SVF}}_{k+1}$ and $\mathbb{E}W^{\text{OPT}'}_{k+1}$. The proofs of these lemmas, that hold for any symmetrically distributed service times, can be found in Appendix A.

Lemma 3.9. *Under the symmetry assumption,*

$$\mathbb{E}W_{k+1} \leqslant \mathbb{E}\left(X_{\tau(1)} + \cdots + X_{\tau(k)}\right)^+ + \mathbb{E}\left(X_{\tau(1)} + \cdots + X_{\tau(k-1)}\right)^+.$$

Lemma 3.10. *Under the symmetry assumption, for any ℓ,*

$$\mathbb{E}W_{k+1} \geqslant \frac{1}{2}\Bigg(\mathbb{E}\left(X_{\tau(1)} + \cdots + X_{\tau(k)}\right)^+ + \mathbb{E}\left(X_{\tau(1)} + \cdots + X_{\tau(\ell)}\right)^+ + \mathbb{E}\left(X_{\tau(\ell+1)} + \cdots + X_{\tau(k)}\right)^+\Bigg).$$

Proof of Theorem 3.8. Note that normal distributions satisfy both the ordering and symmetry assumption. Now the sum $X_1 + \cdots + X_i$ again has a normal distribution, with mean zero and variance $\Sigma_i^2 := \sigma_1^2 + \cdots + \sigma_i^2$. For the SVF sequence we now have, using Lemma 3.9, that

$$\mathbb{E}W_{k+1}^{\text{SVF}} \leqslant \frac{1}{\sqrt{2\pi}}\left(\Sigma_k + \Sigma_{k-1}\right). \tag{8}$$

Now we still need an expression for a lower bound on $\mathbb{E}W_{k+1}^{\text{OPT}'}$. Let $\tilde{\Sigma}_i^2 := \sigma_{\tau(1)}^2 + \cdots + \sigma_{\tau(i)}^2$ be the variance of $X_{\tau(1)} + \cdots + X_{\tau(i)}$. From Lemma 3.10 it then follows that

$$\mathbb{E}W_{k+1} \geqslant \frac{1}{2}\left(\tilde{\Sigma}_k + \tilde{\Sigma}_\ell + \sqrt{\tilde{\Sigma}_k^2 - \tilde{\Sigma}_\ell^2}\right).$$

Recall that $\mathbb{E}W_{k+1}^{\text{OPT}'}$ was the optimal expected waiting time when $\tau(i) \leqslant k$ whenever $i \leqslant k$. Therefore, we have $\tilde{\Sigma}_k = \Sigma_k$ and $\sigma_k^2 = \max\{\sigma_{\tau(1)}^2, \ldots, \sigma_{\tau(k)}^2\}$. Now note that

$$\Sigma_k + \tilde{\Sigma}_\ell + \sqrt{\Sigma_k^2 - \tilde{\Sigma}_\ell^2}$$

is largest when $\tilde{\Sigma}_\ell^2$ is as close to $\frac{1}{2}\Sigma_k^2$ as possible. As σ_k^2 is largest of the $\sigma_{\tau(i)}^2$ with $i \leqslant k$, we can always choose ℓ such that

$$\frac{1}{2}\Sigma_{k-1}^2 \leqslant \tilde{\Sigma}_\ell^2 \leqslant \frac{1}{2}\Sigma_{k-1}^2 + \sigma_k^2.$$

This choice of ℓ provides us with the lower bound

$$\mathbb{E}W_{k+1}^{\text{OPT}'} \geqslant \frac{1}{2}\frac{1}{\sqrt{2\pi}}\left(\Sigma_k + \sqrt{\frac{1}{2}\Sigma_{k-1}^2} + \sqrt{\frac{1}{2}\Sigma_{k-1}^2 + \sigma_k^2}\right),$$

valid for any sequence. Comparing with (8), we obtain

$$\frac{\mathbb{E}W_{k+1}^{\text{SVF}}}{\mathbb{E}W_{k+1}^{\text{OPT}'}} \leqslant 2(\Sigma_k + \Sigma_{k-1}) \Bigg/ \left(\Sigma_k + \sqrt{\frac{1}{2}\Sigma_{k-1}^2} + \sqrt{\frac{1}{2}\Sigma_{k-1}^2 + \sigma_k^2}\right).$$

As $\Sigma_k^2 = \Sigma_{k-1}^2 + \sigma_k^2$, this fraction only depends on the relative size of Σ_{k-1}^2 compared to σ_k^2. Suppose that $\Sigma_{k-1}^2 = c\sigma_k^2$, for some $c \geqslant 0$. Then $\Sigma_k^2 = (c+1)\sigma_k^2$, and the fraction becomes

$$\frac{2(\sqrt{c+1} + \sqrt{c})}{\sqrt{c+1} + \sqrt{\frac{1}{2}c} + \sqrt{\frac{1}{2}c + 1}} =: f(c).$$

It can easily be seen that f is increasing, and that $f(c) \to 4(\sqrt{2} - 1)$ as $c \to \infty$.

We now know that $\mathbb{E}W_{k+1}^{\text{SVF}}/\mathbb{E}W_{k+1}^{\text{OPT}'} \leqslant 4(\sqrt{2} - 1) \approx 1.6569$. By Lemma 3.5 the same is then also true for the cost function. This proves Theorem 3.8. □

3.3. **Numerically improving the bound of Theorem 3.4.** Under the ordering assumption, we have proved that $\varrho_\omega \leqslant 4$, and we also proved that $\varrho_\omega \leqslant 2$ when the service times have symmetric distributions. This suggests that an upper bound on ϱ_ω can be found between 2 and 4 for service time distributions that have some degree of symmetry, but are not completely symmetric.

Here we introduce a method to split the service time distributions into a symmetric and a nonsymmetric part, thus isolating the effect of the asymmetry on the upper bound. This can be used to numerically compute an upper bound on ϱ_ω for given problem instances. We do so for lognormal service time distributions that fit real data in Çayırlı et al. [6]. We still impose the ordering assumption.

We introduce the method for continuously distributed service times to simplify the exposition, noting that extending the method to non-continuous distributions is straightforward. Suppose X_i has density f_i. We set $g_i(x) := \min\{f_i(x), f_i(-x)\}$, $h_i(x) := f_i(x) - g_i(x)$ and $p_i := \int_{\mathbb{R}} h_i(x)\mathrm{d}x$. Then we let U_i be a random variable with density $g_i(x)/(1-p_i)$. We let A_i be a random variable, independent of U_i, with density $h_i(x)/p_i$. Let J_i be a Bernoulli variable taking the value one with probability p_i, independent of A_i and U_i. We thus have

$$X_i \stackrel{\mathrm{d}}{=} U_i(1-J_i) + A_iJ_i.$$

Now U_i has a symmetric distribution around zero, so U_i corresponds to the symmetric part of X_i, and A_i to the nonsymmetric part. Note that $\mathbb{E}X_i = 0$ and $\mathbb{E}U_i = 0$, so we must have $\mathbb{E}A_i = 0$. Let A_i' have the same distribution as A_i, so that A_i' is independent of all the other random variables. Since $\mathbb{E}[A_i'J_i|J_i] = 0$, we have

$$\mathbb{E}\left[U_i(1-J_i) + (A_i - A_i')J_i \middle| U_i(1-J_i) + A_iJ_i\right] = U_i(1-J_i) + A_iJ_i.$$

By Lemma 2.3 we conclude

$$X_i \leqslant_{\mathrm{cx}} U_i(1-J_i) + (A_i - A_i')J_i.$$

As W_{k+1} is a convex function in each of the X_i, we can then replace each X_i by this upper bound in convex order to get an upper bound on $\mathbb{E}W_{k+1}^{\mathrm{SVF}}$. Using Lemma 3.6, we then find the following upper bound:

$$\mathbb{E}W_{k+1}^{\mathrm{SVF}} \leqslant \mathbb{E}|X_1 + \cdots + X_k| + \mathbb{E}|A_1J_1 + \cdots + A_kJ_k|.$$

Group	Mean	Standard deviation
Return	15.50	5.038
New	19.09	6.85

Table 1. Parameters of the lognormal distributions fitted by Çayırlı et al. [6].

The upper bound in this proposition can now be compared numerically to the lower bound $\mathbb{E}W_{k+1}^{\text{OPT}'} \geqslant \mathbb{E}(X_1 + \cdots + X_k)^+$, for each k, valid under the ordering assumption. Combining the above with Lemma 3.5, this leads to the bound given in the next theorem.

Theorem 3.11. *Under the ordering assumption, we have*

$$\varrho_\omega \leqslant 2 + 2\max\big\{\mathbb{E}|A_1 J_1 + \cdots + A_k J_k|/\mathbb{E}|X_1 + \cdots + X_k| : k = 1, \ldots, n\big\}. \tag{9}$$

The more symmetric the service times and thus the random variables X_i, the smaller the p_i and hence also the upper bound in (9). When the service times are completely symmetric, the asymmetric parts A_i will be zero, and we recover the upper bound of 2 of Theorem 3.3.

Note that the upper bound in Theorem 3.11 is much easier to numerically compute or simulate than ϱ_ω itself, as for the latter one needs to go over all $n!$ possible sequences to find the optimal one. Also, this method can be used to find an upper bound on ϱ_ω for any problem instance where the service times come from a finite set of distributions and an upper bound on n is given, as illustrated in the next example.

In Çayırlı et al. [6] patients were divided in two groups: new and return patients. For both groups, lognormal distributions were found as a good fit to the data used in the paper, with parameters as shown in Table 1. We checked that problem instances coming from these two distributions satisfy both $\mathbb{E}[\ln B_1] \leqslant \ldots \leqslant \mathbb{E}[\ln B_n]$ and $\text{Var}(\ln B_1) \leqslant \ldots \leqslant \text{Var}(\ln B_n)$, and so satisfy the ordering assumption. It was also mentioned that the doctor that provided the data sees 10 patients per session.

We now consider 11 problem instances: each problem instance consists of 10 patients, with 0 up to 10 of them being a new patient. Computing the upper bound in (9) for each of these instances through simulation, we find that $\varrho_\omega \leqslant 3.43$ for any problem instance consisting of at most 10 patients, with service times that follow one of the two lognormal distributions. Thus, $\varrho_\omega \leqslant 3.43$ for any problem instance the doctor in [6] might face.

4. Bounds on performance for optimal interarrival times

In the previous section we assumed mean-based schedules. We relax this assumption here, in that we consider the performance of the SVF sequence compared to the optimal combination of sequence and schedule. We will again drop the extra subscript n in the notation that we introduced in Section 2.

The goal of this section is to prove bounds on the approximation ratio r_ω. In Section 4.1, we will do so when the service-time distributions are from the same location-scale family, leading to Theorem 4.2. Then, in Section 4.2, we give some examples in which the upper bound of Theorem 4.2 can be explicitly computed. In Section 4.3, we consider the special case of lognormally distributed service times, that are often seen in practice [6, 21]. These lognormal distributions do not come from one location-scale family. A bound on the approximation ratio r_ω is then given in Theorem 4.5.

4.1. **Location-scale family of service times.**

We impose the following assumption.

Assumption 4.1. *The B_i are from the same location-scale family. In other words, there exists a random variable B having mean zero and variance one such that $B_i \stackrel{\mathrm{d}}{=} \mu_i + \sigma_i B$.*

Note that this assumption implies Assumption 3.1 (by Theorem 3.A.18 in [33]). In Section 6 it is shown that, without any assumption on the service time distributions, no bound on the approximation ratio can be found.

To obtain an upper bound on the cost function under the SVF sequence, we also need to specify the schedule we are using. For the upper bound under Assumption 4.1, we use a schedule of the form $\boldsymbol{x} = \boldsymbol{\mu} + \alpha\boldsymbol{\sigma}$ for some $\alpha > 0$. This means that we plan an amount of time for each appointment equal to the expected time the appointment will take, plus an extra amount of time proportional to its standard deviation, so as to be able to absorb delays. The α will be set to

$$\alpha = \sqrt{\frac{1-\omega}{2\omega} + \frac{\sigma_{n-1}}{2\sum_{i=1}^{n-1}\sigma_i}}, \tag{10}$$

in order to minimize the upper bound. Let Q_B denote the quantile function of B, i.e. $Q_B(y) = \inf\{x : y \leqslant \mathbb{P}(B \leqslant x)\}$. Define $B(\omega) := B - Q_B(1-\omega)$, and

$$K(B,\omega) := \sqrt{2\omega}/\left[\omega\mathbb{E}B(\omega)^- + (1-\omega)\mathbb{E}B(\omega)^+\right].$$

The main result of this section is the following.

Theorem 4.2. *Suppose that, for the SVF sequence, we use the schedule $\boldsymbol{x} = \boldsymbol{\mu} + \alpha\boldsymbol{\sigma}$, with α given by* (10)*. Under Assumption* 4.1*, we have $r_\omega \leqslant K(B,\omega)$.*

This result follows immediately from the bounds on the cost function $C_\omega(\tau, \boldsymbol{x})$ given in the following two propositions, that are proved in Appendix B.

Proposition 4.3. *Suppose* α *is given by* (10). *Under Assumption* 4.1*,*

$$C_\omega(\mathrm{id}, \boldsymbol{\mu} + \alpha\boldsymbol{\sigma}) \leqslant \sqrt{2\omega}\sum_{i=1}^{n-1}\sigma_i.$$

Proposition 4.4. *Under Assumption* 4.1*, for any sequence and schedule,*

$$C_\omega(\tau, \boldsymbol{x}) \geqslant \left[\omega\mathbb{E}B(\omega)^- + (1-\omega)\mathbb{E}B(\omega)^+\right]\sum_{i=1}^{n-1}\sigma_i.$$

The idea behind proving Proposition 4.3 is as follows. We use that the waiting time can be expressed as the maximum of a random walk, as per equation (3). An upper bound for this maximum can now be found by comparing to another random walk that has i.i.d. steps, each distributed as the step of the original random walk with the largest variance. This upper bound is found by noting that if (i) one splits the steps in two parts, (ii) multiplies the last part by some constant larger than one (leaving the first part unchanged), then the maximum increases. For the maximum of the new i.i.d. random walk, the classical *Kingman's bound* can be applied. After thus finding an upper bound on the expected waiting time, the expected idle time can then also be bounded using (5).

The idea behind the proof of Proposition 4.4 is to write

$$\omega\mathbb{E}I_{k+1} + (1-\omega)\mathbb{E}W_{k+1} = \omega\mathbb{E}(W_k + B_{\tau(k)} - x_{\tau(k)})^- + (1-\omega)\mathbb{E}(W_k + B_{\tau(k)} - x_{\tau(k)})^+,$$

and minimize this over $W_k - x_{\tau(k)}$. This minimization problem is the classical *newsvendor problem*, which has a known solution. This results in a lower bound on the cost function that is independent of schedule. This lower bound can also be easily minimized over the sequences, resulting in Proposition 4.4.

When $\omega = \frac{1}{2}$, i.e. when waiting time and idle time are equally important, we know that $Q_B(\frac{1}{2})$ is equal to the median of B. In this case, we have $K\left(B, \frac{1}{2}\right) = 2/\mathbb{E}|B-m|$, where m is the median of B.

4.2. **Examples.** Now we present examples of location-scale families for which we can compute $K(B,\omega)$ or $K\left(B, \frac{1}{2}\right)$ from Theorem 4.2. This way we can obtain some insight into the magnitude of the constant $K(B,\omega)$. For the location-scale families of normal, uniform, shifted exponential and Laplace distributions, the results are shown in Table 2. For normal distributions $K(B,\omega)$ is not shown, as the expression does not simplify (with respect to the one presented in Theorem 4.2).

Location-scale family	$K(B,\omega)$	$K\left(B,\frac{1}{2}\right)$
Normal	-	$\sqrt{2\pi} \approx 2.51$
Uniform	$\frac{1}{1-\omega}\sqrt{\frac{2}{3\omega}}$	$\frac{4}{3}\sqrt{3} \approx 2.31$
Shifted exponential	$-\frac{\sqrt{2}}{\sqrt{\omega}\ln(\omega)}$	$2/\ln(2) \approx 2.89$
Laplace	$\frac{2\sqrt{\omega}}{\min\{\omega,1-\omega\}(1-\ln(2\min\{\omega,1-\omega\}))}$	$2\sqrt{2} \approx 2.83$

Table 2. The values of $K(B,\omega)$ and $K\left(B,\frac{1}{2}\right)$ for some location-scale families.

Now consider the case of Pareto (of type II, that is) distributions. A random variable X has such a distribution if

$$\mathbb{P}(X > x) = \left(1 + \frac{x-\mu}{\sigma}\right)^{-\beta} \quad \text{for } x \geqslant \mu,$$

for certain parameters $\mu, \sigma > 0, \beta$. The Pareto distributions with fixed parameter β form a location-scale family. Suppose that the B_i have Pareto distributions with fixed parameter $\beta > 2$. Then

$$K(B,\omega) = \sqrt{\frac{2\omega\beta}{\beta-2}} \Big/ \left[-2\omega^{-\beta(\beta-1)+1} + \omega^{-\beta(\beta-1)} - (\beta-1)\omega^{\beta+1} + \beta\omega\right].$$

In addition,

$$K\left(B,\tfrac{1}{2}\right) = 2\sqrt{\frac{\beta}{\beta-2}} \Big/ \left[\beta - \left(\tfrac{1}{2}\right)^{\beta}(\beta-1)\right].$$

For most typical location-scale families, the value of $K\left(B,\frac{1}{2}\right)$ is between 2 and 3. However, in the Pareto case the value becomes much larger when β approaches two. Also, for ω close to either one of the extremes 0 or 1, the constant $K(B,\omega)$ blows up.

4.3. **Lognormally distributed service times.** In this subsection we use the notation $m_i := \mathbb{E}[\ln B_i]$ and $s_i^2 := \mathrm{Var}(\ln B_i)$. We have the following result.

Theorem 4.5. *Suppose the B_i are lognormally distributed with $m_1 \leqslant \ldots \leqslant m_n$ and $s_1^2 \leqslant \ldots \leqslant s_n^2$. When we use the schedule $\boldsymbol{x} = (1+\alpha)\boldsymbol{\mu}$ for the* SVF *sequence, with*

$$\alpha = \frac{1}{\sqrt{2\omega}}\sqrt{(\exp(s_{n-1}^2) - 1)},$$

then

$$r_\omega \leqslant 2\omega\alpha \cdot \left[(1-\omega)\mathbb{P}(Z \geqslant Q_Z(1-\omega) - s_1) - \omega\,\mathbb{P}(Z \leqslant Q_Z(1-\omega) - s_1)\right]^{-1}.$$

The proof of this theorem can be found in Appendix B. The ideas behind the proof are similar to those of Theorem 4.2. The main difference is in the upper bound, where it needs to be proved that the i.i.d. random walk used for comparison indeed has a bigger expected maximum. For lognormal distributions, we use a convex ordering among the stepsize distributions to prove this, noting that the maximum of a random walk is a convex function in the stepsizes.

As an example, we apply Theorem 4.5 to the data found in Çayırlı et al. [6]. Recall that the patients were divided into "new" and "return" patients, with service times fitted by lognormal distributions with parameters given in Table 1. It can be checked that any problem instance containing a mix of these patient groups satisfies the assumptions of Theorem 4.5. For any such problem instance, the largest possible s_{n-1} corresponds to a new patient, and the smallest possible s_1 corresponds to a return patient. When setting $\omega = \frac{1}{2}$, calculating the upper bound in Theorem 4.5, we find for the doctor studied in Çayırlı et al. that $r_{\frac{1}{2}} \leqslant 2.90$.

5. Asymptotic optimality of SVF

In this section we assess the performance of the SVF sequence as the number of patients grows large. Throughout this section we assume that the schedule is mean-based: the time planned for each appointment is equal to the corresponding mean service time. The goal in this section is to prove that the SVF sequence is asymptotically optimal as the number of patients tends to infinity, under certain conditions.

5.1. Main result.

We consider the setting in which we are given, for each value of n, a vector $\boldsymbol{B}_n = (B_{n,1}, B_{n,2}, \dots, B_{n,n})$ of service-time distributions. For $i \leqslant n$, let $\mu_{n,i}$ and $\sigma^2_{n,i}$ denote the mean and variance of $B_{n,i}$, and let $X_{n,i} := B_{n,i} - \mu_{n,i}$ for all $i \leqslant n$. Similarly, $S_{n,i}$, $W_{n,i}$, and $I_{n,i}$ are all with respect to the service-time distributions $\boldsymbol{B}_n$.

We require the following assumption, similar to the Lyapunov condition of the Lyapunov-version of the central limit theorem (CLT). The difference between our assumption and the conventional Lyapunov condition is the supremum over all $n \geqslant k$ and all sequences τ.

Assumption 5.1. *We assume that there exists a $\delta > 0$ such that, as $k \to \infty$,*

$$q_k := \sup_{n \geqslant k, \tau \in \mathsf{S}_n} \frac{1}{\sqrt{\sum_{i=1}^k \sigma^2_{n,\tau(i)}}^{\,2+\delta}} \sum_{i=1}^k \mathbb{E}|X_{n,\tau(i)}|^{2+\delta} \to 0.$$

In Section 6 it is shown that it is necessary to make this assumption. The main result of this section is the following.

Theorem 5.2. *Under Assumption 5.1, $\varrho_\omega(\boldsymbol{B}_n) \to 1$ as $n \to \infty$.*

To prove Theorem 5.2, we derive an upper and a lower bound on the expected waiting time, which we then combine. We again view the waiting time as the maximum of the random walk $S_{n,i}$. For both bounds we use the reflected process $\widehat{S}_{n,i}$ in level a, that is also defined in (6), to obtain bounds on the distribution of the waiting time. For the upper bound we can ignore the difference between $S_{n,i}$ and $\widehat{S}_{n,i}$ right after crossing level a for the first time, resulting in the process $\tilde{S}_{n,i}$. The processes $S_{n,i}$, $\widehat{S}_{n,i}$ and $\tilde{S}_{n,i}$ are illustrated in Figure 1. For the lower bound we truncate all steps at some value c. The difference between $\widehat{S}_{n,i}$ and $\tilde{S}_{n,i}$ is then bounded by $2c$. We then choose c small enough that this difference becomes negligible in the limit, but also big enough that the difference between the original random walk and the random walk with truncated steps becomes negligible as well. Using a Berry-Esseen bound for martingales established in [17], we can then estimate the distributions of $S_{n,k}$ and $\tilde{S}_{n,k}$, and thus the distribution of $W_{n,k+1}$, for large k.

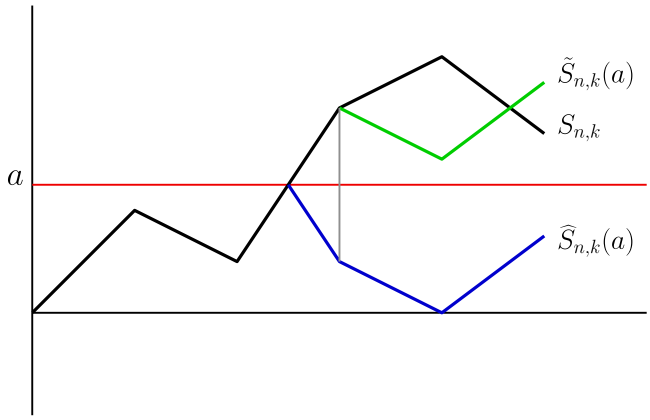

Figure 1. The process $S_{n,j}$ in black, $\widehat{S}_{n,j}(a)$ in blue and $\tilde{S}_{n,j}(a)$ in green. The red line indicates level a.

Define $\Sigma^2_{n,k} := \sum_{i=1}^k \sigma^2_{n,\tau(i)}$. To prove Theorem 5.2, we need the following two propositions, that are proved in Section 5.2 and Section 5.3 respectively. These propositions are then combined in Proposition 5.5, after which we can establish Theorem 5.2.

Proposition 5.3. *For any k and $n \geqslant k$ we have*

$$\frac{\mathbb{E}W_{n,k+1}}{\Sigma_{n,k}} \leqslant \mathbb{E}|Z| + 2(C_\delta + 1)\sqrt{q_k}^{1/(3+\delta)},$$

where Z is a standard normal random variable, and C_δ is a constant that only depends on δ.

Proposition 5.4. *Under Assumption 5.1, for each $\varepsilon > 0$ there exists a K depending on ε only, such that for all $k \geqslant K$ and all $n \geqslant k$,*

$$\frac{\mathbb{E}W_{n,k+1}}{\Sigma_{n,k}} \geqslant (1-\varepsilon)\mathbb{E}|Z|.$$

Proposition 5.5. *Under Assumption* 5.1*, for any* $\varepsilon > 0$ *there exists a* K *depending on* ε *only, such that for all* $k \geqslant K$ *and for all* $n \geqslant k-1$*,*

$$\mathbb{E}W^{\text{SVF}}_{n,k} \leqslant (1+\varepsilon)\mathbb{E}W^{\text{OPT}}_{n,k}.$$

Proof. By Proposition 5.4, for any $\varepsilon > 0$, we can choose k sufficiently large such that

$$\mathbb{E}W_{n,k+1} \geqslant (1-\varepsilon)\sqrt{\sum_{i=1}^{k}\sigma^2_{n,\tau(i)}} \cdot \mathbb{E}|Z| \geqslant (1-\varepsilon)\sqrt{\sum_{i=1}^{k}\sigma^2_{n,i}} \cdot \mathbb{E}|Z|$$

for any sequence τ, in particular for the optimal sequence. Here we also used that $\sigma^2_{n,1}, \ldots, \sigma^2_{n,k}$ are the k smallest variances. For the SVF sequence, Proposition 5.3 gives us that for sufficiently large k we have

$$\mathbb{E}W^{\text{SVF}}_{n,k+1} \leqslant (1+\varepsilon)\sqrt{\sum_{i=1}^{k}\sigma^2_{n,i}} \cdot \mathbb{E}|Z|.$$

Combining these two bounds completes the proof. □

Proof of Theorem 5.2. By the Lindley recursion we have

$$\mathbb{E}W_{n,k+1} = \mathbb{E}(W_{n,k} + X_{n,\tau(k)})^+ \geqslant \mathbb{E}(W_{n,k} + X_{n,\tau(k)}) = \mathbb{E}W_{n,k},$$

so, for any n and τ, $\mathbb{E}W_{n,k}$ is increasing in k. With K as in Proposition 5.5, we consequently have

$$\sum_{k=1}^{n}\mathbb{E}W^{\text{SVF}}_{n,k} \leqslant K\mathbb{E}W^{\text{SVF}}_{n,K} + \sum_{k=K+1}^{n}\mathbb{E}W^{\text{SVF}}_{n,k} \leqslant (1+\varepsilon)\left(K\mathbb{E}W^{\text{OPT}}_{n,K} + \sum_{k=K+1}^{n}\mathbb{E}W^{\text{OPT}}_{n,k}\right).$$

Again using that $\mathbb{E}W_{n,k}$ is increasing in k, we also have

$$K\,\mathbb{E}W^{\text{OPT}}_{n,K} \Big/ \sum_{k=K+1}^{n}\mathbb{E}W^{\text{OPT}}_{n,k} \leqslant \frac{K}{n-K},$$

which is smaller than ε for sufficiently large n. Then we have

$$\sum_{k=1}^{n}\mathbb{E}W^{\text{SVF}}_{n,k} \leqslant (1+\varepsilon)\left(K\mathbb{E}W^{\text{OPT}}_{n,K} + \sum_{k=K+1}^{n}\mathbb{E}W^{\text{OPT}}_{n,k}\right) \leqslant (1+\varepsilon)^2\sum_{k=K+1}^{n}\mathbb{E}W^{\text{OPT}}_{n,k} \leqslant (1+\varepsilon)^2\sum_{k=1}^{n}\mathbb{E}W^{\text{OPT}}_{n,k}.$$

Now note that taking expectations in (5), and using that $x_{n,i} = \mu_{n,i}$, leads to

$$\sum_{i=1}^{n}\mathbb{E}I_{n,i} = \mathbb{E}W_{n,n},$$

so the total expected idle time in the cost function can be replaced by $\mathbb{E}W_{n,n}$. Since also $\mathbb{E}W^{\text{SVF}}_{n,n} \leqslant (1+\varepsilon)W^{\text{OPT}}_{n,n}$ for $n \geqslant K$,

$$\omega \mathbb{E}W^{\text{SVF}}_{n,n} + (1-\omega)\sum_{k=1}^{n} \mathbb{E}W^{\text{SVF}}_{n,k} \leqslant (1+\varepsilon)^2 \left(\omega \mathbb{E}W^{\text{OPT}}_{n,n} + (1-\omega)\sum_{k=1}^{n} \mathbb{E}W^{\text{OPT}}_{n,k}\right)$$

for n sufficiently large (independent of ω). As this holds for any $\varepsilon > 0$, we find $\varrho_\omega(\boldsymbol{B_n}) \to 1$ as $n \to \infty$, and Theorem 5.2 is proved. □

Remark 5.6. As in Remark 3.7, when the scheduled session end time is equal to the expected total service time, the expected overtime $\mathbb{E}W_{n,n+1}$ can be handled similarly to waiting time, and the result of Theorem 5.2 is also valid when some extra term $c\,\mathbb{E}W_{n,n+1}$ (with $c > 0$) is added to the cost function.

Remark 5.7. Note that the mean-based schedule we consider here leads to an unstable queue, implying that waiting times explode when n becomes large. Therefore, one might wonder what happens in the limit when we use a schedule with larger interarrival times (implying that the queue is stable). Insight into such a system can be gained by considering the setting with m groups consisting of patients with i.i.d. service times. Suppose that each group has to be served consecutively, and within each group the interarrival times are constant and larger than the mean service time. Letting the number of patients in each group (say $n_1, \dots, n_m$) grow large, then within each group the corresponding queue effectively behaves as in stationarity. As the stationary behavior does not depend on the groups that came before, the effect of sequencing the groups will vanish in the limit as $n_1, \dots, n_m \to \infty$.

Remark 5.8. Suppose that $\inf_{n,i}\{\sigma^2_{n,i}\} > 0$ and $\sup_{n,i}\{\mathbb{E}|X_{n,i}|^{2+\delta}\} < \infty$. Then $q_k = O(k^{-\delta/2})$ indeed converges to zero. Following the steps of the proof of Theorem 5.2 we then find

$$\varrho_\omega(\boldsymbol{B}_n) = 1 + O\left(\frac{K}{n}\right) + O\left(\sqrt{q_K}^{1/(3+\delta)}\right) = 1 + O\left(\frac{K}{n}\right) + O\left(K^{-\delta/(12+4\delta)}\right).$$

To obtain some insight into the convergence rate, observe that by choosing K in a way that these terms are balanced, it follows that

$$\varrho_\omega(\boldsymbol{B}_n) = 1 + O(n^{-\delta/(12+5\delta)}).$$

5.2. **Proof of Proposition 5.3.** For ease of notation, we assume that τ is the permutation $\tau(i) = i$. Recall that $W_{n,k+1} = \max\{0, S_{n,1}, \dots, S_{n,k}\}$. Let $T(a) = \inf\{j : S_{n,j} \geqslant a\}$, and define

$$\widehat{S}_{n,j}(a) = \begin{cases} S_{n,j} & \text{if } k < T(a) \\ 2a - S_{n,j} & \text{if } k \geqslant T(a); \end{cases}$$

cf. (6). Now in order for $W_{n,k+1}$ to be above a, $T(a)$ must be at most k, and then either $S_{n,k}$ is at least a or $2a - S_{n,k} = \widehat{S}_{n,k}(a)$ is above a. As these are disjoint events, we have

$$\mathbb{P}(W_{n,k+1} \geqslant a) = \mathbb{P}(S_{n,k} \geqslant a) + \mathbb{P}(\widehat{S}_{n,k}(a) > a),$$

which was also found in the proof of Lemma 3.6. Now define $\tilde{S}_{n,j}(a)$ recursively as follows. Let $\tilde{S}_{n,0}(a) = 0$, and put

$$\tilde{S}_{n,j+1}(a) = \begin{cases} \tilde{S}_{n,j}(a) + X_{n,k-j} & \text{if } j < T(a) \\ \tilde{S}_{n,j}(a) - X_{n,k-j} & \text{if } j \geqslant T(a). \end{cases}$$

The three processes $S_{n,j}$, $\widehat{S}_{n,j}(a)$ and $\tilde{S}_{n,j}(a)$ are illustrated in Figure 1.

Note that $\tilde{S}_{n,j}(a)$ is a martingale, with $\mathbb{E}\tilde{S}_{n,k}(a) = 0$ and $\operatorname{Var}\tilde{S}_{n,k}(a) = \sum_{i=1}^{k} \sigma_{n,i}^2$. Also, note that the processes $\widehat{S}_{n,j}(a)$ and $\tilde{S}_{n,j}(a)$ have the same increments, except when the process $S_{n,j}$ crosses level a for the first time. In this step the increments differ by $2(S_{n,T(a)} - a)$, twice the amount by which the random walks "overshoots" level a. As this overshoot is nonnegative and bounded by $\max_{1\leqslant i\leqslant k}\{X_{n,i}\}$, we find that

$$\tilde{S}_{n,k}(a) \geqslant \widehat{S}_{n,k}(a) \geqslant \tilde{S}_{n,k}(a) - 2\max_{1\leqslant i\leqslant k}\{X_{n,i}\}.$$

This leads to the estimates

$$\begin{aligned} \mathbb{P}(S_{n,k} \geqslant a) + \mathbb{P}(\tilde{S}_{n,k}(a) > a) &\geqslant \mathbb{P}(W_{n,k+1} \geqslant a) \\ &\geqslant \mathbb{P}(S_{n,k} > a) + \mathbb{P}\left(\tilde{S}_{n,k}(a) > a + 2\max_{1\leqslant i\leqslant k}\{X_{n,i}\}\right). \end{aligned} \tag{11}$$

As a consequence, we also have

$$\begin{aligned} \mathbb{E}W_{n,k+1} &\leqslant \int_0^\infty \Big(\mathbb{P}(S_{n,k} \geqslant a) + \mathbb{P}(\tilde{S}_{n,k}(a) > a)\Big)\,\mathrm{d}a \\ &= \int_0^\infty \Big(\mathbb{P}(S_{n,k} > a) + \mathbb{P}(\tilde{S}_{n,k}(a) > a)\Big)\,\mathrm{d}a. \end{aligned} \tag{12}$$

The idea is now to estimate the probabilities $\mathbb{P}(S_{n,k} > a)$ and $\mathbb{P}(\tilde{S}_{n,k}(a) > a)$ using a CLT-type result for martingales. Our approach relies on the following result, established in [17].

Theorem 5.9 (Heyde and Brown)**.** *Let $(\xi_i, \mathcal{F}_i)$ be a sequence of martingale differences, and let $Y_j = \xi_1 + \cdots + \xi_j$ be the corresponding martingale. Suppose that the conditional variance,* *given by*

$$\sum_{i=1}^{k} \mathbb{E}[\xi_i^2 | \mathcal{F}_{i-1}],$$

is equal to one for some k. Then for any $\delta > 0$ there exists a constant C_δ that depends on δ only, such that, with Z denoting a standard normal random variable,

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(Y_k > x) - \mathbb{P}(Z > x)| \leqslant C_\delta \left(\sum_{i=1}^{k} \mathbb{E}|\xi_i|^{2+\delta} \right)^{1/(3+\delta)}.$$

Proof of Proposition 5.3. Note that $\Sigma_{n,k}^2 := \sum_{i=1}^{k} \sigma_{n,i}^2$ is the variance of both $S_{n,k}$ and $\tilde{S}_{n,k}(a)$. In order to apply Theorem 5.9 we scale all steps, and hence $S_{n,k}$ and $\tilde{S}_{n,k}(a)$, by a factor $1/\Sigma_{n,k}$. For both martingales the squared increments $(X_{n,k-i+1})^2$ are independent of the previous increments, hence after rescaling the conditional variance after k steps equals one for both martingales. Note that we can recognize the q_k from Assumption 5.1 in the upper bound, so we find for any x that

$$\left| \mathbb{P}\left(\frac{S_{n,k}}{\Sigma_{n,k}} > x \right) - \mathbb{P}(Z > x) \right| \leqslant C_\delta q_k^{1/(3+\delta)}, \tag{13}$$

$$\left| \mathbb{P}\left(\frac{\tilde{S}_{n,k}(a)}{\Sigma_{n,k}} > x \right) - \mathbb{P}(Z > x) \right| \leqslant C_\delta q_k^{1/(3+\delta)}. \tag{14}$$

Using inequality (13) and Chebyshev's inequality, we find

$$\begin{aligned} \frac{1}{\Sigma_{n,k}} \int_0^\infty \mathbb{P}(S_{n,k} > a)\mathrm{d}a &= \int_0^\infty \mathbb{P}\left(\frac{S_{n,k}}{\Sigma_{n,k}} > x \right) \mathrm{d}x \\ &\leqslant \int_0^{1/\sqrt{q_k}^{1/(3+\delta)}} \left(\mathbb{P}(Z > x) + C_\delta q_k^{1/(3+\delta)} \right) \mathrm{d}x + \int_{1/\sqrt{q_k}^{1/(3+\delta)}}^\infty \frac{1}{x^2} \mathrm{d}x \\ &\leqslant \mathbb{E}Z^+ + (C_\delta + 1)\sqrt{q_k}^{1/(3+\delta)}. \end{aligned}$$

A similar reasoning using (14) finds the same upper bound for

$$\frac{1}{\Sigma_{n,k}} \int_0^\infty \mathbb{P}(\tilde{S}_{n,k}(a) > a)\mathrm{d}a.$$

Now adding these bounds together and using the bound in (12) proves Proposition 5.3. □

5.3. **Proof of Proposition 5.4.** Again we assume $\tau(i) = i$ for ease of notation, which, in this section, does not necessarily correspond to the SVF sequence.

Recall that

$$\widehat{S}_{n,k}(a) \geqslant \tilde{S}_{n,k}(a) - 2 \max_{1\leqslant i\leqslant k} \{X_{n,i}\}.$$

An upper bound on $\max_{1\leqslant i\leqslant k}\{X_{n,i}\}$ could be used to find a lower bound on $\widehat{S}_{n,k}(a)$. If we would change the steps in such a way that all steps are at most $c_{n,k}$, for some $c_{n,k}$ depending on n and k but not on i, then this $c_{n,k}$ would give an upper bound on $\max_{1\leqslant i\leqslant k}\{X_{n,i}\}$, and we would have

$$\widehat{S}_{n,k}(a) \geqslant \tilde{S}_{n,k}(a) - 2c_{n,k}.$$

We will now consider the random walk with steps $X_{n,i}\mathbb{1}_{X_{n,i}\leqslant c_{n,k}}$ instead of $X_{n,i}$ (where $\mathbb{1}_E$ denotes the indicator of an event E). Let $W'_{n,k+1}$ be the maximum of the new random walk. When $c_{n,k} > 0$, we now have

$$\mathbb{E}W_{n,k+1} \geqslant \mathbb{E}W'_{n,k+1}$$

as we are lowering all the steps. The $c_{n,k}$ are picked later in a way that balances the need to sufficiently bound the overshoot and the need to not affect the steps too much.

Now let $S'_{n,k}$, $\widehat{S}'_{n,k}(a)$, $\tilde{S}'_{n,k}(a)$ be the random variables defined for this new random walk in the same way as $S_{n,k}$, $\widehat{S}_{n,k}(a)$ and $\tilde{S}_{n,k}(a)$ were defined for the old random walk. We thus have

$$\mathbb{P}(W'_{n,k+1} \geqslant a) \geqslant \mathbb{P}(S'_{n,k} > a) + \mathbb{P}(\widehat{S}'_{n,k}(a) > a) \geqslant \mathbb{P}(S'_{n,k} > a) + \mathbb{P}(\tilde{S}'_{n,k}(a) > a + 2c_{n,k}).$$

Note that the steps no longer have mean zero, and so $\tilde{S}'_{n,j}(a)$ is no longer a martingale. To repair this issue, we must know how much the change in steps due to the indicator affects the mean and variance of all the steps. For this we have the following lemma, that is proved in Appendix C. Define $(\Sigma'_{n,k})^2 := \operatorname{Var} S'_{n,k}$ and

$$\gamma_{n,k} := \left(\frac{\Sigma_{n,k}}{c_{n,k}}\right)^{\delta}.$$

Lemma 5.10. *For any k and $n \geqslant k$,*

(1) $$\frac{\sum_{i=1}^{k} \mathbb{E}\left[(X_{n,i})\mathbb{1}_{X_{n,i}>c_{n,k}}\right]}{\Sigma_{n,k}} \leqslant q_k \gamma_{n,k}^{(1+\delta)/\delta}\ ; \quad \textit{and}$$

(2) $$\frac{(\Sigma'_{n,k})^2}{\Sigma^2_{n,k}} \geqslant 1 - 2q_k\gamma_{n,k}.$$

Now we have the following result. Define, with Z is a standard normal random variable, and C_δ the constant featuring in Theorem 5.9,

$$D_k := \int_0^{1/\sqrt{q_k}^{1/(3+\delta)}} \mathbb{P}(Z > x)\mathrm{d}x - C_\delta \sqrt{q_k}^{1/(3+\delta)}.$$

Lemma 5.11. *When $c_{n,k} > 0$,*

$$\frac{\mathbb{E}W_{n,k+1}}{\Sigma_{n,k}} \geqslant 2D_k\sqrt{1-2q_k\gamma_{n,k}} - \frac{2c_{n,k}}{\Sigma_{n,k}} - 2q_k\gamma_{n,k}^{(1+\delta)/\delta}.$$

Proof. Again we want to apply Theorem 5.9. This time we not only need to divide by the standard deviation $\Sigma'_{n,k}$ of $S'_{n,k}$, but also subtract the (negative) mean. We then find

$$\begin{aligned}\frac{1}{\Sigma'_{n,k}}\int_0^\infty \mathbb{P}(S'_{n,k} > a)\mathrm{d}a &= \int_{-\mathbb{E}S'_{n,k}/\Sigma'_{n,k}}^\infty \mathbb{P}\left(\frac{S'_{n,k}-\mathbb{E}S'_{n,k}}{\Sigma'_{n,k}} > x\right)\mathrm{d}x \\ &\geqslant \int_0^{1/\sqrt{q_k}^{1/(3+\delta)}} \left(\mathbb{P}(Z > x) - C_\delta q_k^{1/(3+\delta)}\right)\mathrm{d}x \\ &\quad - \int_0^{-\mathbb{E}S'_{n,k}/\Sigma'_{n,k}} \mathbb{P}\left(\frac{S'_{n,k}-\mathbb{E}S'_{n,k}}{\Sigma'_{n,k}} > x\right)\mathrm{d}x \\ &\geqslant D_k + \frac{\mathbb{E}S'_{n,k}}{\Sigma'_{n,k}},\end{aligned}$$

where we used in the last step that probabilities are bounded by one.

Next, we need a bound for

$$\frac{1}{\Sigma'_{n,k}}\int_0^\infty \mathbb{P}(\tilde{S}'_{n,k}(a) > a + 2c_{n,k})\mathrm{d}a. \tag{15}$$

Note that $\tilde{S}'_{n,j}(a)$ is no longer a martingale, as the mean step size is no longer zero. However, we can remedy this by noting that $\tilde{S}'_{n,k}(a)$ is bounded below by

$$\tilde{S}_{n,k}(a) - \sum_{i=1}^k (X_{n,i})\mathbb{1}_{X_{n,i}>c_{n,k}} = \tilde{S}_{n,k}(a) - S_{n,k} + S'_{n,k},$$

and $\tilde{S}_{n,j}(a) - S_{n,j} + S'_{n,j} - \mathbb{E}S'_{n,j}$ is again a martingale with mean zero. Applying Theorem 5.9 to this martingale and taking into account the shift $\mathbb{E}S'_{n,k}$ and overshoot $2c_{n,k}$, we find that (15) is bounded below by

$$D_k + \frac{\mathbb{E}S'_{n,k}}{\Sigma'_{n,k}} - \frac{2c_{n,k}}{\Sigma'_{n,k}}.$$

Now adding up these two lower bounds, we find

$$\frac{\mathbb{E}W'_{n,k+1}}{\Sigma'_{n,k}} \geqslant 2D_k + \frac{2\,\mathbb{E}S'_{n,k}}{\Sigma'_{n,k}} - \frac{2c_{n,k}}{\Sigma'_{n,k}}.$$

Now note that, due to Lemma 5.10.(2), $\Sigma'_{n,k}/\Sigma_{n,k} \geqslant \sqrt{1-2q_k\gamma_{n,k}}$. In addition, $\mathbb{E}W_{n,k+1} \geqslant \mathbb{E}W'_{n,k+1}$. Consequently,

$$\frac{\mathbb{E}W_{n,k+1}}{\Sigma_{n,k}} \geqslant 2D_k\sqrt{1-2q_k\gamma_{n,k}} + \frac{2\,\mathbb{E}S'_{n,k}}{\Sigma_{n,k}} - \frac{2c_{n,k}}{\Sigma_{n,k}}.$$

Applying Lemma 5.10.(1),

$$\frac{\mathbb{E}W_{n,k+1}}{\Sigma_{n,k}} \geqslant 2D_k\sqrt{1-2q_k\gamma_{n,k}} - 2q_k\gamma_{n,k}^{(1+\delta)/\delta} - \frac{2c_{n,k}}{\Sigma_{n,k}},$$

which is the bound we wanted to prove. □

Proof of Proposition 5.4. We can still choose $c_{n,k} > 0$, as each choice gives a lower bound on $\mathbb{E}W_{n,k+1}$. We would like to have $c_{n,k}/\Sigma_{n,k} \to 0$, $q_k\gamma_{n,k}^{(1+\delta)/\delta} \to 0$ and $q_k\gamma_{n,k} \to 0$ as $k \to \infty$. A choice that achieves this goal is

$$c_{n,k} := \sqrt{q_k}^{1/(\delta+1)}\Sigma_{n,k},$$

so that $\gamma_{n,k} = q_k^{-\delta/(2\delta+2)}$. We thus obtain, with $\bar{q}_k := \sqrt{q_k}^{1/(\delta+1)} + \sqrt{q_k}$,

$$\frac{\mathbb{E}W_{n,k+1}}{\Sigma_{n,k}} \geqslant 2\left(\int_0^{1/\sqrt{q_k}^{1/(3+\delta)}} \mathbb{P}(Z > x)\mathrm{d}x - C_\delta\sqrt{q_k}^{1/(3+\delta)}\right)\sqrt{1-2\sqrt{q_k}^{2-\delta/(\delta+1)}} - 2\bar{q}_k.$$

Note that this converges to $\mathbb{E}|Z|$ as $q_k \to 0$, so this completes the proof of Proposition 5.4. □

6. Two examples with lower bounds on performance

In this section we present two insightful examples that give lower bounds on ϱ_ω and r_ω for particular problem instances. The first example shows the necessity of certain assumptions in being able to give any upper bound on the approximation ratio. The second example gives a lower bound on ϱ_ω for problem instances that satisfy Assumption 3.1, thus complementing Theorem 3.4.

Example 6.1. Suppose we have two patients, and the service time of patient i is given by

$$B_i = \begin{cases} \mu_i + \frac{1}{a_i} & \text{with probability } a_i^2 \\ \mu_i - \frac{1}{a_i} & \text{with probability } a_i^2 \\ \mu_i & \text{with probability } 1-2a_i^2, \end{cases} \tag{16}$$

for some values $\mu_i > 0$ and $a_i \leqslant 1/\sqrt{2}$. Then $\mathbb{E}B_i = \mu_i$, and $\mathrm{Var}B_i = 1$, and so either of the two possible sequences could be considered the SVF sequence. We take the SVF sequence to be given by $\tau(i) = i$ (one could of course perturb the distributions so that this is the unique SVF ordering).

Now assume that we use the mean-based schedule given by $\boldsymbol{x} = \boldsymbol{\mu}$. The cost function for the SVF sequence $\tau(i) = i$ is then

$$\omega\mathbb{E}I_2 + (1-\omega)\mathbb{E}W_2 = \omega\mathbb{E}(B_1-\mu_1)^- + (1-\omega)\mathbb{E}(B_1-\mu_1)^+ = \omega a_1 + (1-\omega)a_1 = a_1.$$

In the same way, the cost function for the other sequence is equal to a_2. If we take a_1 to be bigger than a_2, we conclude $\varrho_\omega = a_1/a_2$, which can be arbitrarily large. This shows that it is necessary in Section 3 to impose Assumption 3.1, even under Assumption 3.2.

The construction can easily be extended to one with any larger number of patients, by introducing additional patients with deterministic service times. Therefore, this example also shows that Assumption 5.1 is necessary in Section 5.

The example applies also when an optimal, rather than mean-based, scheduling rule is used. Fixing $\omega = \frac{1}{2}$, for two patients with service times as in (16), the cost function is

$$\tfrac{1}{2}\mathbb{E}\left(B_{\tau(1)} - x_{\tau(1)}\right)^{-} + \tfrac{1}{2}\mathbb{E}\left(B_{\tau(1)} - x_{\tau(1)}\right)^{+}.$$

By Lemma B.4, the $x_{\tau(1)}$ that minimizes this cost function is given by $x_{\tau(1)} = \mu_{\tau(1)}$, which is the mean of $B_{\tau(1)}$. So the situation is unchanged, and also for the optimal scheduling rule no bound on the approximation ratio can be found without imposing further assumptions. This justifies why we use Assumption 4.1 in Section 4.

Example 6.2. This example serves to give a lower bound on $\sup_{\omega,\boldsymbol{B}} \varrho_\omega(\boldsymbol{B})$, where the supremum is taken over all problem instances (including n and ω) that satisfy Assumption 3.1. We consider the situation where $\omega = 1$, that is, where the cost function is given by the total idle time. The conclusion which can be drawn from the example is that it is not possible to prove a better constant bound on ϱ_ω than 1.28, if we only make Assumption 3.1, thus complementing Theorem 3.4.

To prove this bound of 1.28, we again consider a sequence of problem instances $\boldsymbol{B}_n$ and use the same notation as has been used in Section 5. Suppose we have $n := n' + 1$ patients, and the service time of patient i for $i \leqslant n'$ is given by

$$B_{n,i} = \begin{cases} \mu_{n,i} + b_{n,i} & \text{with probability } 1 - \frac{C}{n'} \\ \mu_{n,i} - b_{n,i}\left(\frac{n'}{C} - 1\right) & \text{with probability } \frac{C}{n'}, \end{cases}$$

for some values $\mu_{n,i} > 0$, $C \in (0,1)$. Let $b_{n,i} := C/(n' - C)$ for $i = 1, \dots, n' - 1$ and $b_{n,n'} := 1 - C(n'-1)/(n'-C)$. We assume that n' is large enough to have $b_{n,n'} \geqslant b_{n,1}$. Then these service times satisfy Assumption 3.1, as each $B_{n,n'} - \mu_{n,n'}$ is in distribution equal to $c(B_{n,1} - \mu_{n,1})$ for some $c \geqslant 1$, and the $B_{n,i} - \mu_{n,i}$ are identically distributed for $i \leqslant n' - 1$. The service time of patient n is assumed to have the largest variance among all patients, and be such that Assumption 3.1 is still satisfied. We assume that we use the mean-based schedule, so the interarrival times are given by the mean service times.

Now suppose that the sequence τ is such that $\tau(n) = n$. Recall that $X_{n,i} = B_{n,i} - \mu_{n,i}$. By equations (3) and (5) we then have that

$$\sum_{i=1}^{n} \mathbb{E} I_{n,i} = \mathbb{E} W_{n,n'+1}$$

is the expected maximum of the random walk with steps $X_{n,\tau(n')}, X_{n,\tau(n'-1)}, \ldots, X_{n,\tau(1)}$. For sufficiently large n, any downward jump in the random walk will be at least as large as all possible upward jumps together, as $\sum_{i=1}^{n'} b_{n,i} = 1$ and

$$b_{n,i} \left(\frac{n'}{C} - 1 \right) \geqslant \frac{C}{n' - C} \left(\frac{n'}{C} - 1 \right) = 1.$$

Therefore, we can compute the maximum of the random walk by conditioning on the first downward jump, so

$$\sum_{i=1}^{n} I_{n,i} = \left(1 - \frac{C}{n'}\right)^{n} + \sum_{k=1}^{n'} \left(1 - \frac{C}{n'}\right)^{k-1} \frac{C}{n'} \sum_{i=1}^{k-1} b_{n,\tau(n'+1-i)}.$$

We will use this to compare the SVF sequence, given by $\tau(i) = i$, with the reverse sequence given by $\tau(i) = n' + 1 - i$ for $i \leqslant n'$ (and $\tau(n) = n$), where we let n tend to infinity.

First consider the reverse sequence. Then the total expected idle time is equal to the expected maximum of the random walk with steps $X_{n,1}, X_{n,2}, \ldots, X_{n,n'}$. Letting $I_{n,i}^{\text{REV}}$ denote the idle times under the reverse sequence, we now have

$$\begin{aligned}
\lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E} I_{n,i}^{\text{REV}} &= \lim_{n'\to\infty} \left(1 - \frac{C}{n'}\right)^{n'} + \sum_{k=1}^{n'} \left(1 - \frac{C}{n'}\right)^{k-1} \frac{C}{n'} (k-1) \frac{C}{n' - C} \\
&= \mathrm{e}^{-C} + \lim_{n'\to\infty} \frac{C^2}{n'^2} \sum_{k=1}^{n'} (k-1) \left(1 - \frac{C}{n'}\right)^{k-1} \\
&= \mathrm{e}^{-C} - \lim_{n'\to\infty} n' \frac{C}{n'} \left(1 - \frac{C}{n'}\right)^{n'} - \frac{C}{n'} \left(1 - \frac{C}{n'}\right)^{n'} + \left(1 - \frac{C}{n'}\right)^{n'} + \frac{C}{n'} - 1 \\
&= 1 - C\mathrm{e}^{-C}.
\end{aligned}$$

Now consider the SVF sequence, for which the total idle time equals the maximum of the random walk with steps $X_{n,n'}, \ldots, X_{n,1}$. We then have

$$\begin{aligned}
\lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E} I_{n,i}^{\text{SVF}} &= \lim_{n'\to\infty} \left(1 - \frac{C}{n'}\right)^{n'} + \sum_{k=1}^{n'} \left(1 - \frac{C}{n'}\right)^{k-1} \frac{C}{n'} \left[(k-1) \frac{C}{n' - C} + b_{n,n'} - b_{n,1} \right] \\
&= 1 - C\mathrm{e}^{-C} + \lim_{n'\to\infty} \left(1 - \left(1 - \frac{C}{n'}\right)^{n'} \right) (b_{n,n'} - b_{n,1}) \\
&= 1 - C\mathrm{e}^{-C} + (1 - \mathrm{e}^{-C})(1 - C) = 2 - C - \mathrm{e}^{-C}.
\end{aligned}$$

For this sequence of problem instances we now have

$$\lim_{n\to\infty} \varrho_1(\boldsymbol{B}_n) \geqslant \frac{2 - C - \mathrm{e}^{-C}}{1 - C\mathrm{e}^{-C}}.$$

Minimizing over C then gives $\sup_{\omega,\boldsymbol{B}} \varrho_\omega(\boldsymbol{B}) \geqslant 1.28$, as desired.

7. Discussion and directions for further research

We have shown that under quite general conditions, the SVF sequence yields a constant-factor approximation. Furthermore, we have seen that additional information about the instance, such as knowing that the service-time distributions fall within a certain class, or that the number of patients is large, can lead to substantial improvements in our worst-case bounds.

For mean-based schedules, Theorem 3.4 and Example 6.2 show that the worst-case approximation ratio lies between 1.28 and 4. It would be interesting to reduce this gap; we suspect neither bound is tight. In particular, the upper bound on the cost of the SVF sequence appears to be a strong bound only in the regime of many patients, with service times of similar variances; the lower bound on the cost of arbitrary sequences, on the other hand, appear strong in situations where only a few service times with large variance have a significant impact on the cost function. This suggests that more refined arguments, possibly considering multiple regimes, could lead to an improved upper bound. Improving our bounds for special cases (such as normal and lognormal distributions), or considering other practically relevant service-time distributions, would also be of interest.

When optimizing over both the sequence and the schedule, we obtained bounds for location-scale families and for lognormally-distributed service times. These bounds are, however, not uniform: in the former case, the bounds depend on ω and the location-scale family, and in the latter case, on the parameters of the lognormal distributions. A constant-factor approximation that does not depend on these quantities, or that holds in greater generality (e.g., to all distributions satisfying the ordering assumption), remains an open question. The SVF sequencing rule remains a promising candidate, but a more sophisticated choice of scheduling rule will certainly be needed.

To the best of our knowledge, this is the first paper that assesses whether an easily computed sequence performs provably well, rather than trying to find the optimal sequence for a special (typically low-dimensional) instance, or comparing heuristics through simulation. Finding the optimal sequence is an important (but inherently difficult) problem, and we hope our approach triggers more research in this direction.

Appendix A. Proofs corresponding to Section 3

Here we prove Lemma 3.5, Lemma 3.9 and Lemma 3.10, that are used in Section 3.

Lemma 3.5. *Let $\mathbb{E}W_{k+1}^{\text{OPT}'}$ denote the expected waiting time of the patient in appointment slot $k+1$, under the sequence that minimizes this expected waiting time, subject to the constraint that $\tau(i) \leqslant k$ for $i = 1, \ldots, k$, i.e. the first k patients are assigned to the first k slots. Suppose $\mathbb{E}W_{k+1}^{\text{SVF}}/\mathbb{E}W_{k+1}^{\text{OPT}'} \leqslant \varrho'$ for all k. Then, under Assumption 3.1, $\varrho_\omega \leqslant \varrho'$.*

Proof. Taking expectations in (5) and using that $x_i = \mu_i$, we find that $\sum_{i=1}^n \mathbb{E}I_i = \mathbb{E}W_n$. Hence, the cost function (2) equals $\omega\mathbb{E}W_n + (1-\omega)\sum_{i=1}^n \mathbb{E}W_i$. Our goal is thus to bound the ratio

$$\varrho_\omega = \frac{\omega\mathbb{E}W_n^{\text{SVF}} + (1-\omega)\sum_{i=1}^n \mathbb{E}W_i^{\text{SVF}}}{\omega\mathbb{E}W_n^{\text{OPT}} + (1-\omega)\sum_{i=1}^n \mathbb{E}W_i^{\text{OPT}}}. \tag{17}$$

Now note that $W_{k+1} = \max\{0, S_1, \ldots, S_k\}$ is a convex function in each of the X_i, as it is the maximum of linear functions in X_i. Under Assumption 3.1 this implies that $\mathbb{E}W_{k+1}^{\text{OPT}} \geqslant \mathbb{E}W_{k+1}^{\text{OPT}'}$, as each step X_i with $i > k$ can be replaced by some X_j with $j \leqslant k$ and as $X_j \leqslant_{\text{cx}} X_i$ this lowers the expected waiting time. Now

$$\mathbb{E}W_{k+1}^{\text{SVF}}/\mathbb{E}W_{k+1}^{\text{OPT}} \leqslant \mathbb{E}W_{k+1}^{\text{SVF}}/\mathbb{E}W_{k+1}^{\text{OPT}'} \leqslant \varrho',$$

and so also the ratio in (17) is bounded by ϱ', which completes the proof. □

Lemma 3.9. *Under Assumption 3.2,*

$$\mathbb{E}W_{k+1} \leqslant \mathbb{E}\left(X_{\tau(1)} + \cdots + X_{\tau(k)}\right)^+ + \mathbb{E}\left(X_{\tau(1)} + \cdots + X_{\tau(k-1)}\right)^+.$$

Proof. By Lemma 3.6, W_k is stochastically dominated by $|X_{\tau(1)} + \cdots + X_{\tau(k-1)}|$. As a consequence,

$$\mathbb{E}W_{k+1} = \mathbb{E}\left(W_k + X_{\tau(k)}\right)^+ \leqslant \mathbb{E}\left(|X_{\tau(1)} + \cdots + X_{\tau(k-1)}| + X_{\tau(k)}\right)^+. \tag{18}$$

If Y and Z are independent and both have symmetric distributions, then

$$\mathbb{E}(|Y| + Z)^+ = \mathbb{E}(Y+Z)^+ + \mathbb{E}Y^+.$$

This is easily checked by conditioning on $|Y| = a$ and $|Z| = b$, as then Y is either a or $-a$ with probability $\frac{1}{2}$, and similarly for Z. Applying this result to the upper bound in (18), we find the upper bound in the lemma. □

Lemma 3.10. *Under Assumption 3.2, for any ℓ,*

$$\mathbb{E}W_{k+1} \geqslant \tfrac{1}{2}\Big(\mathbb{E}\left(X_{\tau(1)} + \cdots + X_{\tau(k)}\right)^+ + \mathbb{E}\left(X_{\tau(1)} + \cdots + X_{\tau(\ell)}\right)^+ + \mathbb{E}\left(X_{\tau(\ell+1)} + \cdots + X_{\tau(k)}\right)^+\Big).$$

Proof. Let $S'_\ell = X_{\tau(1)}+X_{\tau(2)}+\cdots+X_{\tau(\ell)}$, so that $S'_\ell = S_k - S_{k-\ell}$. As $W_{k+1} = \max\{0, S_1, \ldots, S_k\}$, we then have

$$W_{k+1} \geqslant \max\{0, S_{k-\ell}, S_k\} = \max\{0, S_{k-\ell}, S_{k-\ell} + S'_\ell\} = (S_{k-\ell} + (S'_\ell)^+)^+. \tag{19}$$

If Y and Z are independent and both have symmetric distributions, then

$$\mathbb{E}(Y + Z^+)^+ = \tfrac{1}{2}\left(\mathbb{E}(Y+Z)^+ + \mathbb{E}Y^+ + \mathbb{E}Z^+\right).$$

Again, this is easily checked by conditioning on $|Y| = a$ and $|Z| = b$, as then Y is either a or $-a$ with probability $\frac{1}{2}$, and similarly for Z. Applying this result to the lower bound in (19), we find the lower bound in the lemma. □

Appendix B. Proofs corresponding to Section 4

Here we prove Proposition 4.3, Proposition 4.4, and Theorem 4.5 from Section 4. In order to prove Proposition 4.3, we need a number of lemmas.

Lemma B.1. *Let M be the maximum of a random walk with steps $Y_1, \ldots, Y_k$. Now let $i \in \{1, \ldots, k\}$, and let $c \geqslant 1$. Let M' be defined as the maximum of a random walk with steps $Y_1, \ldots, Y_i, c\,Y_{i+1}, \ldots, c\,Y_k$. Then $M \leqslant M'$.*

Proof. Suppose that the maximum of the first random walk is attained at time j, that is, $M = Y_1+\cdots+Y_j$. If $j \leqslant i$, then the second random walk also attains the value $Y_1+\cdots+Y_j$, so $M \leqslant M'$. If $j > i$, then the second random walk attains the value $Y_1+\cdots+Y_i+cY_{i+1}+\cdots+cY_j$. Now $Y_{i+1} + \cdots + Y_j \geqslant 0$, as otherwise $Y_1 + \cdots + Y_i > Y_1 + \cdots + Y_j$, in contradiction with M being the maximum. From this and $c \geqslant 1$ it follows that

$$Y_1 + \cdots + Y_i + c\,Y_{i+1} + \cdots + c\,Y_j \geqslant Y_1 + \cdots + Y_j,$$

so $M \leqslant M'$. □

Lemma B.2. *Suppose Assumption* 4.1 *holds, and we use the schedule $\boldsymbol{x} = \boldsymbol{\mu} + \alpha\boldsymbol{\sigma}$. Let M_k be the all-time maximum of a random walk with i.i.d. steps distributed as $\sigma_k(B-\alpha)$. Then W^{SVF}_{k+1} is stochastically dominated by M_k, for all k.*

Proof. By (3), we know that W^{SVF}_{k+1} is the maximum of a random walk with steps $B_k - x_k, B_{k-1} - x_{k-1}, \ldots, B_1 - x_1$. Now note that $B_i \overset{\text{d}}{=} \mu_i + \sigma_i B$ and $x_i = \mu_i + \alpha\sigma_i$, hence $B_i - x_i \overset{\text{d}}{=} \sigma_i(B-\alpha)$. So W^{SVF}_{k+1} can be represented as the maximum of a random walk with steps distributed as $\sigma_k(B-\alpha), \ldots, \sigma_1(B-\alpha)$.

We first multiply the last step of this random walk σ_2/σ_1. By Lemma B.1, we then see that W_{k+1} is stochastically dominated by the maximum of a random walk with steps distributed as $\sigma_k(B-\alpha),\ldots,\sigma_2(B-\alpha),\sigma_2(B-\alpha)$. The next step is to multiply the last two steps with σ_3/σ_2. Again by Lemma B.1, W_{k+1}^{SVF} is stochastically dominated by the maximum of a random walk with steps $\sigma_k(B-\alpha),\ldots,\sigma_3(B-\alpha)$, $\sigma_3(B-\alpha)$, $\sigma_3(B-\alpha)$. Continuing in this way, we find that W_{k+1}^{SVF} is stochastically dominated by the maximum of a random walk with k steps distributed as $\sigma_k(B-\alpha)$.

Now note that adding extra steps to a random walk can only increase the maximum. Therefore, we now find that W_{k+1}^{SVF} is stochastically dominated by M_k. □

Lemma B.3. *Under Assumption* 4.1*, when using the schedule* $\boldsymbol{x}=\boldsymbol{\mu}+\alpha\boldsymbol{\sigma}$*, we have for all* k *that*

$$\mathbb{E}W_{k+1}^{\text{SVF}}\leqslant\frac{\sigma_k}{2\alpha}.$$

Proof. We need the following bound by Kingman [20]. Let M be the all-time maximum of a random walk with i.i.d. steps distributed as Y, with $\mathbb{E}Y<0$. Then

$$\mathbb{E}M\leqslant\frac{\operatorname{Var}(Y)}{2|\mathbb{E}Y|}.$$

Let M_k be as in Lemma B.2. Then, by Lemma B.3 and Kingman's bound, we have

$$\mathbb{E}W_{k+1}^{\text{SVF}}\leqslant\mathbb{E}M_k\leqslant\frac{\operatorname{Var}(\sigma_k(B-\alpha))}{2|\mathbb{E}(\sigma_k(B-\alpha))|}=\frac{\sigma_k^2}{2\sigma_k\alpha}=\frac{\sigma_k}{2\alpha},$$

as claimed. □

Now we are ready to prove Proposition 4.3.

Proposition 4.3. *Suppose* α *is given by* (10)*. Under Assumption* 4.1*,*

$$C_\omega(\text{id},\boldsymbol{\mu}+\alpha\boldsymbol{\sigma})\leqslant\sqrt{2\omega}\sum_{i=1}^{n-1}\sigma_i.$$

Proof. We already have a bound on the mean waiting time in Lemma B.3, so we proceed by considering the mean idle time. Taking expectations in (5) and using the fact that $x_i=\mu_i+\alpha\sigma_i$,

$$\sum_{i=1}^{n}\mathbb{E}I_i^{\text{SVF}}+\sum_{i=1}^{n}\mu_i=\sum_{i=1}^{n}\mu_i+\alpha\sum_{i=1}^{n-1}\sigma_i+\mathbb{E}W_n^{\text{SVF}}.$$

Hence, by virtue of Lemma B.3,

$$\sum_{i=1}^{n}\mathbb{E}I_i^{\text{SVF}}=\alpha\sum_{i=1}^{n-1}\sigma_i+\mathbb{E}W_n^{\text{SVF}}\leqslant\alpha\sum_{i=1}^{n-1}\sigma_i+\frac{\sigma_{n-1}}{2\alpha}.$$

Upon combining the bounds for the mean waiting times and the total mean idle time, we find that

$$C_\omega(\mathrm{id}, \boldsymbol{\mu} + \alpha\boldsymbol{\sigma}) \leqslant \alpha\omega \sum_{i=1}^{n-1} \sigma_i + \frac{\omega}{2\alpha}\sigma_{n-1} + \frac{1-\omega}{2\alpha} \sum_{i=1}^{n-1} \sigma_i.$$

Through standard calculus we find that this upper bound is minimized for the α given in (10). The corresponding upper bound for this α is

$$\sqrt{2\omega(1-\omega)} \sqrt{\sum_{i=1}^{n-1} \sigma_i \left(\sum_{i=1}^{n-1} \sigma_i + \frac{\omega}{1-\omega}\sigma_{n-1} \right)}.$$

Using the fact that $\sigma_{n-1} \leqslant \sum_{i=1}^{n-1} \sigma_i$, we find the upper bound in Proposition 4.3. □

In order to prove Proposition 4.4, we need the following lemma, which is known from the classical newsvendor problem (see e.g. [19]).

Lemma B.4. *Let X be a random variable, with Q_X its quantile function. Then for $\omega \in (0,1)$,*

$$\omega\mathbb{E}(X-c)^- + (1-\omega)\mathbb{E}(X-c)^+$$

is minimal for $c = Q_X(1-\omega) = \inf\{x : 1-\omega \leqslant \mathbb{P}(X \leqslant x)\}$.

Proposition 4.4. *Under Assumption* 4.1*, for any sequence and schedule,*

$$C_\omega(\tau, \boldsymbol{x}) \geqslant \left[\omega\mathbb{E}B(\omega)^- + (1-\omega)\mathbb{E}B(\omega)^+\right] \sum_{i=1}^{n-1} \sigma_i.$$

Proof. Consider some arbitrary sequence τ and schedule $\boldsymbol{x}$. Recall that the idle and waiting times satisfy the recursions in (1). We consider

$$\omega\mathbb{E}I_{k+1} + (1-\omega)\mathbb{E}W_{k+1} = \omega\mathbb{E}\left(W_k + B_{\tau(k)} - x_{\tau(k)}\right)^- + (1-\omega)\mathbb{E}\left(W_k + B_{\tau(k)} - x_{\tau(k)}\right)^+. \tag{20}$$

Now note that $W_k + B_{\tau(k)} - x_{\tau(k)} = B_{\tau(k)} - (x_{\tau(k)} - W_k)$, so by minimizing (20) over all possible values of $x_{\tau(k)} - W_k$, we find by Lemma B.4 that

$$\omega\mathbb{E}I_{k+1} + (1-\omega)\mathbb{E}W_{k+1} \geqslant \omega\mathbb{E}(B_{\tau(k)} - Q_{B_{\tau(k)}}(1-\omega))^- + (1-\omega)\mathbb{E}(B_{\tau(k)} - Q_{B_{\tau(k)}}(1-\omega))^+.$$

Because $B_{\tau(k)} \overset{\mathrm{d}}{=} \mu_{\tau(k)} + \sigma_{\tau(k)}B$, it follows that $Q_{B_{\tau(k)}}(1-\omega) = \mu_{\tau(k)} + \sigma_{\tau(k)}Q_B(1-\omega)$. Then

$$B_{\tau(k)} - Q_{B_{\tau(k)}}(1-\omega) \overset{\mathrm{d}}{=} \mu_{\tau(k)} + \sigma_{\tau(k)}B - (\mu_{\tau(k)} + \sigma_{\tau(k)}Q_B(1-\omega)) = \sigma_{\tau(k)}(B - Q_B(1-\omega)),$$

which, recalling that $B(\omega) = B - Q_B(1-\omega)$, leads to

$$\omega\mathbb{E}I_{k+1} + (1-\omega)\mathbb{E}W_{k+1} \geqslant \sigma_{\tau(k)}\left[\omega\mathbb{E}B(\omega)^- + (1-\omega)\mathbb{E}B(\omega)^+\right].$$

Note that $\sum_{i=1}^{n-1}\sigma_{\tau(i)} \geqslant \sum_{i=1}^{n-1}\sigma_i$, as the σ_i were put in increasing order. Now summing over k we find the lower bound of Proposition 4.4. □

To prove Theorem 4.5 we need the following lemma, which fulfills the same role for lognormally distributed service times as Lemma B.2 does for location-scale families. We then prove the upper and lower bounds needed for Theorem 4.5 in two propositions.

Lemma B.5. *Suppose the B_i are lognormally distributed with $m_1 \leqslant \ldots \leqslant m_n$ and $s_1^2 \leqslant \ldots \leqslant s_n^2$, and that we use the schedule $\boldsymbol{x} = (1+\alpha)\boldsymbol{\mu}$. Let M_k be the all-time maximum of a random walk with i.i.d. steps distributed as $B_k - x_k$. Then $\mathbb{E}W_{k+1}^{\text{SVF}} \leqslant \mathbb{E}M_k$.*

Proof. Let $M_k^{(i)}$ be the maximum of a random walk with steps distributed as $B_k - x_k, B_{k-1} - x_{k-1}, \ldots, B_{i+1} - x_{i+1}$ followed by i steps distributed as $B_i - x_i$. We will prove that $\mathbb{E}M_k^{(i)} \leqslant \mathbb{E}M_k^{(i+1)}$ for $i = 1, \ldots, k-1$. As $\mathbb{E}W_{k+1}^{\text{SVF}} = \mathbb{E}M_k^{(1)}$ and $\mathbb{E}M_k^{(k)} \leqslant \mathbb{E}M_k$ (adding extra steps increases the maximum), it then follows that $\mathbb{E}W_{k+1}^{\text{SVF}} \leqslant \mathbb{E}M_k$.

Consider the random walk with steps distributed as $B_k - x_k, B_{k-1} - x_{k-1}, \ldots, B_{i+1} - x_{i+1}$, followed by i steps distributed as $B_i - x_i$. Let Z be a standard normal random variable. Note that

$$\begin{aligned} B_i - x_i = B_i - (1+\alpha)\mu_i &\overset{\mathrm{d}}{=} \exp(m_i + s_i Z) - (1+\alpha)\exp(m_i + s_i^2/2) \\ &= \exp(m_i)\left[\exp(s_i Z) - (1+\alpha)\exp(s_i^2/2)\right]. \end{aligned}$$

Use Lemma B.1 to show we get an upper bound on $M_k^{(i)}$ by replacing all steps distributed as $B_i - x_i$ by steps distributed as

$$X_i' := \exp(m_{i+1} + (s_{i+1}^2 - s_i^2)/2)\left[\exp(s_i Z) - (1+\alpha)\exp(s_i^2/2)\right].$$

Let Z' be another standard normal random variable independent of Z. Then

$$B_{i+1} - x_{i+1} \overset{\mathrm{d}}{=} \exp(m_{i+1} + s_i Z + \epsilon Z') - (1+\alpha)\exp(m_{i+1} + (s_i^2 + \epsilon^2)/2)$$

with $\epsilon := \sqrt{s_{i+1}^2 - s_i^2}$, and

$$\begin{aligned} &\mathbb{E}[\exp(m_{i+1} + s_i Z + \epsilon Z') - (1+\alpha)\exp(m_{i+1} + (s_i^2 + \epsilon^2)/2)|X_i'] \\ &= \mathbb{E}[\exp(m_{i+1} + s_i Z + \epsilon Z') - (1+\alpha)\exp(m_{i+1} + (s_i^2 + \epsilon^2)/2)|Z] \\ &= \exp(m_i + \epsilon^2/2)\left[\exp(s_i Z) - (1+\alpha)\exp(s_i^2/2)\right] = X_i'. \end{aligned}$$

It follows by Lemma 2.3 that $X_i' \leqslant_{\mathrm{cx}} B_{i+1} - x_{i+1}$. As the maximum of a random walk is a convex function in each of the individual stepsizes, we can replace each step distributed as X_i' by one distributed as $B_{i+1} - x_{i+1}$. Therefore, $\mathbb{E}M_k^{(i)} \leqslant \mathbb{E}M_k^{(i+1)}$, which completes the proof. □

Proposition B.6. *Suppose the B_i are lognormally distributed with $m_1 \leqslant \ldots \leqslant m_n$ and $s_1^2 \leqslant \ldots \leqslant s_n^2$. Suppose α is given by*

$$\alpha = \frac{1}{\sqrt{2\omega}}\sqrt{(\exp(s_{n-1}^2) - 1)}.$$

Then

$$C_\omega(\mathrm{id}, (1+\alpha)\boldsymbol{\mu}) \leqslant 2\omega\alpha \sum_{i=1}^{n-1} \exp(m_i + s_i^2/2).$$

Proof. With M_k as in Lemma B.5, we can now apply Kingman's bound to find

$$\begin{aligned}\mathbb{E}W_{k+1}^{\mathrm{SVF}} \leqslant \mathbb{E}M_k \leqslant \frac{\mathrm{Var}(B_k)}{2|\mathbb{E}B_k - (1+\alpha)\mathbb{E}B_k|} &= \frac{1}{2\alpha}(\exp(s_k^2) - 1)\exp(m_k + s_k^2/2) \\ &\leqslant \frac{1}{2\alpha}(\exp(s_{n-1}^2) - 1)\exp(m_k + s_k^2/2).\end{aligned}$$

By equation (5) we also have

$$\begin{aligned}\sum_{i=1}^{n} \mathbb{E}I_i^{\mathrm{SVF}} &= \alpha \sum_{i=1}^{n-1} \mathbb{E}B_i + \mathbb{E}W_n^{\mathrm{SVF}} \\ &\leqslant \alpha \sum_{i=1}^{n-1} \exp(m_i + s_i^2/2) + \frac{1}{2\alpha}(\exp(s_{n-1}^2) - 1)\exp(m_{n-1} + s_{n-1}^2/2) \\ &\leqslant \alpha \sum_{i=1}^{n-1} \exp(m_i + s_i^2/2) + \frac{1}{2\alpha}(\exp(s_{n-1}^2) - 1)\sum_{i=1}^{n-1} \exp(m_i + s_i^2/2).\end{aligned}$$

In total, we then find

$$C_\omega(\mathrm{id}, (1+\alpha)\boldsymbol{x}) \leqslant \omega\alpha \sum_{i=1}^{n-1} \exp(m_i + s_i^2/2) + \frac{1}{2\alpha}(\exp(s_{n-1}^2) - 1)\sum_{i=1}^{n-1} \exp(m_i + s_i^2/2).$$

Minimizing this upper bound over α, we obtain the result. □

Proposition B.7. *Suppose the B_i are lognormally distributed with $m_1 \leqslant \ldots \leqslant m_n$ and $s_1^2 \leqslant \ldots \leqslant s_n^2$. Then, for any sequence and schedule,*

$$C_\omega(\tau, \boldsymbol{x}) \geqslant [(1-\omega)\mathbb{P}(Z \geqslant Q_Z(1-\omega) - s_1) - \omega\mathbb{P}(Z \leqslant Q_Z(1-\omega) - s_1)] \sum_{i=1}^{n-1} \exp(m_i + s_i^2/2).$$

Proof. Similar as in the location-scale family case, we can use Lemma B.4 to find

$$\omega\mathbb{E}I_{k+1} + (1-\omega)\mathbb{E}W_{k+1} = \omega\mathbb{E}(B_{\tau(k)} - (x_{\tau(k)} - W_k))^- + (1-\omega)\mathbb{E}(B_{\tau(k)} - (x_{\tau(k)} - W_k))^+$$

$$\geqslant \omega\mathbb{E}(B_{\tau(k)} - Q_{B_{\tau(k)}}(1-\omega))^- + (1-\omega)\mathbb{E}(B_{\tau(k)} - Q_{B_{\tau(k)}}(1-\omega))^+.$$

Computing this lower bound, we find with Z being a standard normal random variable that

$$\begin{aligned}&\omega\mathbb{E}I_{k+1} + (1-\omega)\mathbb{E}W_{k+1}\\&\quad\geqslant \exp(m_{\tau(k)} + s_{\tau(k)}^2/2)\left[(1-\omega)\mathbb{P}(Z \geqslant Q_Z(1-\omega) - s_{\tau(k)}) - \omega\mathbb{P}(Z \leqslant Q_Z(1-\omega) - s_{\tau(k)})\right].\end{aligned}$$

It can be easily seen that

$$(1-\omega)\mathbb{P}(Z \geqslant Q_Z(1-\omega) - s_{\tau(k)}) - \omega\mathbb{P}(Z \leqslant Q_Z(1-\omega) - s_{\tau(k)})$$

is an increasing function in $s_{\tau(k)}$, and so is minimal for $s_{\tau(k)} = s_1$. Using this and summing over k, we find the lower bound of the proposition. □

Proof of Theorem 4.5. This follows directly from Proposition B.6 and Proposition B.7. □

Appendix C. Proofs corresponding to Section 5

Here we prove Lemma 5.10, that is used in Section 5.

Lemma 5.10. *For any k and $n \geqslant k$:*

(1) $\dfrac{\sum_{i=1}^k \mathbb{E}\left[(X_{n,i})\mathbb{1}_{X_{n,i}>c_{n,k}}\right]}{\Sigma_{n,k}} \leqslant q_k\gamma_{n,k}^{(1+\delta)/\delta}$; *and*

(2) $\dfrac{(\Sigma'_{n,k})^2}{\Sigma_{n,k}^2} \geqslant 1 - 2q_k\gamma_{n,k}$.

Proof. (1) Using that $X_{n,i} > c_{n,k}$ implies $|X_{n,i}|^{1+\delta}/c_{n,k}^{1+\delta} > 1$, we find

$$\mathbb{E}\left[(X_{n,i})\mathbb{1}_{X_{n,i}>c_{n,k}}\right] \leqslant \mathbb{E}\left[\left(|X_{n,i}|^{2+\delta}/c_{n,k}^{1+\delta}\right)\mathbb{1}_{X_{n,i}>c_{n,k}}\right] \leqslant \mathbb{E}|X_{n,i}|^{2+\delta}/c_{n,k}^{1+\delta}.$$

Summing over i and dividing both sides by $\Sigma_{n,k}$ we find

$$\frac{\sum_{i=1}^k \mathbb{E}\left[(X_{n,i})\mathbb{1}_{X_{n,i}>c_{n,k}}\right]}{\Sigma_{n,k}} \leqslant \frac{\sum_{i=1}^n \mathbb{E}|X_{n,i}|^{2+\delta}}{\Sigma_{n,k}c_{n,k}^{1+\delta}} = \frac{\sum_{i=1}^n \mathbb{E}|X_{n,i}|^{2+\delta}}{\Sigma_{n,k}^{2+\delta}}\left(\frac{\Sigma_{n,k}}{c_{n,k}}\right)^{1+\delta} \leqslant q_k\gamma_{n,k}^{(1+\delta)/\delta},$$

as was claimed.

(2) Analogous to the proof of part (1), we can also deduce that

$$\frac{\sum_{i=1}^k \mathbb{E}\left[X_{n,i}^2\mathbb{1}_{X_{n,i}>c_{n,k}}\right]}{\Sigma_{n,k}^2} \leqslant q_k\gamma_{n,k}. \tag{21}$$

Using that $\mathbb{E}X_{n,i} = 0$, we find

$$\frac{(\Sigma'_{n,k})^2}{\Sigma_{n,k}^2} = \frac{\sum_{i=1}^n \mathbb{E}\left[X_{n,i}^2\mathbb{1}_{X_{n,i}\leqslant c_{n,k}}\right]}{\Sigma_{n,k}^2} - \frac{\sum_{i=1}^n \left(\mathbb{E}\left[X_{n,i}\mathbb{1}_{X_{n,i}\leqslant c_{n,k}}\right]\right)^2}{\Sigma_{n,k}^2}$$

$$= \frac{\Sigma_{n,k}^2 - \sum_{i=1}^n \mathbb{E}\left[X_{n,i}^2 \mathbb{1}_{X_{n,i}>c_{n,k}}\right]}{\Sigma_{n,k}^2} - \frac{\sum_{i=1}^n \left(\mathbb{E}\left[X_{n,i}\mathbb{1}_{X_{n,i}>c_{n,k}}\right]\right)^2}{\Sigma_{n,k}^2}.$$

Now applying Jensen's inequality to the last term and (21),

$$\frac{(\Sigma'_{n,k})^2}{\Sigma_{n,k}^2} \geqslant \frac{\Sigma_{n,k}^2 - 2\sum_{i=1}^n \mathbb{E}\left[X_{n,i}^2 \mathbb{1}_{X_{n,i}>c_{n,k}}\right]}{\Sigma_{n,k}^2} \geqslant 1 - 2q_k\gamma_{n,k},$$

which is what we wanted to prove. □

Appendix D. Lognormal distributions that satisfy the dilation order

Proposition D.1. *Suppose that A and B are lognormally distributed random variables, such that $\mathbb{E}[\ln A] \leqslant \mathbb{E}[\ln B]$ and $\mathrm{Var}(\ln A) \leqslant \mathrm{Var}(\ln B)$. Then $A \leqslant_{\mathrm{dil}} B$.*

Proof. Let $m_1 = \mathbb{E}[\ln A]$, $m_2 = \mathbb{E}[\ln B]$, $s_1^2 = \mathrm{Var}(\ln A)$ and $s_2^2 = \mathrm{Var}(\ln B)$. Let Z and Z' be two independent standard normal random variables, and let $\epsilon = \sqrt{s_2^2 - s_1^2}$. Then

$$A - \mathbb{E}A \overset{\mathrm{d}}{=} \widehat{X} := \exp(m_1 + s_1 Z) - \exp(m_1 + s_1^2/2),$$
$$B - \mathbb{E}B \overset{\mathrm{d}}{=} \widehat{Y} := \exp(m_2 + s_1 Z + \epsilon Z') - \exp(m_2 + (s_1^2 + \epsilon^2)/2).$$

Now note that

$$\begin{aligned}\mathbb{E}[\widehat{Y}|\widehat{X}] = \mathbb{E}[\widehat{Y}|Z] &= \mathbb{E}\left[\exp(m_2 + s_1 Z + \epsilon Z') - \exp(m_2 + (s_1^2 + \epsilon^2)/2)|Z\right] \\ &= \exp(m_2 + \epsilon^2/2)\left(\exp(s_1 Z) - \exp(s_1^2/2)\right) \\ &= \exp(m_2 - m_1 + \epsilon^2/2)\widehat{X}.\end{aligned}$$

Thus by Lemma 2.3,

$$\exp(m_2 - m_1 + \epsilon^2/2)(A - \mathbb{E}A) \leqslant_{\mathrm{cx}} B - \mathbb{E}B.$$

(This is similar to the argument that $X'_i \leqslant_{\mathrm{cx}} B_{i+1} - x_{i+1}$ in the proof of Lemma B.5.) Now we can apply Theorem 3.A.18 from Shaked and Shanthikumar [33], that says that $X \leqslant_{\mathrm{dil}} aX$ for $a \geqslant 1$, to see that

$$A - \mathbb{E}A \leqslant_{\mathrm{cx}} \exp(m_2 - m_1 + \epsilon^2/2)(A - \mathbb{E}A) \leqslant_{\mathrm{cx}} B - \mathbb{E}B.$$

This proves that $A \leqslant_{\mathrm{dil}} B$. □

References

[1] A. Ahmadi-Javid, Z. Jalali, K. Klassen (2017). *Outpatient appointment systems in healthcare: A review of optimization studies*, European Journal of Operational Research, **258**(1), 3-34.

[2] M. A. Begen, M. Queyranne (2011). *Appointment scheduling with discrete random durations*, Mathematics of Operations Research, **36**(2), 240-257.

[3] B. Berg, B. Denton, S. Erdogan, T. Rohleder, T. Huschka (2014). *Optimal booking and scheduling in outpatient procedure centers*, Computers & Operations Research, **50**, 24-37.

[4] P. Billingsley (1995). *Probability and Measure*, Third edition. Wiley, New York.

[5] T. Çayırlı, E. Veral (2003). *Outpatient scheduling in health care: a review of literature*, Production and Operations Management, **12**(4), 519-549.

[6] T. Çayırlı, E. Veral, H. Rosen (2006). *Designing appointment scheduling systems for ambulatory care services*, Health Care Management Science, **9**(1), 47-58.

[7] T. Çayırlı, E. Veral, H. Rosen (2008). *Assessment of patient classification in appointment system design*, Production and Operations Management, **17**(3), 338-353.

[8] J. Charnetski (1984). *Scheduling operating room surgical procedures with early and late completion penalty costs*, Journal of Operations Management, **5**(1), 91-102.

[9] B. Denton, D. Gupta (2003). *A sequential bounding approach for optimal appointment scheduling*, IIE Transactions, **35**(11), 1003-1016.

[10] B. Denton, J. Viapiano, A. Vogl (2007). *Optimization of surgery sequencing and scheduling decisions under uncertainty*, Health Care Management Science, **10**(1), 13-24.

[11] A. Erdogan, A. Gose, B. Denton (2015). *Online appointment sequencing and scheduling*, IIE Transactions, **47**(11), 1267-1286.

[12] H. Guda, M. Dawande, G. Janakiraman, K. S. Jung (2016). *Optimal policy for a stochastic scheduling problem with applications to surgical scheduling*, Production and Operations Management, **25**(7), 1194-1202.

[13] D. Gupta (2007). *Surgical suites' operations management*, Production and Operations Management, **16**(6), 689-700.

[14] D. Gupta, B. Denton (2008). *Appointment scheduling in health care: Challenges and opportunities*, IIE Transactions, **40**(9), 800-819.

[15] A. Gupta, A. Kumar, V. Nagarajan, X. Shen (2018). *Stochastic load balancing on unrelated machines*, Proceedings of the SODA 2018 Conference, 1274-1285.

[16] V. Gupta, B. Moseley, M. Uetz, Q. Xie (2017). *Stochastic online scheduling on unrelated machines*, Integer Programming and Combinatorial Optimization, Springer International Publishing, 228-240.

[17] C. C. Heyde, B. M. Brown (1970). *On the departure from normality of a certain class of martingales*, Annals of Mathematical Statistics, **41**(6), 2161-2165.

[18] B. Kemper, C. A. J. Klaassen, M. Mandjes (2014). *Optimized appointment scheduling*, European Journal of Operational Research, **239**(1), 243-255.

[19] M. Khouja (1999). *The single-period (news-vendor) problem: literature review and suggestions for future research*, Omega, **27**, 537-553.

[20] J. F. C. Kingman (1962). *Some inequalities for the queue GI/G/1*, Biometrika, **49**(3-4), 315-324.

[21] K. Klassen, T. Rohleder (1996). *Scheduling outpatient appointments in a dynamic environment*, Journal of Operations Management, **14**(2), 83-101.

[22] Q. Kong, C. Lee, C. Teo, Z. Zheng (2013). *Scheduling arrivals to a stochastic service delivery system using copositive cones*, Operations Research, **61**(3), 711-726.

[23] Q. Kong, C. Lee, C. Teo, Z. Zheng (2016). *Appointment sequencing: Why the smallest-variance-first rule may not be optimal*, European Journal of Operational Research, **255**(3), 809-821.

[24] A. Kuiper, B. Kemper, M. Mandjes (2015). *A computational approach to optimized appointment scheduling*, Queueing Systems, **79**(1), 5-36.

[25] D. V. Lindley (1952). *The theory of queues with a single server*, Mathematical Proceedings of the Cambridge Philosophical Society, **48**(2), 227-289.

[26] H. Mak, Y. Rong, J. Zhang (2014). *Sequencing appointments for service systems using inventory approximations*, Manufacturing & Service Operations Management, **16**(2), 251-262.

[27] H. Mak, Y. Rong, J. Zhang (2015). *Appointment scheduling with limited distributional information*, Management Science, **61**(2), 316-334.

[28] C. Mancilla, R. Storer (2012). *A sample average approximation approach to stochastic appointment sequencing and scheduling*, IIE Transactions, **44**(8), 655-670.

[29] S. Mittal, A. S. Schulz, S. Stiller (2014). *Robust appointment scheduling*, Proceedings of APPROX/RANDOM 2014, 356-370.

[30] M. Pinedo (2012). *Scheduling: Theory, Algorithms, and Systems*, Springer, New York.

[31] L. W. Robinson, R. R. Chen (2011). *Estimating the implied value of the customer's waiting time*, Manufacturing & Service Operations Management, **13**(1), 53-57.

[32] T. Rohleder, K. Klassen (2000). *Using client-variance information to improve dynamic appointment scheduling performance*, Omega, **28**(3), 293-302.

[33] M. Shaked, J.G. Shanthikumar (2007). *Stochastic Orders*. Springer, New York.

[34] M. Skutella, M. Sviridenko, M. Uetz (2016). *Stochastic scheduling on unrelated machines*, Mathematics of Operations Research, **41**(3), 851-864.

[35] P. M. Vanden Bosch, D. C. Dietz (2000). *Minimizing expected waiting in a medical appointment system*, IIE Transactions, **32**(9), 841-848.

[36] E. N. Weiss (1990). *Models for determining estimated start times and case orderings in hospital operating rooms*, IIE Transactions, **22**(2), 143-150.