



# Non-indexability of the stochastic appointment scheduling problem<sup>☆</sup>

Mehdi Jafarnia-Jahromi<sup>a,\*</sup>, Rahul Jain<sup>b</sup>

<sup>a</sup> ECE Department, University of Southern California, Los Angeles, CA, USA

<sup>b</sup> ECE, ISE & CS Department, University of Southern California, Los Angeles, CA, USA



## ARTICLE INFO

### Article history:

Received 15 March 2019

Received in revised form 9 January 2020

Accepted 12 April 2020

Available online xxxx

### Keywords:

Operations research applications  
Stochastic appointment scheduling  
Sequencing  
Sample average approximation

## ABSTRACT

Consider a set of jobs with independent random service times to be scheduled on a single machine. The jobs can be surgeries in an operating room, patients' appointments in outpatient clinics, etc. The challenge is to determine the optimal sequence and appointment times of jobs to minimize some function of the server idle time and service start-time delay. We introduce a generalized objective function of delay and idle time, and consider  $l_1$ -type and  $l_2$ -type cost functions as special cases of interest. Determining an index-based policy for the optimal sequence in which to schedule jobs has been an open problem for many years. For example, it was conjectured that 'least variance first' (LVF) policy is optimal for the  $l_1$ -type objective. This is known to be true for the case of two jobs with specific distributions. A key result in this paper is that the optimal sequencing problem is non-indexable, i.e., neither the variance, nor any other such index can be used to determine the optimal sequence in which to schedule jobs for  $l_1$  and  $l_2$ -type objectives. We then show that given a sequence in which to schedule the jobs, sample average approximation yields a solution which is statistically consistent.

© 2020 Elsevier Ltd. All rights reserved.

## 1. Introduction

Scheduling is an important aspect for efficient resource utilization, and there is a vast literature on the topic dating back several decades (see for example, Baker, 1974; Conway, Maxwell, & Miller, 2003; Erdogan, Gose, & Denton, 2015; Pinedo, 2012). The problem considered in this paper is stochastic appointment scheduling which has various applications in scheduling of surgeries at operating rooms, appointments in outpatient clinics, cargo ships at seaports, etc.

**Problem Statement:** Stochastic appointment scheduling problem (ASP) has a simple statement: Consider a finite set of  $n$  jobs to be scheduled on a single machine. Job durations are random with known distributions. If a job is completed before the appointment time of the subsequent job, the server will remain idle. Conversely, if it lasts beyond the allocated slot, the following job will be delayed. We need to determine optimal appointment times so that the expectation of a function of idle time and delay is minimized. Thus, the ASP addresses two important questions. First, given the sequence of jobs, what are the optimal appointment

times? This is called the *scheduling* problem. Second, in what sequence should the jobs be served? This is called the *sequencing* problem. We note that while we call it the *sequencing problem* for brevity, it really refers to the joint sequencing–scheduling problem of determining both the optimal sequence as well as the appointment times.

Examples of the scheduling problem are cargo ships at seaports that must be given time to berth in the order in which they arrive. Examples of the sequencing problem are surgeries in a single operating room that must be scheduled the day before in an order that minimizes certain metrics such as wait times and idle time.

Intuitively speaking, scheduling jobs with more uncertain durations first may lead to delay propagation through the schedule. This intuition has motivated many researchers to prove optimality of 'least variance first' (LVF) policy. However, efforts beyond the case of  $n = 2$  have not been fruitful for specific distributions such as uniform and exponential (Wang, 1999; Weiss, 1990). Recent numerical work of Mansourifard, Mansourifard, Ziyadi, and Krishnamachari (2018) has argued that LVF is not the best heuristic for the sequencing problem in the case that idle time and delay unit costs are not balanced. They introduce 'newsvendor' index as another heuristic that outperforms variance. However, no proof of optimality is provided. This controversy raises an important open question that: *Is there an index (a map from a random variable to the reals) that yields the optimal sequence?*

In this paper, both the *scheduling* and *sequencing* problems are addressed. In the *sequencing* problem, we introduce an index

<sup>☆</sup> The material in this paper was partially presented at INFORMS 19th Applied Probability Society Conference, July 10–12, 2017, Northwestern University, Evanston, Illinois, USA. This paper was recommended for publication in revised form by Associate Editor Kok Lay Teo under the direction of Editor Ian R. Petersen.

\* Corresponding author.

E-mail addresses: [mjafarni@usc.edu](mailto:mjafarni@usc.edu) (M. Jafarnia-Jahromi), [rahul.jain@usc.edu](mailto:rahul.jain@usc.edu) (R. Jain).

(a map from a random variable to the reals) and prove that it is the only possible candidate to return the optimal sequence. This candidate index reduces to the 'Newsvendor' index and variance index for  $l_1$  and  $l_2$ -type cost functions, respectively. However, by providing counterexamples for optimality of variance and newsvendor index, we show that the *sequencing* problem is not indexable in general. Moreover, the candidate index illuminates that the heuristic sub-optimal indexing policy one might use depends on the objective function. In contrast to what has been used in the literature for a long time, variance is not the candidate index for  $l_1$ -type cost function. Instead, newsvendor should be used. This theoretical result confirms the numerical evidence of Mansourifard et al. (2018) that newsvendor outperforms variance.

In the *scheduling* problem, we prove that the  $l_1$ -type objective function introduced by Weiss (1990) is convex and there exists a solution to the stochastic optimization problem. Moreover, sample average approximation (SAA) can be used to approximate the optimization problem. We prove that SAA gives a solution that is statistically consistent. These provide a feasible computational method to compute optimal appointment times given the sequence. To the best of our knowledge this result is new from two perspectives. First, the result is proved for a generalized objective function  $c$  (as long as it is convex) which includes previously considered objective function in the literature ( $l_1$ -type objective) as a special case of interest. Second, the assumptions for consistency of SAA are considerably relaxed compared to the standard literature of SAA. In fact we prove that SAA gives a consistent result as long as there exists a schedule with finite cost. The main contributions of this paper are:

- It is rigorously proved that there exists no index that yields the optimal sequence in the stochastic appointment scheduling problem (Theorem 11). However, a candidate index is introduced that yields a heuristic sub-optimal sequence. This candidate index reduces to newsvendor and variance for  $l_1$ -type and  $l_2$ -type objective functions, respectively.
- It is proved that the objective function of Weiss (1990) for stochastic appointment scheduling problem is convex and there exists an optimal schedule (Proposition 21 in Appendix A and Theorem 17).
- It is proved that for a fixed sequence of jobs, sample average approximation yields an approximate solution for the optimal schedule that is statistically consistent (Theorem 20).

The remainder of the paper is organized as follows. In Section 2, literature is reviewed. Section 3 provides the problem formulation. In Section 4, the *sequencing* problem is discussed. The *scheduling* problem is addressed in Section 5. Section 6 provides numerical results followed by conclusions in Section 7.

## 2. Literature review

Extensive application of appointment scheduling in transportation such as bus scheduling (Wu, Liu, Jin, & Ma, 2019), as well as healthcare applications (Erdogan & Denton, 2013) has led to a vast body of literature on the topic. We note that the problem considered in this paper is a variant of a broader scheme of *railway* scheduling (Herroelen & Leus, 2004; Tian & Demeulemeester, 2014) where jobs are not allowed to start earlier than their schedule. Other variants of scheduling such as *roadrunner* scheduling (Newbold, 1998) are beyond the scope of this work. Among all the related papers, we focus on the most relevant work and classify it into *sequencing* and *scheduling*. The interested reader is referred to Ahmadi-Javid, Jalali, and

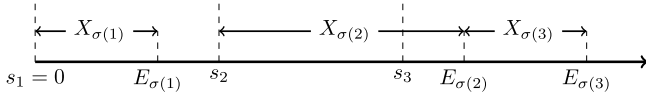
Klassen (2017), Baker (1974), Cardoen, Demeulemeester, and Bellén (2010), Cayirli and Veral (2003), Chen and Robinson (2014), Conway et al. (2003), Erdogan and Denton (2013), Erdogan et al. (2015), Gupta and Denton (2008), Hulshof, Kortbeek, Boucherie, Hans, and Bakker (2012), Kuiper and Mandjes (2015a, 2015b), Kuiper, Mandjes, and de Mast (2017), Pinedo (2012), Tian and Demeulemeester (2014) and Wu et al. (2019) and references therein for other aspects of the problem.

**Sequencing.** The *sequencing* problem we consider was first formulated by Weiss (1990). Intuitively speaking, jobs with less uncertainty should be placed first to avoid delay propagation throughout the schedule. Motivated by this intuition, a large body of literature suggested 'least variance first' (LVF) rule as a sequencing policy (see Denton, Viapiano, & Vogl, 2007; Mak, Rong, & Zhang, 2014; Qi, 2016; Wang, 1999; Weiss, 1990). However, optimality of LVF rule is only proved for the case of two jobs  $n = 2$  for certain distributions (Wang, 1999; Weiss, 1990). Mak et al. (2014) and Guda, Dawande, Janakiraman, and Jung (2016) tried to impose conditions under which LVF rule is optimal but the conditions are relatively restrictive and unlikely to hold in most scenarios of interest. A variant of the problem where jobs are allowed to start before scheduled appointments (no idle time is allowed) is studied by Baker (2014) and Guda et al. (2016). In particular, Guda et al. (2016) show that LVF rule is optimal if there exists a dilation ordering for service durations. Berg, Denton, Erdogan, Rohleder, and Hushcha (2014) and Gupta (2007) proved that if there exists a convex ordering for job durations, it is optimal to schedule smaller in convex order first for  $n = 2$ . However, their efforts for  $n > 2$  have not been fruitful. Kong, Lee, Teo, and Zheng (2016) considered likelihood ratio as a measure of variability and obtained some insights into why smallest variability first may not be optimal. Based on the insights, they provided a counterexample for non-optimality of LVF rule in the case of  $n = 6$ .

Besides the theoretical work, some papers have resorted to extensive simulation studies to investigate optimality of heuristics (see Denton et al., 2007; Klassen & Rohleder, 1996; Lebowitz, 2003; Marcon & Dexter, 2006). In particular, Denton et al. (2007) numerically showed that LVF outperforms sequencing in increasing order of mean and coefficient of variation. However, Mancilla and Storer (2012) and Mansourifard et al. (2018) argued that LVF is not the best sequencing policy especially when idle time and delay cost units are not balanced. Alternatively, Mansourifard et al. (2018) proposed a 'newsvendor' index and supported its better performance in simulations. No proof of optimality was provided.

**Scheduling.** The *scheduling* problem is also intensively studied in the literature. The seminal work of Bailey (1952) recommended to set appointment intervals equal to the average service time of each job. This approach was further pursued by Soriano (1966) and Choi and Banerjee (2016). However, letting job slots to be average service time can be near optimal only in the case that waiting cost is about 10% to 50% of the idle cost (see Denton & Gupta, 2003).

Starting with (Weiss, 1990), some papers modeled the problem using stochastic optimization to optimize on slot duration. He considered weighted sum of idle time and delay as the objective function and noticed that for the case of  $n = 2$ , the problem is equivalent to the newsvendor problem. Based on that, he proposed a heuristic estimate of the job start times for  $n > 2$ . This heuristic was extended by Kemper, Klaassen, and Mandjes (2014) to general convex function of idle time and delay. Wang (1993) and Wang (1999) considered another objective function as the weighted sum of jobs' flow time (delay and service time) and server completion time and proved its convexity. Assuming



**Fig. 1.** Appointment scheduling.  $s_i$  denotes appointment time of job  $i$ . For the realization shown in this figure, server remains idle between  $E_{\sigma(1)}$  and  $s_2$  and third job is delayed for  $E_{\sigma(2)} - s_3$  amount of time.

that the job durations are exponentially distributed, he provided a set of nonlinear equations to derive the optimum slot durations. Vink, Kuiper, Kemper, and Bhulai (2015) presented a lag order approximation by ignoring the effect of previous jobs past a certain point. Kong, Lee, Teo, and Zheng (2013) adopted a robust method over all distributions with a given mean and covariance matrix of job durations. They computationally solved for 36 jobs and showed that their solution is within 2% of the approximate optimal solution given by Denton and Gupta (2003). We refer the reader to Begen and Queyranne (2011) and Kaandorp and Koole (2007) for approaches in discrete time.

### 3. Problem formulation

We start with providing a mathematical formulation of the problem. Let  $\mathcal{X}^+$  be the space of nonnegative random variables and  $\mathbf{X} = (X_1, \dots, X_n)$  be a vector of independent random variables with components in  $\mathcal{X}^+$  and known distributions denoting jobs durations to be served on a single server. Let  $\sigma \cdot \mathbf{X}$  denote a permutation of the jobs in the order they are served, such that  $\sigma(i)$  is the  $i$ th job receiving service. Without loss of generality, we can assume that the first job starts at time zero, i.e.,  $s_1 = 0$ . Let  $\mathbf{s} = (s_2, \dots, s_n)$  be appointment times for job 2 through  $n$  in the order  $\sigma$  and let  $E_{\sigma(i)}$  be a random variable denoting the end time of job  $i$  in this order (see Fig. 1). Job  $\sigma(i)$  may finish before or after scheduled start time of the subsequent job. In the case that  $E_{\sigma(i)} \leq s_{i+1}$ , job  $\sigma(i+1)$  starts according to the schedule and server is idle between  $E_{\sigma(i)}$  and  $s_{i+1}$ . In the case where  $E_{\sigma(i)} > s_{i+1}$ , job  $\sigma(i+1)$  is delayed by  $E_{\sigma(i)} - s_{i+1}$  and will start as soon as the previous job is finished. Hence,

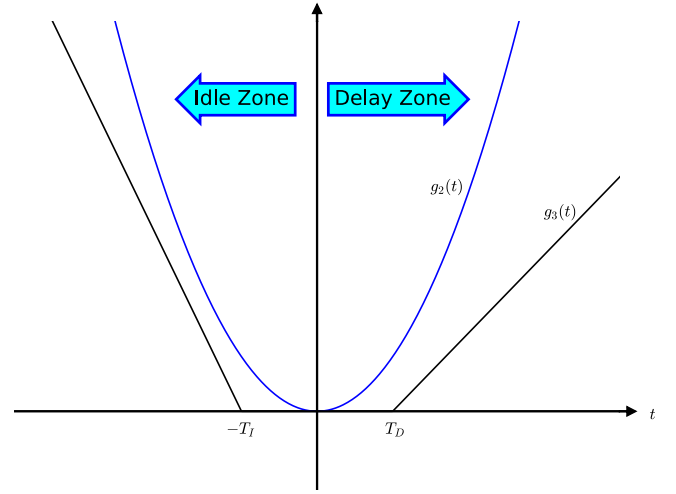
$$\begin{aligned} E_{\sigma(1)} &= X_{\sigma(1)}, \\ E_{\sigma(i)} &= \max(E_{\sigma(i-1)}, s_i) + X_{\sigma(i)}, \quad i = 2, \dots, n. \end{aligned} \quad (1)$$

Our goal is to determine appointment times such that a combination of both delay and idle time is optimized. Consider an objective function of form,

$$C(\mathbf{s}, \sigma \cdot \mathbf{X}) = \sum_{i=2}^n g(E_{\sigma(i-1)} - s_i) \quad (2)$$

where  $g: \mathbb{R} \rightarrow \mathbb{R}$  is a nonnegative, continuous and coercive function (i.e.,  $\lim_{|t| \rightarrow \infty} g(t) = \infty$ ). Furthermore, we can assume that  $g(0) = 0$  since a perfect scenario where  $E_{\sigma(i-1)} = s_i$  should not impose any cost. However, this assumption is not technically necessary. For the special case of  $g = g_1$  in Example 1, this objective function reduces to that of Weiss (1990). However, (2) is not the most general objective function one can consider. For example, in some applications, it is useful to distinguish between jobs by considering different delay per unit costs for different jobs. Moreover, (2) does not account for overtime, a related quantity that is important in some applications.

Given the schedule  $\mathbf{s}$ ,  $C(\mathbf{s}, \sigma \cdot \mathbf{X})$  captures the associative cost of the realization of job durations  $\mathbf{X}$  in the order  $\sigma$ . Thus,  $c^\sigma(\mathbf{s}) = \mathbb{E}[C(\mathbf{s}, \sigma \cdot \mathbf{X})]$  denotes the expected cost of schedule  $\mathbf{s}$  when the jobs are served in the order  $\sigma$ .



**Fig. 2.** Examples of function  $g$ .

In the scheduling problem in Section 5, we assume that the sequence of jobs is given, and we are looking for a schedule that minimizes the expected cost, i.e.,

$$\inf_{\mathbf{s} \in \mathcal{S}} c^\sigma(\mathbf{s}) \quad (3)$$

where  $\mathcal{S} = \{(s_2, \dots, s_n) \in \mathbb{R}^{n-1} \mid 0 \leq s_2 \leq \dots \leq s_n\}$  is a closed and convex subset of  $\mathbb{R}^{n-1}$ . The sequencing problem discussed in Section 4, addresses the question of finding the optimal order and appointment times of the jobs, i.e.,

$$\min_{\sigma} \inf_{\mathbf{s} \in \mathcal{S}} c^\sigma(\mathbf{s}).$$

Before proceeding with the sequencing problem, let us see some possible choices for the function  $g$ .

**Example 1.** Let  $g_1(t) = \beta(t)^+ + \alpha(-t)^+$  where  $(\cdot)^+ = \max(\cdot, 0)$  and  $\alpha, \beta > 0$ . Thus, the objective function would be

$$c_1^\sigma(\mathbf{s}) = \sum_{i=2}^n \mathbb{E}[\alpha(s_i - E_{\sigma(i-1)})^+ + \beta(E_{\sigma(i-1)} - s_i)^+]. \quad (4)$$

$(s_i - E_{\sigma(i-1)})^+$  denotes idle time before job  $i$  and  $(E_{\sigma(i-1)} - s_i)^+$  indicates its possible delay. Cost function  $c_1^\sigma$  is the same cost function used by Weiss (1990). If  $\alpha \neq \beta$ , it captures potential different costs associated with idle time and delay. We call this function  $l_1$ -type objective function.

**Example 2.** Let  $g_2(t) = t^2$ . The objective function reduces to

$$c_2^\sigma(\mathbf{s}) = \sum_{i=2}^n \mathbb{E}[(E_{\sigma(i-1)} - s_i)^2]. \quad (5)$$

Cost function  $c_2^\sigma$  penalizes both idle time and delay equally. However, due to the nonlinearity of  $c_2^\sigma$ , long idle time and delay are less tolerable. We call this function  $l_2$ -type objective function.

**Example 3.** Let

$$g_3(t) = \begin{cases} \beta(t - T_D), & \text{if } t \geq T_D \\ -\alpha(t + T_I), & \text{if } t \leq -T_I \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

where  $T_D, T_I \geq 0$  are delay and idle time tolerance, respectively (see Fig. 2). In this case, no cost is exposed for delay and idle time under a certain threshold. This situation arises in some applications such as operating room scheduling where some small amount of delay is tolerable.

## 4. Sequencing problem

### 4.1. Non-indexability

In this section, we first consider the joint sequencing-scheduling problem (referred to as just the ‘sequencing problem’ since the optimal sequence cannot be determined without also determining the optimal appointment times). Intuitively, scheduling jobs with higher uncertainty in durations first may lead to delay propagation through the schedule. Considering objective function  $c_1^\sigma$ , this intuition has motivated many researchers to prove optimality of least variance first (LVF) policy. However, the efforts have not been fruitful beyond the case of two jobs ( $n = 2$ ) for some typical distributions such as exponential and uniform. Most related papers thus have resorted to numerical evaluation to analyze the performance of the LVF rule. In particular, Denton et al. (2007) compared three ordering policies, namely, increasing mean, increasing variance, and increasing coefficient of variation. Using numerical experiment with real surgery duration data, they argued that ordering with increasing variance outperforms the other two heuristics. However, Mansourifard et al. (2018) claimed that variance does not distinguish the potential difference between idle time and delay for  $c_1^\sigma$ . They introduced the *newsvendor* index defined as

$$I_1^*(X) = \alpha \mathbb{E}[(s^* - X)^+] + \beta \mathbb{E}[(X - s^*)^+] \quad (7)$$

where  $F_X$  is cumulative distribution function of  $X$ ,  $s^* := F_X^{-1}(\frac{\beta}{\alpha+\beta})$ , and numerically verified that sequencing in increasing order of  $I_1^*$  outperforms LVF, and conjectured that it returns the optimal sequence. No proof of optimality was given. These conjectures will be evaluated in this section. In particular, we will prove that there exists no index (a map from a random variable to the reals) that yields the optimal sequence for objective functions  $c_1^\sigma$  and  $c_2^\sigma$ .

Moreover, we rigorously prove that the only candidate to provide the optimal sequence is newsvendor index for objective function  $c_1^\sigma$  and variance for objective function  $c_2^\sigma$ . This provides a theoretical support for numerical evidence of Mansourifard et al. (2018). Moreover, it completely eliminates variance as a candidate heuristic for objective function  $c_1^\sigma$ .

Let us first start with a simple example of sequencing two jobs.

**Example 4.** Consider the case of scheduling two jobs with durations  $X_1, X_2$ . The optimization problem to determine optimal appointment times given the sequence  $(X_1, X_2)$  would be:

$$\inf_{s_2 \geq 0} \mathbb{E}[g(X_1 - s_2)] \quad (8)$$

The optimal cost given by the above equation is indeed an index that maps random variable  $X_1$  to a real number. Moreover, sorting in increasing order of this index yields the optimal sequence for  $n = 2$ .

Motivated by this example, we have a candidate index for general  $n$ :

$$I_g^*(X) = \inf_{s \geq 0} \mathbb{E}[g(X - s)]. \quad (9)$$

One can verify that this index reduces to variance ( $I_2^*$ ) and newsvendor index ( $I_1^*$ ) in the case that  $g(t) = t^2$  and  $g(t) = \beta(t)^+ + \alpha(-t)^+$ , respectively. The natural question is whether this index provides the optimal sequence for  $n > 2$ . And if not, whether there is any other index that yields the optimal sequence. In the ensuing, we will show that the answer to both of these questions is negative. In fact, we first prove in Proposition 8 that  $I_g^*$  is the only possible candidate to return the optimal sequence and then through counterexamples 9 and 10 show that it is not optimal.

To prepare the setup for Proposition 8, let  $\bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$  be the extended real line. We say  $I : \mathcal{X}^+ \rightarrow \bar{\mathbb{R}}$  is an index and denote the space of all indexes by  $\mathcal{I}$ . For example, mean, variance, newsvendor and  $I_g^*$  are examples of elements in  $\mathcal{I}$ . First, we define an equivalence relation on  $\mathcal{I}$ .

**Definition 5.** Let  $I_1, I_2 \in \mathcal{I}$ . We say  $I_1$  is in relation with  $I_2$  denoting by  $I_1 R I_2$  if for any  $X_1, X_2 \in \mathcal{X}^+$ ,  $I_1(X_1) \leq I_1(X_2)$  if and only if  $I_2(X_1) \leq I_2(X_2)$ .

It is straightforward to check that  $R$  is an equivalence relation on  $\mathcal{I}$ . Hence,  $R$  splits  $\mathcal{I}$  into disjoint equivalence classes. Next, we define a notation for sorting random variables in increasing order of an index.

**Definition 6.** Let  $\mathbf{X} = (X_1, \dots, X_n)$  be a random vector where  $X_i \in \mathcal{X}^+$  for all  $i$ , and  $I \in \mathcal{I}$  be an index. We say  $\sigma \cdot \mathbf{X} = (X_{\sigma(1)}, \dots, X_{\sigma(n)})$  is a valid permutation of  $\mathbf{X}$  with respect to  $I$  if  $I(X_{\sigma(1)}) \leq \dots \leq I(X_{\sigma(n)})$ . We denote the set of valid permutations by  $\mathcal{P}_I(\mathbf{X})$ .

In the case that  $I(X_1), \dots, I(X_n)$  take distinct values,  $\mathcal{P}_I(\mathbf{X})$  includes only one element.

If  $I_1$  is equivalent to  $I_2$ , then  $\mathcal{P}_{I_1}(\mathbf{X}) = \mathcal{P}_{I_2}(\mathbf{X})$  for any random vector  $\mathbf{X}$  with components in  $\mathcal{X}^+$ .

**Definition 7.** Index  $I$  is optimal for cost function  $c$  if for any  $n \geq 2$  and any random vector  $\mathbf{X} = (X_1, \dots, X_n)$  with components in  $\mathcal{X}^+$ ,  $\inf_{\sigma \in \mathcal{P}_I(\mathbf{X})} \mathbb{E}[C(\mathbf{s}, \sigma \cdot \mathbf{X})] \leq \inf_{\sigma \in \mathcal{P}_I(\mathbf{X})} \mathbb{E}[C(\mathbf{s}, \mathbf{X})]$  for all  $\sigma \in \mathcal{P}_I(\mathbf{X})$ .

Thus, by the above remark if an index of a class is optimal, all equivalent indices are also optimal. Hence, optimality is a class property.

We already observed that  $I_g^*$  is optimal for the case of  $n = 2$ . The following Proposition provides a result for general  $n$ .

**Proposition 8.** If there exists an optimal index for cost function  $c$ , it is equivalent to  $I_g^*$ .

**Proof.** Assume by contradiction that there exists index  $J$  which is optimal but not equivalent to  $I_g^*$ . Hence, there exist random variables  $X_1, X_2 \in \mathcal{X}^+$  such that  $I_g^*(X_1) < I_g^*(X_2)$  but  $J(X_1) \geq J(X_2)$ . Note that  $I_g^*(X_1) = \inf_{s \geq 0} \mathbb{E}[g(X_1 - s)]$  and  $I_g^*(X_2) = \inf_{s \geq 0} \mathbb{E}[g(X_2 - s)]$ . Hence,  $I_g^*(X_1) < I_g^*(X_2)$  implies that  $\inf_{s \geq 0} \mathbb{E}[g(X_1 - s)] < \inf_{s \geq 0} \mathbb{E}[g(X_2 - s)]$ . However, optimality of  $J$  implies that  $\inf_{s \geq 0} \mathbb{E}[g(X_1 - s)] \geq \inf_{s \geq 0} \mathbb{E}[g(X_2 - s)]$  which is a contradiction.  $\square$

Note that  $I_g^*$  reduces to  $I_1^*$  and  $I_2^*$  for objective functions  $c_1^\sigma$  and  $c_2^\sigma$ , respectively. We also notice that contrary to widely believed conjectures in the literature that  $I_2^*$  (LVF rule) is an optimal index-based policy for cost function  $c_1^\sigma$ , Proposition 8 states that variance can only be a candidate for  $c_2^\sigma$ . However, note that this proposition does not say anything about the existence of an optimal index. In the following, we provide counter examples which show that sequencing (and optimally scheduling) in increasing order of  $I_1^*$  and  $I_2^*$  is not optimal for  $c_1^\sigma$  and  $c_2^\sigma$ , respectively.

**Example 9.** Let  $X_1, X_2, X_3$  be independent random variables in  $L^1$  and assume that  $X_1 \sim U(0, 1)$  and  $X_2$  follows the following distribution (see Fig. 3):

$$F_{X_2}(x) = \begin{cases} 0, & \text{if } x \leq 0 \\ 2x^2, & \text{if } 0 < x < 0.5 \\ 2(x - 0.5)^2 + 0.5, & \text{if } 0.5 \leq x < 1 \\ 1, & \text{otherwise} \end{cases}$$



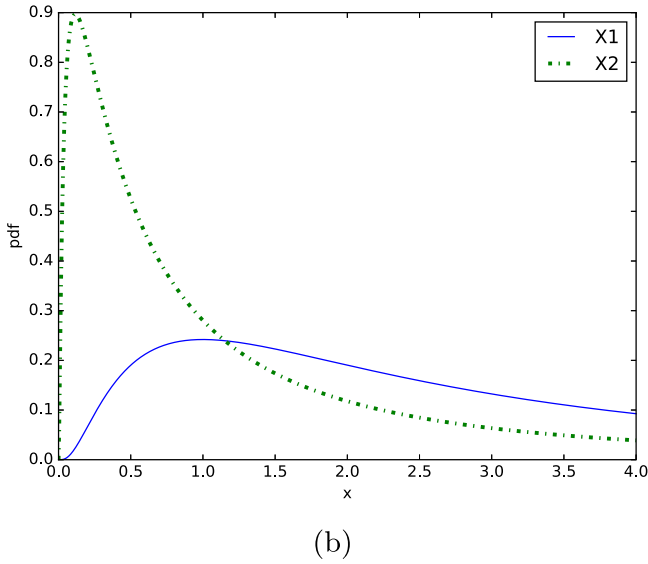
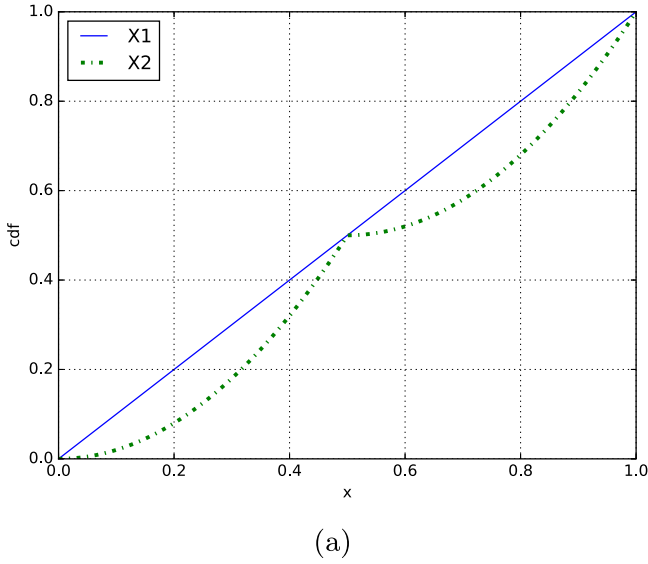


Fig. 3. Distribution of  $X_1$  and  $X_2$  for (a) Example 9 and (b) Example 10.

Consider objective function  $c_1^\sigma$  with  $\alpha = \beta = 1$ .  $I_1^*$  reduces to  $\mathbb{E}[|X - F_X^{-1}(\frac{1}{2})|]$ . First we claim that  $I_1^*(X_1) = I_1^*(X_2) = \frac{1}{4}$ :

$$\begin{aligned}
 \mathbb{E}[|X_1 - F_{X_1}^{-1}(\frac{1}{2})|] &= \mathbb{E}[|X_1 - \frac{1}{2}|] \\
 &= \int_0^{\frac{1}{2}} (\frac{1}{2} - x) dx + \int_{\frac{1}{2}}^1 (x - \frac{1}{2}) dx \\
 &= \int_0^{\frac{1}{2}} (\frac{1}{2} - x) dx + \int_0^{\frac{1}{2}} x dx = \frac{1}{4}, \\
 \mathbb{E}[|X_2 - F_{X_2}^{-1}(\frac{1}{2})|] &= \mathbb{E}[|X_2 - \frac{1}{2}|] \\
 &= \int_0^{\frac{1}{2}} (\frac{1}{2} - x) 4x dx + \int_{\frac{1}{2}}^1 (x - \frac{1}{2})(4x - 2) dx \\
 &= \int_0^{\frac{1}{2}} (\frac{1}{2} - x) 4x dx + \int_0^{\frac{1}{2}} 4x^2 dx = \frac{1}{4}.
 \end{aligned}$$

Distribution of  $X_3$  can be arbitrary as long as  $I_1^*(X_3) > \frac{1}{4}$  to make sure that it comes last. In order to have  $I_1^*$  as the optimal index, changing the order of  $X_1$  and  $X_2$  should not affect the optimal value of  $c_1^\sigma$ . However, for the sequence  $\sigma_1 \cdot \mathbf{X} = (X_1, X_2, X_3)$ ,  $\inf_{\mathbf{s} \in \mathcal{S}} c_1^{\sigma_1}(\mathbf{s}) \approx 0.3946$  but sequence  $\sigma_2 \cdot \mathbf{X} = (X_2, X_1, X_3)$  yields  $\inf_{\mathbf{s} \in \mathcal{S}} c_1^{\sigma_2}(\mathbf{s}) \approx 0.3872$ .

Thus, the index  $I_1^*$  is not optimal for cost function  $c_1^\sigma$ .

**Example 10.** Consider objective function  $c_2^\sigma$  and let  $X_1 \sim \ln \mathcal{N}(1, 1)$  and  $X_2 \sim \ln \mathcal{N}(\frac{1}{2} \ln(\frac{e}{e+1}), 2)$  be independent (see Fig. 3).

Note that  $I_2^*(X_1) = I_2^*(X_2) = e^3(e-1)$ . Distribution of  $X_3$  can be arbitrary as long as  $I_2^*(X_3) > I_2^*(X_1) = I_2^*(X_2)$  to make sure that it comes last. In order to have  $I_2^*$  as the optimal index, changing the order of  $X_1$  and  $X_2$  should not affect the optimal value of  $c_2^\sigma$ . However, for the sequence  $\sigma_1 \cdot \mathbf{X} = (X_1, X_2, X_3)$ ,  $\inf_{\mathbf{s} \in \mathcal{S}} c_2^{\sigma_1}(\mathbf{s}) \approx 94.158$  but sequence  $\sigma_2 \cdot \mathbf{X} = (X_2, X_1, X_3)$  yields  $\inf_{\mathbf{s} \in \mathcal{S}} c_2^{\sigma_2}(\mathbf{s}) \approx 99.096$ .

Thus, the index  $I_2^*$  is not optimal for cost function  $c_2^\sigma$ . The above leads us to the following conclusion.

**Theorem 11.** *There exists no index that yields the optimal sequence for cost functions  $c_1^\sigma$  and  $c_2^\sigma$ .*

**Proof.** Proposition 8 implies that  $I_k^*$  is the only possible optimal index for cost functions  $c_k$ ,  $k = 1, 2$ . But counterexamples 9 and 10 show that these need not be optimal. This leads us to the conclusion that optimal indices may not exist, i.e., the problem is non-indexable.  $\square$

**Remark 12.** It is worth mentioning that Proposition 8 still holds even if we restrict the space of random variables to a certain family. Therefore, although Theorem 11 states that the sequencing problem is not indexable in general, it does not preclude the possibility of indexability in a restricted space. Nevertheless, Proposition 8 ensures that one should not investigate indices other than  $I_g^*$ . Finding a family of distributions for which  $I_g^*$  is an optimal index is still an open research problem. In particular, Example 10 ensures that even if we restrict the space of random variables to exponential family, the problem remains non-indexable. In fact, we are unable to conclude about indexability if we further restrict to the exponential distribution. Moreover, Theorem 11 does not exclude the possibility of existence of non-index-based optimal policies.

#### 4.2. Bounds on the optimal cost

It is disappointing that contrary to long-held conjectures in the literature, the sequencing problem is non-indexable in general. Nevertheless,  $I_g^*$  can be considered as a heuristic to order the random variables and achieve a suboptimal solution. We next provide lower and upper bounds on the optimum cost with  $c_1^\sigma$  and  $c_2^\sigma$  objective functions. Note that the increasing order of  $I_k^*$ ,  $k = 1, 2$  minimizes the upper bound.

**Theorem 13.** *For  $k = 1, 2$ , the optimum cost of objective function  $c_k^\sigma$  can be bounded by:*

$$\sum_{i=1}^{n-1} I_k^*(X_{\sigma(i)}) \leq \inf_{\mathbf{s} \in \mathcal{S}} c_k^\sigma(\mathbf{s}) \leq \sum_{i=1}^{n-1} (n-i) I_k^*(X_{\sigma(i)}) \quad (10)$$

**Proof.** We need two Lemmas for the proof of the theorem. Their proofs are relegated to Appendix A. Lemma 14 proves a sub-additive property of the index functions while Lemma 15 is a technical lemma.

**Lemma 14.** Let  $X_1, X_2 \in \mathcal{X}^+$  be independent. Then, for  $k = 1, 2$

$$I_k^*(X_1 + X_2) \leq I_k^*(X_1) + I_k^*(X_2). \quad (11)$$

**Lemma 15.** Assume  $g(0) = 0$  and

- (i) let  $X \in \mathcal{X}^+$ . Then,  $\sup_{x \in \mathbb{R}} I_g^*(\max(x, X)) \leq I_g^*(X)$ .
- (ii) let  $X_1, X_2 \in \mathcal{X}^+$  be independent. Then,  $\max(I_g^*(X_1), I_g^*(X_2)) \leq I_g^*(X_1 + X_2)$ .

Lemmas 14 and 15 can now be used to bound  $I_k^*(E_{\sigma(j)})$ :

$$I_k^*(E_{\sigma(j)}) = I_k^*(\max(s_j, E_{\sigma(j-1)}) + X_{\sigma(j)}) \quad (12)$$

$$\leq I_k^*(\max(s_j, E_{\sigma(j-1)})) + I_k^*(X_{\sigma(j)}) \quad (13)$$

$$\leq I_k^*(E_{\sigma(j-1)}) + I_k^*(X_{\sigma(j)}) \quad (14)$$

for  $k = 1, 2$  where the first and the second inequality follow from Lemmas 14 and 15, respectively. Using the fact that  $E_{\sigma(1)} = X_{\sigma(1)}$ , one can write:

$$I_k^*(E_{\sigma(j)}) \leq \sum_{i=1}^j I_k^*(X_{\sigma(i)}) \quad (15)$$

By lower bound in Lemma 15,  $I_k^*(\max(s_j, E_{\sigma(j-1)}) + X_j) \geq I_k^*(X_j)$ . Hence,  $I_k^*(E_{\sigma(j)})$  can be bounded by:

$$I_k^*(X_{\sigma(j)}) \leq I_k^*(E_{\sigma(j)}) \leq \sum_{i=1}^j I_k^*(X_{\sigma(i)}). \quad (16)$$

Now, to prove the upper bound let  $\tilde{\mathbf{s}} = (\tilde{s}_2, \dots, \tilde{s}_n)$  where  $\tilde{s}_i = F_{E_{\sigma(i-1)}}^{-1}(\frac{\beta}{\alpha+\beta})$  for the case that  $k = 1$  and  $\tilde{s}_i = \mathbb{E}[E_{\sigma(i-1)}]$  for the case that  $k = 2$ . Note that  $\tilde{s}_i$  can be calculated recursively because  $E_{\sigma(i-1)}$  is a function of  $\tilde{s}_2$  through  $\tilde{s}_{i-1}$ . We have:

$$\begin{aligned} \inf_{\mathbf{s}} c_k^\sigma(\mathbf{s}) &\leq c_k^\sigma(\tilde{\mathbf{s}}) = \sum_{j=2}^n I_k^*(E_{\sigma(j-1)}) \\ &\leq \sum_{j=1}^{n-1} \sum_{i=1}^j I_k^*(X_{\sigma(i)}) \\ &= \sum_{i=1}^{n-1} \sum_{j=i}^{n-1} I_k^*(X_{\sigma(i)}) \\ &= \sum_{i=1}^{n-1} (n-i) I_k^*(X_{\sigma(i)}). \end{aligned}$$

To prove the lower bound, note that  $\mathbb{E}[g_k(E_{\sigma(i-1)} - s_i)] \geq I_k^*(E_{\sigma(i-1)}) \geq I_k^*(X_{\sigma(i-1)})$  where  $g_1(t) = \beta(t)^+ + \alpha(-t)^+$  and  $g_2(t) = t^2$ . Thus,

$$c_k^\sigma(\mathbf{s}) = \sum_{i=2}^n \mathbb{E}[g_k(E_{\sigma(i-1)} - s_i)] \geq \sum_{i=2}^n I_k^*(X_{\sigma(i-1)}). \quad \square$$

**Remark 16.** Note that the upper and lower bounds in (10) coincide when  $n = 2$ , and this is the result we already expected from Example 4. For general  $n$ , sequencing with respect to increasing order of  $I_k^*$  minimizes the upper bound in (10).

## 5. Scheduling problem

In many problems, the sequence in which to schedule is given and only the appointment times are to be determined optimally. In this section, we assume that the sequence of  $n$  random variables  $\mathbf{X} = (X_1, \dots, X_n)$  is fixed and without loss of generality (by possibly renaming jobs) remove the notation  $\sigma$  for simplicity. We call this problem the scheduling problem. We propose sample

average approximation (SAA) as an algorithm to find the optimal appointment times and prove it is statistically consistent in the case that the objective function is convex (e.g.,  $c_1^\sigma$ ). This result is significant because the only assumption required for consistency of SAA is the existence of a schedule with finite cost. This assumption significantly relaxes the typical assumptions required for consistency of SAA in the literature (see e.g., Theorem 5.4 of Shapiro, Dentcheva, & Ruszczyński, 2009).

### 5.1. Existence of solution

We first show that there exists a solution to the optimization problem in (3).

**Theorem 17.**

- (i) For any particular realization of  $\mathbf{X}$ ,  $C(\cdot, \mathbf{X})$  is nonnegative and coercive.
- (ii)  $c(\cdot)$  is nonnegative, coercive and lower semi-continuous. Furthermore, if  $c(\mathbf{s}) < \infty$  for some  $\mathbf{s} \in \mathcal{S}$ , then there exists a solution to the optimization problem in (3) and the set of minimizers is compact.

The proof is relegated to Appendix A.

One of the essential conditions in Theorem 17 is that  $c(\mathbf{s}) < \infty$  for some  $\mathbf{s} \in \mathcal{S}$ . The question is how to check whether this condition is satisfied. Should we explore the entire set  $\mathcal{S}$  in the hope of finding such  $\mathbf{s}$ ? Let us illuminate this condition: First of all it is easy to see that for  $p \geq 1$  and  $g(t) = |t|^p$ , this condition is equivalent to  $X_i \in L^p$  (i.e.,  $\mathbb{E}[|X_i|^p] < \infty$ ) for  $i = 1, \dots, n-1$ . This is also true for some other variations where  $g$  is a piecewise function of the form  $|\cdot|^p$  such as  $g_1$  and  $g_3$  in Examples 1 and 3. Moreover, if  $c(\mathbf{s}) < \infty$  for some  $\mathbf{s} \in \mathcal{S}$ , it is finite for all  $\mathbf{s} \in \mathcal{S}$ . It is mainly due to the fact that  $L^p$  is a vector space. Therefore, in such cases, there is no need to explore the set  $\mathcal{S}$ . However, for general  $g$ , the set  $\{X \in \mathcal{X}^+ \mid \mathbb{E}[g(X)] < \infty\}$  may not be a vector space (see Birnbaum–Orlicz space, Birnbaum & Orlicz, 1931) and  $c(\mathbf{s})$  may be infinite for some  $\mathbf{s}$ . In that case, random exploration may yield  $\mathbf{s} \in \mathcal{S}$  such that  $c(\mathbf{s}) < \infty$ .

### 5.2. Sample average approximation

The next question is how to calculate the optimal appointment times. Theorem 17 assures that there exists an optimal schedule under mild condition. However, calculating expectation is very costly in our problem due to the convolution nature of the distribution of the service completion times. In fact, for a given schedule  $\mathbf{s}$ , distribution of  $E_{\sigma(i)}$  is convolution of distributions of  $\max(s_i, E_{\sigma(i-1)})$  and  $X_i$ . An alternative is to use sample average approximation (SAA) to approximate the optimization problem. SAA is a well studied topic in stochastic programming (see for example, Royset, 2013; Shapiro et al., 2009). In the following, we discuss SAA and provide a theoretical guarantee for convergence of the solution in stochastic appointment scheduling problem. We assume that

**Assumption 18.** For any realization of  $\mathbf{X}$ ,  $C(\cdot, \mathbf{X})$  is convex.

This assumption holds for the  $l_1$ -type objective function (see Proposition 21 in Appendix A) which is widely considered in the literature. However, it does not hold for the case of  $l_2$ -type objective (see Example 22 in Appendix A).

Let  $(\mathbf{X}^j)_{j=1}^m$  be an independently and identically distributed (i.i.d.) random sample of size  $m$  for durations  $\mathbf{X}$  and define

$$C_m(\mathbf{s}) = \frac{1}{m} \sum_{j=1}^m C(\mathbf{s}, \mathbf{X}^j) \quad (17)$$

Instead of solving the optimization problem in Eq. (3), we are going to solve

$$\inf_{\mathbf{s} \in \mathcal{S}} C_m(\mathbf{s}). \quad (18)$$

Convexity and coercivity of  $C(\cdot, \mathbf{X})$  imply convexity and coercivity of  $C_m(\cdot)$ . Therefore, there exists a solution to the optimization problem in (18). In addition, Strong Law of Large Numbers implies that for each  $\mathbf{s}$ ,  $C_m(\mathbf{s}) \rightarrow c(\mathbf{s})$  a.s. as  $m \rightarrow \infty$ . Nevertheless, optimization over the set  $\mathcal{S}$  requires some stronger result to guarantee  $\inf_{\mathbf{s} \in \mathcal{S}} C_m(\mathbf{s}) \rightarrow \inf_{\mathbf{s} \in \mathcal{S}} c(\mathbf{s})$  a.s. as  $m \rightarrow \infty$ . Moreover, it would be useful to see if the set of minimizers of the SAA also converges to the set of true minimizers in some sense. To reach that goal, we need the following definition of deviation for sets (see Eq. (7.4) in Shapiro et al., 2009).

**Definition 19.** Let  $(M, d)$  be a metric space and  $A, B \subseteq M$ . We define distance of  $a \in A$  from  $B$  by

$$\text{dist}(a, B) := \inf\{d(a, b) \mid b \in B\} \quad (19)$$

and deviation of  $A$  from  $B$  by

$$\mathbb{D}(A, B) := \sup_{a \in A} \text{dist}(a, B). \quad (20)$$

Note that  $\mathbb{D}(A, B) = 0$  implies  $A \subseteq \text{cl}(B)$  (i.e.  $A$  is a subset of closure of  $B$  with respect to  $M$ ). The next theorem guarantees that SAA is a consistent estimator for the scheduling problem.

**Theorem 20.** Suppose Assumption 18 holds and  $c(\mathbf{s}) < \infty$  for some  $\mathbf{s} \in \mathcal{S}$  and let  $S^* = \arg\inf_{\mathbf{s} \in \mathcal{S}} c(\mathbf{s})$  and  $S_m^* = \arg\inf_{\mathbf{s} \in \mathcal{S}} C_m(\mathbf{s})$ . Then,  $\inf_{\mathbf{s} \in \mathcal{S}} C_m(\mathbf{s}) \rightarrow \inf_{\mathbf{s} \in \mathcal{S}} c(\mathbf{s})$  and  $\mathbb{D}(S_m^*, S^*) \rightarrow 0$  a.s. as  $m \rightarrow \infty$ .

The proof is available in Appendix A.

Theorem 20 proves the consistent behavior of SAA as the number of samples tends to infinity. Let us now observe how it behaves in terms of bias. For any  $\mathbf{s}' \in \mathcal{S}$ , we can write  $\inf_{\mathbf{s} \in \mathcal{S}} C_m(\mathbf{s}) \leq C_m(\mathbf{s}')$ . By taking expectation and then minimizing over  $\mathbf{s}'$ , we conclude that  $\mathbb{E}[\inf_{\mathbf{s} \in \mathcal{S}} C_m(\mathbf{s})] \leq \inf_{\mathbf{s} \in \mathcal{S}} \mathbb{E}[C_m(\mathbf{s})]$ . Since samples are i.i.d.,  $\mathbb{E}[C_m(\mathbf{s})] = c(\mathbf{s})$ . Therefore,  $\mathbb{E}[\inf_{\mathbf{s} \in \mathcal{S}} C_m(\mathbf{s})] \leq \inf_{\mathbf{s} \in \mathcal{S}} c(\mathbf{s})$  which means SAA is negatively biased. Does this bias decrease as the number of samples increases? The answer is affirmative. Theorem 2 in Mak, Morton, and Wood (1999) proves that  $\mathbb{E}[\inf_{\mathbf{s} \in \mathcal{S}} C_m(\mathbf{s})] \leq \mathbb{E}[\inf_{\mathbf{s} \in \mathcal{S}} C_{m+1}(\mathbf{s})]$ .

## 6. Numerical results

It has become a standard practice to evaluate performance on operating room data due to the immediate application of stochastic appointment scheduling in healthcare. Denton et al. (2007) used real surgery scheduling data collected at Fletcher Allen Health Care of New York. In this paper, we consider surgery scheduling dataset from Keck hospital of USC.

The dataset includes 38,000 surgeries performed in 25 operating rooms over the course of 3 years. More than 800 different procedure types were performed by 200 surgeons. Surgeries with the same procedure type performed by the same surgeon are assumed to be samples of the same distribution. Our numerical analysis is restricted to those distributions that have at least 30 samples. We stick to 30 samples because we observed that they are sufficient for a close enough SAA of the optimal solution. This is much fewer than the theoretically required number of samples given by Begen, Levi, and Queyranne (2012). In some practical scenarios, there are not enough samples to directly apply SAA. In such scenarios, similar cases based on the nature of the procedure type can be aggregated to build distributions with enough number of samples. In this paper, we focus on the surgeon-procedure pairs that have enough number of samples.

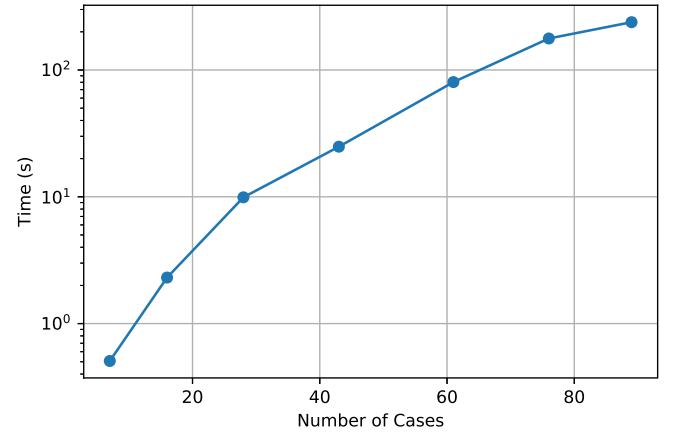


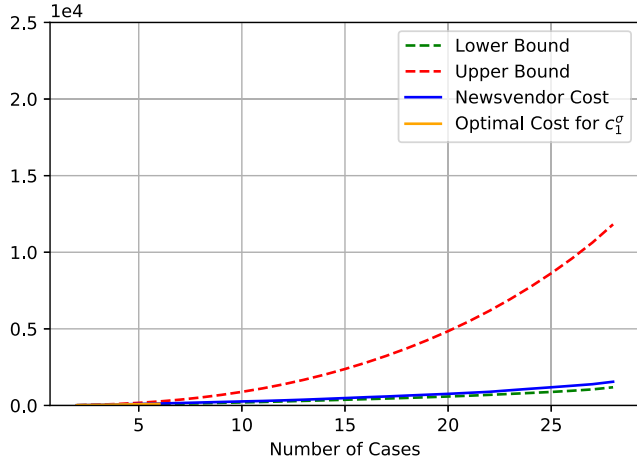
Fig. 4. SAA running time (in seconds) to find approximate optimal schedule for a given sequence. 30 samples/job are used for SAA though no appreciable difference even if 10x more samples used.

We first show that given a sequence, SAA-based optimization algorithm is fast enough for all practical purposes to find an approximate solution. To do so, we use the Powell method (Brent, 2013) to solve the SAA-based optimization problem numerically. The experiments are performed in Python on a 2015 Macbook Pro with 2.7 GHz Intel Core i5 processor and 16 GB 1867 MHz DDR3 memory. Fig. 4 confirms that appointments for a given sequence of  $n = 80$  jobs can be calculated in about 3 min. Moreover, we observed that changing the number of samples from 10 to 300 does not change the run time of the SAA-based optimization significantly.

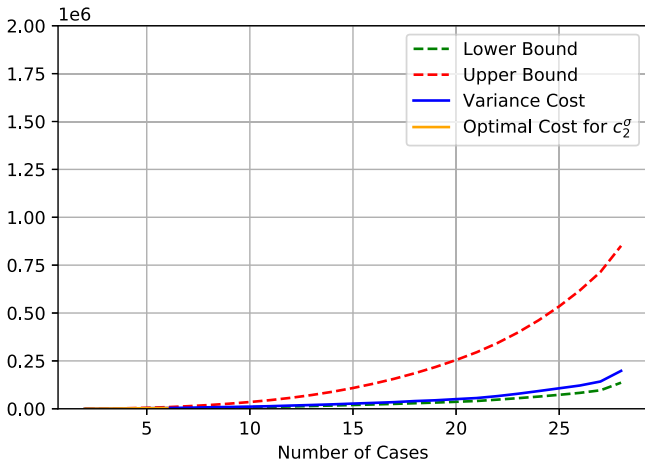
Secondly, the bounds provided in Theorem 13 are evaluated. Bounds in Theorem 13 are for general distribution and may be useful in the worst case scenarios. However, Fig. 5 shows that the upper bound is loose as the number of jobs increases on Keck dataset. The upper bound of Theorem 13 uses the complete delay propagation through the schedule, i.e., potential cost of each job affects all the future jobs equally. Although this situation might arise in the worst case, we have observed that on Keck dataset, it does not happen. Indeed, the gaps between jobs prevent the delay to have full effect on subsequent jobs.

Non-indexability shown in Theorem 11 is for general distribution. One might wonder if non-indexability is actually observed in practice. We verify that the optimal sequence is indeed different from the one given by heuristic policies (see Table 1) using Keck dataset. Newsvendor and LVF indexes are considered as heuristic policies for  $c_1^o$  and  $c_2^o$  objective functions, respectively since they are the only possible candidates to return the optimal sequence (Proposition 8). The true optimal sequence is calculated by comparing all  $n!$  choices. In operating room scheduling, the number of surgeries performed in a typical day hardly exceeds 6 which leaves the door open for exhaustive search to find the optimal sequence. However, other applications such as outpatient clinics have much larger number of jobs and it may not be feasible to exhaustively search over all possible sequences.

Cost function  $c_1^o$  depends on the idle time and delay per unit costs  $\alpha$  and  $\beta$ . Mansourifard et al. (2018) analyzed how newsvendor index outperforms variance in different regimes of these parameters. In Fig. 6, we evaluate the gap between newsvendor index and the optimal sequence as  $\alpha$  and  $\beta$  change. The optimal sequence is obtained by exhaustive search over all  $n!$  possible sequences. It can be seen that as the ratio of  $\alpha/\beta$  increases, the sub-optimality gap of newsvendor index increases on Keck dataset.

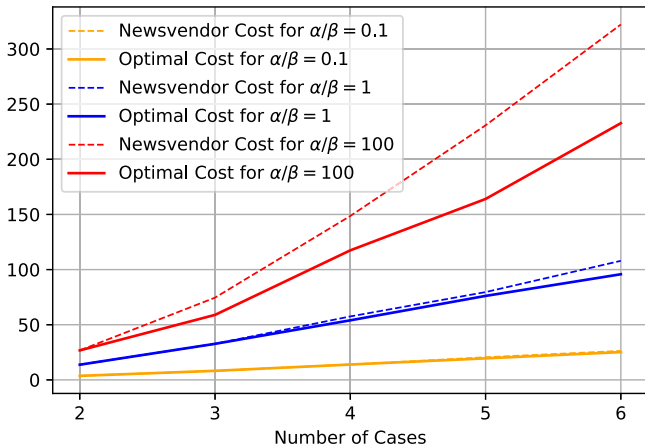


(a)



(b)

**Fig. 5.** Upper and lower bounds on optimal cost for (a)  $c_1^\sigma$  and (b)  $c_2^\sigma$  cost functions. As shown in the figure, upper bound is quite loose on USC Keck dataset. Table 1 provides numerical values for  $n \leq 6$  to compare optimal sequence with index-based heuristic policy.



**Fig. 6.** Optimality gap of newsvendor index increases as the ratio of  $\alpha/\beta$  increases on Keck dataset. The optimal sequence is found by exhaustive search over all  $n!$  possible sequences.  $\beta = 1$  is fixed and  $\alpha$  changes from 0.1 to 100. The dashed lines show the cost for the sequence obtained by least newsvendor first. The cost of the optimal sequence is shown by solid lines.

**Table 1**

Non-optimality of least newsvendor first for  $c_1^\sigma$  and least variance first for  $c_2^\sigma$ . Optimal sequence found by exhaustive search is different from the sequence given by heuristic index-based policies.

		$n$ (Number of jobs)				
		2	3	4	5	6
$c_1^\sigma$	lower bound	14.9	32.8	51.2	71.4	92.3
	optimal cost	14.9	36.6	51.6	79.0	105.3
	newsvendor cost	14.9	36.6	64.4	95.4	126.5
	upper bound	14.9	47.7	98.9	170.4	262.7
$c_2^\sigma$	lower bound	368.0	817.5	1534.3	2430.0	3562.3
	optimal cost	368.0	1036.3	1763.8	2760.1	3912.5
	variance cost	368.0	1081.7	1923.1	2853.7	3939.1
	upper bound	368.0	1185.5	2719.9	5149.8	8712.2

## 7. Conclusions

In this paper, we considered the optimal stochastic appointment scheduling problem. Each job potentially has a different service time distribution and the objective is to minimize the expectation of a function of idle time and start-time delay. There are two sub-problems. (i) The sequencing problem: the optimal sequence in which to schedule the jobs. We show that this problem in general is non-indexable. (ii) The scheduling problem: finding the optimal appointment times given a sequence or order of jobs. We show that there exists a solution to the scheduling problem. Moreover, the  $l_1$ -type objective function is convex. Further, we give a sample average approximation-based algorithm that yields an approximately optimal solution which is asymptotically consistent.

It has been an open problem for many years to find the index that yields the optimal sequence of jobs. Following the work of Weiss (1990), who showed that Least Variance First (LVF) is optimal for two cases for specific distributions, it had been conjectured that the problem is indexable and LVF may be optimal for the general problem with the  $l_1$ -type objective. In fact, several simulation studies and approximation algorithms are based on such policies. In this paper, we have settled the open question of the optimal index-type policy, namely that the problem is non-indexable in general, and no such index exists. Indeed, we show that if the problem is indexable, then a 'Newsvendor index' would be optimal for the  $l_1$  cost objective, a variance index would be optimal for  $l_2$  objective, and we also give form of an index  $I_g^*$  that would be optimal for a generalized cost function  $g$ . But we provide counterexamples that show that an optimal index-based policy does not exist for some problems. It is quite possible that the problem is indexable for specific distribution classes. That remains an open research question.

## Acknowledgements

The authors would like to thank the reviewers for great suggestions to improve the manuscript. Support from NSF SBIR I Award 1549113 is gratefully acknowledged.

## Appendix A. Proofs

**Proof of Lemma 14.** For  $k = 2$  the statement is obvious. For  $k = 1$ , using the fact that  $(a + b)^+ \leq a^+ + b^+$  for  $a, b \in \mathbb{R}$ , we have:

$$\begin{aligned}
 I_1^*(X_1 + X_2) &= \inf_s \mathbb{E}[\alpha(s - X_1 - X_2)^+ + \beta(X_1 + X_2 - s)^+] \\
 &\leq \mathbb{E}[\alpha(F_{X_1}^{-1}(\frac{\beta}{\alpha + \beta}) + F_{X_2}^{-1}(\frac{\beta}{\alpha + \beta}) - X_1 - X_2)^+]
 \end{aligned}$$



$$\begin{aligned}
& + \beta(X_1 + X_2 - F_{X_1}^{-1}(\frac{\beta}{\alpha + \beta}) - F_{X_2}^{-1}(\frac{\beta}{\alpha + \beta}))^+ ] \\
& \leq \mathbb{E}[\alpha(F_{X_1}^{-1}(\frac{\beta}{\alpha + \beta}) - X_1)^+ + \beta(X_1 - F_{X_1}^{-1}(\frac{\beta}{\alpha + \beta}))^+] \\
& + \mathbb{E}[\alpha(F_{X_2}^{-1}(\frac{\beta}{\alpha + \beta}) - X_2)^+ + \beta(X_2 - F_{X_2}^{-1}(\frac{\beta}{\alpha + \beta}))^+] \\
& = I_1^*(X_1) + I_1^*(X_2) \quad \square
\end{aligned}$$

**Proof of Lemma 15.** (i) We can write  $g(t) = g_r(t) + g_l(t)$  where  $g_r(t) = g(t^+)$  and  $g_l(t) = g(-(-t)^+)$  capture  $g$  for positive and negative values of  $t$ , respectively. Since  $g$  is nonnegative, convex and  $g(0) = 0$ , we can conclude that  $g_r$  is nondecreasing and  $g_l$  is nonincreasing. Moreover,  $I_g^*(X) = \inf_{s \geq 0} \mathbb{E}[g(X - s)] = \inf_{s \in \mathbb{R}} \mathbb{E}[g(X - s)]$ . Suppose  $s^*$  is a minimizer for  $I_g^*$  and let  $\mathcal{X} = \{x \in \mathbb{R} : x \leq s^*\}$ . We prove the lemma for  $x \in \mathcal{X}$  and  $x \notin \mathcal{X}$  separately.

Let  $x \in \mathcal{X}$ . We can write:

$$\begin{aligned}
I_g^*(\max(x, X)) &= \inf_s \mathbb{E}[g(\max(x, X) - s)] \\
&= \inf_s \mathbb{E}[g_r(\max(x, X) - s) + g_l(\max(x, X) - s)] \\
&\leq \inf_{s \geq x} \mathbb{E}[g_r(\max(x, X) - s) + g_l(\max(x, X) - s)] \\
&= \inf_{s \geq x} \mathbb{E}[g_r(X - s) + g_l(\max(x, X) - s)] \\
&\leq \inf_{s \geq x} \mathbb{E}[g_r(X - s) + g_l(X - s)] \\
&= \inf_{s \geq x} \mathbb{E}[g(X - s)] \\
&= \mathbb{E}[g(X - s^*)] = I_g^*(X)
\end{aligned}$$

For the case that  $x \notin \mathcal{X}$ , we can write:

$$\begin{aligned}
I_g^*(\max(x, X)) &= \inf_s \mathbb{E}[g(\max(x, X) - s)] \\
&= \inf_s \mathbb{E}[g_r(\max(x, X) - s) + g_l(\max(x, X) - s)] \\
&\leq \inf_{s < x} \mathbb{E}[g_r(\max(x, X) - s) + g_l(\max(x, X) - s)] \\
&= \inf_{s < x} \mathbb{E}[g_r(\max(x, X) - s)] \\
&= \mathbb{E}[g_r(\max(x, X) - x)] \\
&= \mathbb{E}[g_r(X - x)] \\
&\leq \mathbb{E}[g_r(X - s^*)] \\
&\leq \mathbb{E}[g_r(X - s^*) + g_l(X - s^*)] \\
&= \mathbb{E}[g(X - s^*)] = I_g^*(X)
\end{aligned}$$

(ii) Note that  $I_g^*(X) = \inf_{s \geq 0} \mathbb{E}[g(X - s)] = \inf_{s \in \mathbb{R}} \mathbb{E}[g(X - s)]$ . To prove  $\max(I_g^*(X_1), I_g^*(X_2)) \leq I_g^*(X_1 + X_2)$ , by symmetry, suffices to prove  $I_g^*(X_1) \leq I_g^*(X_1 + X_2)$ . Let  $x_2 \geq 0$ , we have:

$$\begin{aligned}
I_g^*(X_1) &= \inf_s \mathbb{E}[g(X_1 - s)] \\
&= \inf_s \mathbb{E}[g(X_1 + x_2 - s)] \\
&= \inf_s \mathbb{E}[g(X_1 + X_2 - s) \mid X_2 = x_2] \\
&= \inf_s \phi(s, x_2)
\end{aligned}$$

where  $\phi(s, x_2) = \mathbb{E}[g(X_1 + X_2 - s \mid X_2 = x_2)]$ . The above equality holds for any value of  $x_2 \geq 0$ . Hence,  $I_g^*(X_1) \leq \phi(s, x_2)$  for any  $s \in \mathbb{R}$ . Therefore,  $I_g^*(X_1) \leq \mathbb{E}[\phi(s, x_2)] = \mathbb{E}[g(X_1 + X_2 - s)]$  by smoothing property of conditional expectation. Thus,  $I_g^*(X_1) \leq \inf_s \mathbb{E}[g(X_1 + X_2 - s)] = I_g^*(X_1 + X_2)$ .  $\square$

**Proof of Theorem 17.** (i) Since  $g$  is nonnegative, it is obvious that  $C(\cdot, \mathbf{X})$  is also nonnegative.

To prove coercivity of  $C(\cdot, \mathbf{X})$ , let  $(\mathbf{s}^m)_{m \geq 1} \subseteq \mathbb{R}^{n-1}$  be a sequence such that  $\|\mathbf{s}^m\| \rightarrow \infty$ . We need to show that  $C(\mathbf{s}^m, \mathbf{X}) \rightarrow$

$\infty$  as  $m \rightarrow \infty$ . Let  $j$  be the smallest integer such that  $\|\mathbf{s}_j^m\| \rightarrow \infty$ . Note that for any particular realization of  $\mathbf{X}$ , there exists  $M \in \mathbb{R}$  such that  $|E_{j-1}^m| \leq M$  for all  $m$  where  $E_{j-1}^m$  denotes finish time of job  $j - 1$  with schedule  $\mathbf{s}^m$ . By triangle inequality,

$$|E_{j-1}^m - s_j^m| \geq |s_j^m| - |E_{j-1}^m| \geq |s_j^m| - M \rightarrow \infty \quad (\text{A.1})$$

as  $m \rightarrow \infty$ . Coercivity of  $g$  implies that  $g(E_{j-1}^m - s_j^m) \rightarrow \infty$  as  $m \rightarrow \infty$ . On the other hand, since  $g$  is nonnegative we can write  $C(\mathbf{s}^m, \mathbf{X}) \geq g(E_{j-1}^m - s_j^m)$  for all  $m$ . Hence,  $C(\mathbf{s}^m, \mathbf{X}) \rightarrow \infty$  as  $m \rightarrow \infty$ .

(ii) Clearly,  $c(\cdot)$  is nonnegative. To prove coercivity, let  $(\mathbf{s}^m)_{m \geq 1}$  be as defined in the previous part, by Fatou's Lemma and coercivity of  $C(\cdot, \mathbf{X})$  we have:

$$\begin{aligned}
\liminf_m c(\mathbf{s}^m) &= \liminf_m \mathbb{E}[C(\mathbf{s}^m, \mathbf{X})] \\
&\geq \mathbb{E}[\liminf_m C(\mathbf{s}^m, \mathbf{X})] = \infty.
\end{aligned}$$

To prove lower semi-continuity, let  $(\mathbf{s}^k)_{k \geq 1} \subseteq \mathbb{R}^{n-1}$  be a sequence converging to  $\mathbf{s} \in \mathbb{R}^{n-1}$ . By Fatou's Lemma we can write

$$\begin{aligned}
\liminf_k c(\mathbf{s}^k) &= \liminf_k \mathbb{E}[C(\mathbf{s}^k, \mathbf{X})] \\
&\geq \mathbb{E}[\liminf_k C(\mathbf{s}^k, \mathbf{X})] \geq \mathbb{E}[C(\mathbf{s}, \mathbf{X})] = c(\mathbf{s}).
\end{aligned}$$

Since  $c$  is coercive and  $c(\mathbf{s}) < \infty$  for some  $\mathbf{s} \in \mathcal{S}$ , without loss of generality we can assume that the minimization is over a compact set. Moreover,  $c(\mathbf{s})$  is lower semi-continuous. Thus, the set of minimizers is nonempty and compact.  $\square$

**Proof of Theorem 20.** Define the extended real valued functions

$$\begin{aligned}
\bar{C}_m(\mathbf{s}) &= C_m(\mathbf{s}) + \mathbb{I}_{\mathcal{S}}(\mathbf{s}) \\
\bar{c}(\mathbf{s}) &= c(\mathbf{s}) + \mathbb{I}_{\mathcal{S}}(\mathbf{s})
\end{aligned}$$

where

$$\mathbb{I}_{\mathcal{S}}(\mathbf{s}) = \begin{cases} 0, & \text{if } \mathbf{s} \in \mathcal{S} \\ +\infty, & \text{Otherwise} \end{cases}$$

Note that  $\bar{C}_m, \bar{c}$  are nonnegative, convex and lower semicontinuous because  $C_m, c$  are lower semicontinuous and  $\mathcal{S}$  is closed and convex. By Theorem 2.3 of [Artstein and Wets \(1994\)](#) (see [Appendix B](#)),  $\bar{C}_m(\cdot)$  epi-converges to  $\bar{c}(\cdot)$  (denoted by  $\bar{C}_m(\cdot) \xrightarrow{e} \bar{c}(\cdot)$ ) for a.e.  $\omega \in \Omega$ .

Note that  $S^* = \arg\inf_{\mathbf{s} \in \mathcal{S}} c(\mathbf{s}) = \arg\inf_{\mathbf{s} \in \mathbb{R}^{n-1}} \bar{c}(\mathbf{s})$  and  $S_m^* = \arg\inf_{\mathbf{s} \in \mathcal{S}} C_m(\mathbf{s}) = \arg\inf_{\mathbf{s} \in \mathbb{R}^{n-1}} \bar{C}_m(\mathbf{s})$ . Since  $c(\mathbf{s}) < \infty$  for some  $\mathbf{s} \in \mathcal{S}$ , by [Theorem 17](#) we know that  $S^*$  is nonempty and compact. Let  $K$  be a compact subset of  $\mathbb{R}^{n-1}$  such that  $S^*$  lies in the interior of  $K$ . Let  $\hat{S}_m^* = \arg\inf_{\mathbf{s} \in K} \bar{C}_m(\mathbf{s})$ . We first show that for a.e.  $\omega \in \Omega$ ,  $\hat{S}_m^*$  is nonempty for large enough  $m$ . Let  $\mathbf{s}^* \in S^*$  and consider  $\omega \in \Omega$  for which  $\bar{C}_m(\cdot) \xrightarrow{e} \bar{c}(\cdot)$ . By definition of epi-convergence,  $\limsup_m \bar{C}_m(\mathbf{s}_m) \leq \bar{c}(\mathbf{s}^*)$  for some  $\mathbf{s}_m \rightarrow \mathbf{s}^*$ . Therefore, there exists  $M \geq 1$  such that for  $m \geq M$ ,  $\bar{C}_m(\mathbf{s}_m) \leq \bar{c}(\mathbf{s}^*) + 1 < \infty$ . Moreover, it follows from  $\mathbf{s}_m \rightarrow \mathbf{s}^*$  that for large enough  $m$ ,  $\mathbf{s}_m$  lies in the interior of  $K$ . Since  $\bar{C}_m(\cdot)$  is convex and lower semicontinuous and  $K$  is compact,  $\hat{S}_m^*$  is nonempty a.s. (see [Appendix B](#) for Proposition 2.3.2 of [Bertsekas, Nedić, & Ozdaglar, 2003](#)).

Now, let us show that  $\mathbb{D}(\hat{S}_m^*, S^*) \rightarrow 0$  a.s. Consider  $\omega \in \Omega$  for which  $\bar{C}_m(\cdot) \xrightarrow{e} \bar{c}(\cdot)$ . We claim that for such  $\omega$ ,  $\mathbb{D}(\hat{S}_m^*, S^*) \rightarrow 0$ . Assume by contradiction that  $\mathbb{D}(\hat{S}_m^*, S^*) \not\rightarrow 0$ . Thus, there exist  $\epsilon > 0$  and  $\mathbf{y}_m \in \hat{S}_m^*$  (for large enough  $m$ ) such that  $\text{dist}(\mathbf{y}_m, S^*) \geq \epsilon$ . Let  $\mathbf{y}_{m_l} \rightarrow \mathbf{y}$  be a convergent subsequence of  $(\mathbf{y}_m)_{m \geq 1}$ . Such a subsequence exists because  $K$  is compact. It follows from  $\text{dist}(\mathbf{y}_m, S^*) \geq \epsilon$  that  $\mathbf{y} \notin S^*$ . On the other hand, Proposition 7.26 of [Shapiro et al. \(2009\)](#) (see [Appendix B](#)) implies that  $\mathbf{y} \in \arg\inf_{\mathbf{s} \in K} \bar{c}(\mathbf{s}) = S^*$  which is a contradiction.

Note that  $S^*$  is in the interior of  $K$ . It follows from  $\mathbb{D}(\hat{S}_m^*, S^*) \rightarrow 0$  that for large enough  $m$ ,  $\hat{S}_m^*$  lies in the interior of  $K$ . Hence,  $\hat{S}_m^*$  is a local minimizer. Convexity of  $\bar{C}_m(\cdot)$  implies that  $\hat{S}_m^*$  is a global minimizer i.e.  $\hat{S}_m^* = S_m^*$ . Therefore,  $\mathbb{D}(S_m^*, S^*) \rightarrow 0$  a.s. as  $m \rightarrow \infty$ .

It remains to prove that  $\inf_{s \in S} C_m(s) \rightarrow \inf_{s \in S} c(s)$  a.s. Fix  $\omega \in \Omega$  for which  $\bar{C}_m(\cdot) \xrightarrow{e} \bar{c}(\cdot)$  and let  $s_m^* \in S_m^*$  be a convergent sequence. Such a sequence exists because for large enough  $m$ ,  $S_m^*$  falls inside the compact set  $K$ . Then, by Proposition 7.26 of Shapiro et al. (2009) (see below),  $\inf_{s \in S} C_m(s) \rightarrow \inf_{s \in S} \bar{c}(s)$  or equivalently,  $\inf_{s \in S} C_m(s) \rightarrow \inf_{s \in S} c(s)$ .  $\square$

**Proposition 21.** Let  $\mathbf{X}$  be a fixed sequence of jobs and  $C_1(\cdot, \mathbf{X}) = \sum_{i=2}^n [\alpha(s_i - E_{i-1})^+ + \beta(E_{i-1} - s_i)^+]$  as in Example 1. For any realization of  $\mathbf{X}$ ,  $C_1(\cdot, \mathbf{X})$  is convex and thus,  $c_1(\cdot) = \mathbb{E}[C_1(\cdot, \mathbf{X})]$  is also convex.

**Proof.** We proceed by writing  $C_1(\cdot, \mathbf{X})$  as maximum of  $2^{n-1}$  affine functions. Since an affine function is convex, so is the maximum. To define these functions, we first split  $\mathbb{R}^{n-1}$  into  $2^{n-1}$  regions and then define an affine function in each region. These functions are then extended to the entire  $\mathbb{R}^{n-1}$ . The details are given in the following.

Fix a realization of  $\mathbf{X}$  and note that for any schedule  $\mathbf{s} = (s_2, \dots, s_n) \in \mathbb{R}^{n-1}$ , either  $(s_i - E_{i-1})^+ > 0$  or  $(E_{i-1} - s_i)^+ > 0$  for  $i = 2, \dots, n$ . Let  $b_i$  be a binary variable that indicates which of the two happens. More specifically,  $b_i = 1$  if  $s_i \geq E_{i-1}$ , and  $b_i = 0$  if  $s_i < E_{i-1}$ . These binary variables are used to split  $\mathbb{R}^{n-1}$  into  $2^{n-1}$  regions  $R_{b_2, \dots, b_n}$ . More precisely, if  $b_i = 1$ , then  $s_i \geq E_{i-1}$  denotes the range of  $s_i$  in  $R_{b_2, \dots, b_n}$  and if  $b_i = 0$ , then  $s_i < E_{i-1}$  determines its range. For example for  $n = 4$ , region  $R_{101}$  would be

$$R_{101} := \{(s_2, s_3, s_4) \in \mathbb{R}^3 \mid s_2 \geq E_1, s_3 < E_2, s_4 \geq E_3\}.$$

Corresponding to each region, one can define a function  $\bar{f}_{b_2, \dots, b_n} : R_{b_2, \dots, b_n} \rightarrow \mathbb{R}$  that consists of sum of  $n - 1$  terms associated with each  $b_i$ . If  $b_i = 1$ , then the corresponding term would be  $\alpha(s_i - E_{i-1})$  and if  $b_i = 0$ , it would be  $\beta(E_{i-1} - s_i)$ . For example for  $n = 4$ ,  $\bar{f}_{101}$  would be

$$\begin{aligned} \bar{f}_{101}(\mathbf{s}) &:= \alpha(s_2 - E_1) + \beta(E_2 - s_3) + \alpha(s_4 - E_3) \\ &= \alpha(s_2 - X_1) + \beta(s_2 + X_2 - s_3) + \alpha(s_4 - s_2 - X_2). \end{aligned} \quad (\text{A.2})$$

Note that  $C_1(\mathbf{s}, \mathbf{X}) = \bar{f}_{b_2, \dots, b_n}(\mathbf{s})$  on  $R_{b_2, \dots, b_n}$ . Moreover, restricting the domain of  $\bar{f}_{b_2, \dots, b_n}$  to  $R_{b_2, \dots, b_n}$  allowed us to write the last equality in (A.2) which can now be used for an affine extension to the entire  $\mathbb{R}^{n-1}$ . Let  $f_{b_2, \dots, b_n} : \mathbb{R}^{n-1} \rightarrow \mathbb{R}$  be such an extension. We claim that  $C_1(\mathbf{s}, \mathbf{X}) = \max_{b_2, \dots, b_n} f_{b_2, \dots, b_n}(\mathbf{s})$  for all  $\mathbf{s} \in \mathbb{R}^{n-1}$  and thus convex. To prove this claim, it suffices to show that on  $R_{b_2, \dots, b_n}$ ,  $\bar{f}_{b_2, \dots, b_n}(\mathbf{s}) \geq f_{b'_2, \dots, b'_n}(\mathbf{s})$  (because  $C_1(\mathbf{s}, \mathbf{X}) = \bar{f}_{b_2, \dots, b_n}(\mathbf{s})$  on  $R_{b_2, \dots, b_n}$ ). This is indeed true because if  $b'_i \neq b_i$ , the corresponding term would be negative in  $f_{b'_2, \dots, b'_n}(\mathbf{s})$ . Finally, note that  $c_1(\cdot) = \mathbb{E}[C_1(\cdot, \mathbf{X})]$  is also convex since expectation preserves convexity.  $\square$

Proposition 21 shows that the  $I_1$ -type objective function is convex. The following example shows that this may not be true for the objective function  $c_2$ .

**Example 22.** Consider the special case of  $n = 3$  and let  $X_1, X_2 > 0$  be positive scalars (which can be seen as degenerate distributions). We show that the function  $c_2(s_2, s_3) := (X_1 - s_2)^2 + (\max\{X_1, s_2\} + X_2 - s_3)^2$  is not convex. Let  $t = 0.5$  and  $\mathbf{s}^1 = (X_1 - \gamma, X_1 + 10X_2)$ ,  $\mathbf{s}^2 = (X_1, X_1 + 10X_2)$  and  $\mathbf{s}^3 = (X_1 + \gamma, X_1 + 10X_2)$  for some  $0 < \gamma < \min\{X_1, 6X_2\}$ . Substituting these values, we observe that  $c_2(\mathbf{s}^2) = c_2(t\mathbf{s}^1 + (1-t)\mathbf{s}^3) > tc_2(\mathbf{s}^1) + (1-t)c_2(\mathbf{s}^3)$ .

## Appendix B. Useful theorems and propositions

**Theorem 23** (Theorem 2.3 of Artstein & Wets, 1994). Let  $F : S \times \mathcal{E} \rightarrow (-\infty, \infty]$  be a measurable function and  $P(d\xi)$  be a probability measure over the space  $\mathcal{E}$  of random elements. We assume that  $S$  is a metric space. Define  $f(s) := \mathbb{E}[F(s, \xi)] = \int F(s, \xi)P(d\xi)$  and let  $\xi_1, \dots, \xi_m$  be independent samples of  $\mathcal{E}$  drawn according to  $P$ . Suppose (1)  $F(\cdot, \xi)$  is lower semicontinuous for fixed  $\xi \in \mathcal{E}$  and (2) for each  $s_0 \in S$  there exist an open set  $N_0 \subseteq S$  and an integrable function  $g_0 : \mathcal{E} \rightarrow (-\infty, \infty)$  such that the inequality

$$F(s, \xi) \geq g_0(\xi)$$

holds for all  $s \in N_0$ . Then,  $\frac{1}{m} \sum_{j=1}^m F(\cdot, \xi_j)$  almost surely epi-converges to  $f(\cdot)$ .

**Proposition 24** (Proposition 7.26 of Shapiro et al., 2009). Let  $f_m, f : S \rightarrow (-\infty, \infty]$  where  $S \subseteq \mathbb{R}^n$ . Suppose that  $f_m(\cdot)$  epi-converges to  $f(\cdot)$ . Then,

$$\limsup_m [\inf_s f_m(s)] \leq \inf_s f(s).$$

Suppose further that (1) for some  $\epsilon_m \downarrow 0$  there exists an  $\epsilon_m$ -minimizer  $s_m$  of  $f_m(\cdot)$  such that the sequence  $s_m$  converges to a point  $\bar{s}$ . Then,  $\bar{s} \in \text{argmin} f$  and

$$\lim_{m \rightarrow \infty} [\inf_s f_m(s)] = \inf_s f(s)$$

**Proposition 25** (Proposition 2.3.2 of Bertsekas et al., 2003). Let  $S$  be a closed convex subset of  $\mathbb{R}^n$ , and let  $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$  be a closed convex function such that  $f(s) < \infty$  for some  $s \in S$ . The set of minimizing points of  $f$  over  $S$  is nonempty and compact if and only if  $S$  and  $f$  have no common nonzero direction of recession.

## References

- Ahmadi-Javid, Amir, Jalali, Zahra, & Klassen, Kenneth J. (2017). Outpatient appointment systems in healthcare: A review of optimization studies. *European Journal of Operational Research*, 258(1), 3–34.
- Artstein, Zvi, & Wets, Roger J. -B (1994). Consistency of minimizers and the SLLN for stochastic programs. IBM Thomas J. Watson Research Division.
- Bailey, Norman T. J. (1952). A study of queues and appointment systems in hospital out-patient departments, with special reference to waiting-times. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 185–199.
- Baker, Kenneth R. (1974). *Introduction to sequencing and scheduling*. John Wiley & Sons.
- Baker, Kenneth R. (2014). Minimizing earliness and tardiness costs in stochastic scheduling. *European Journal of Operational Research*, 236(2), 445–452.
- Begen, Mehmet A., Levi, Retsef, & Queyranne, Maurice (2012). A sampling-based approach to appointment scheduling. *Operations Research*, 60(3), 675–681.
- Begen, Mehmet A., & Queyranne, Maurice (2011). Appointment scheduling with discrete random durations. *Mathematics of Operations Research*, 36(2), 240–257.
- Berg, Bjorn P., Denton, Brian T., Erdogan, S. Ayca, Rohleder, Thomas, & Huschka, Todd (2014). Optimal booking and scheduling in outpatient procedure centers. *Computers & Operations Research*, 50, 24–37.
- Bertsekas, D. P., Nedić, A., & Ozdaglar, A. E. (2003). *Athena scientific optimization and computation series, Convex analysis and optimization*. Athena Scientific, <https://books.google.com/books?id=DaOFQgAACAAJ>.
- Birnbaum, Z., & Orlicz, W. -f. (1931). Über die verallgemeinerung des begriffes der zueinander konjugierten potenzien. *Studia Mathematica*, 3(1), 1–67.
- Brent, Richard P. (2013). *Algorithms for minimization without derivatives*. Courier Corporation.
- Cardoen, Brecht, Demeulemeester, Erik, & Beliën, Jeroen (2010). Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, 201(3), 921–932.
- Cayirli, Tugba, & Veral, Emre (2003). Outpatient scheduling in health care: a review of literature. *Production and Operations Management*, 12(4), 519–549.
- Chen, Rachel R., & Robinson, Lawrence W. (2014). Sequencing and scheduling appointments with potential call-in patients. *Production and Operations Management*, 23(9), 1522–1538.

- Choi, Sangdo Sam, & Banerjee, Amarnath Andy (2016). Comparison of a branch-and-bound heuristic, a newsvendor-based heuristic and periodic bailey rules for outpatients appointment scheduling systems. *The Journal of the Operational Research Society*, 67(4), 576–592.
- Conway, Richard Walter, Maxwell, William L., & Miller, Louis W. (2003). *Theory of scheduling*. Courier Corporation.
- Denton, Brian, & Gupta, Diwakar (2003). A sequential bounding approach for optimal appointment scheduling. *IIE Transactions*, 35(11), 1003–1016.
- Denton, Brian, Viapiano, James, & Vogl, Andrea (2007). Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science*, 10(1), 13–24.
- Erdogan, S. Ayca, & Denton, Brian (2013). Dynamic appointment scheduling of a stochastic server with uncertain demand. *INFORMS Journal on Computing*, 25(1), 116–132.
- Erdogan, S. Ayca, Gose, Alexander, & Denton, Brian T. (2015). Online appointment sequencing and scheduling. *IIE Transactions*, 47(11), 1267–1286.
- Guda, Harish, Dawande, Milind, Janakiraman, Ganesh, & Jung, Kyung Sung (2016). Optimal policy for a stochastic scheduling problem with applications to surgical scheduling. *Production and Operations Management*, 25(7), 1194–1202.
- Gupta, Diwakar (2007). Surgical suites' operations management. *Production and Operations Management*, 16(6), 689–700.
- Gupta, Diwakar, & Denton, Brian (2008). Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions*, 40(9), 800–819.
- Herroelen, Willy, & Leus, Roel (2004). The construction of stable project baseline schedules. *European Journal of Operational Research*, 156(3), 550–565.
- Hulshof, Peter J. H., Kortbeek, Nikky, Boucherie, Richard J., Hans, Erwin W., & Bakker, Piet J. M. (2012). Taxonomic classification of planning decisions in health care: a structured review of the state of the art in OR/MS. *Health Systems*, 1(2), 129–175.
- Kaandorp, Guido C., & Koole, Ger (2007). Optimal outpatient appointment scheduling. *Health Care Management Science*, 10(3), 217–229.
- Kemper, Benjamin, Klaassen, Chris A. J., & Mandjes, Michel (2014). Optimized appointment scheduling. *European Journal of Operational Research*, 239(1), 243–255.
- Klassen, Kenneth J., & Rohleder, Thomas R. (1996). Scheduling outpatient appointments in a dynamic environment. *Journal of Operations Management*, 14(2), 83–101.
- Kong, Qingxia, Lee, Chung-Yee, Teo, Chung-Piaw, & Zheng, Zhichao (2013). Scheduling arrivals to a stochastic service delivery system using copositive cones. *Operations Research*, 61(3), 711–726.
- Kong, Qingxia, Lee, Chung-Yee, Teo, Chung-Piaw, & Zheng, Zhichao (2016). Appointment sequencing: Why the smallest-variance-first rule may not be optimal. *European Journal of Operational Research*, 255(3), 809–821.
- Kuiper, Alex, & Mandjes, Michel (2015a). Appointment scheduling in tandem-type service systems. *Omega*, 57, 145–156.
- Kuiper, Alex, & Mandjes, Michel (2015b). Practical principles in appointment scheduling. *Quality and Reliability Engineering International*, 31(7), 1127–1135.
- Kuiper, Alex, Mandjes, Michel, & de Mast, Jeroen (2017). Optimal stationary appointment schedules. *Operations Research Letters*, 45(6), 549–555.
- Lebowitz, Philip (2003). Schedule the short procedure first to improve OR efficiency. *AORN Journal*, 78(4), 651–659.
- Mak, Wai-Kei, Morton, David P., & Wood, R. Kevin (1999). Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24(1), 47–56.
- Mak, Ho-Yin, Rong, Ying, & Zhang, Jiawei (2014). Appointment scheduling with limited distributional information. *Management Science*, 61(2), 316–334.
- Mancilla, Camilo, & Storer, Robert (2012). A sample average approximation approach to stochastic appointment sequencing and scheduling. *IIE Transactions*, 44(8), 655–670.
- Mansourifard, Farzaneh, Mansourifard, Parisa, Ziyadi, Morteza, & Krishnamachari, Bhaskar (2018). A heuristic policy for outpatient surgery appointment sequencing: newsvendor ordering. In *2nd IEOM European conference on industrial engineering and operations management*. Paris.
- Marcon, Eric, & Dexter, Franklin (2006). Impact of surgical sequencing on post anesthesia care unit staffing. *Health Care Management Science*, 9(1), 87–98.
- Newbold, Robert C. (1998). *Project management in the fast lane: applying the theory of constraints*. CRC Press.
- Pinedo, Michael (2012). *Scheduling*. Springer.
- Qi, Jin (2016). Mitigating delays and unfairness in appointment systems. *Management Science*, 63(2), 566–583.
- Royset, Johannes O. (2013). On sample size control in sample average approximations for solving smooth stochastic programs. *Computational Optimization and Applications*, 55(2), 265–309.
- Shapiro, Alexander, Dentcheva, Darinka, & Ruszczyński, Andrzej (2009). Lectures on stochastic programming. *MPS/SIAM Series on Optimization*, 9, 1.
- Soriano, Alfonso (1966). Comparison of two scheduling systems. *Operations Research*, 14(3), 388–397.
- Tian, Wendi, & Demeulemeester, Erik (2014). Railway scheduling reduces the expected project makespan over roadrunner scheduling in a multi-mode project scheduling environment. *Annals of Operations Research*, 213(1), 271–291.
- Vink, Wouter, Kuiper, Alex, Kemper, Benjamin, & Bhulai, Sandjai (2015). Optimal appointment scheduling in continuous time: The lag order approximation method. *European Journal of Operational Research*, 240(1), 213–219.
- Wang, P. Patrick (1993). Static and dynamic scheduling of customer arrivals to a single-server system. *Naval Research Logistics*, 40(3), 345–360.
- Wang, P. Patrick (1999). Sequencing and scheduling n customers for a stochastic server. *European Journal of Operational Research*, 119(3), 729–738.
- Weiss, Elliott N. (1990). Models for determining estimated start times and case orderings in hospital operating rooms. *IIE Transactions*, 22(2), 143–150.
- Wu, Weitiao, Liu, Ronghui, Jin, Wenzhou, & Ma, Changxi (2019). Stochastic bus schedule coordination considering demand assignment and rerouting of passengers. *Transportation Research, Part B (Methodological)*, 121, 275–303.



**Mehdi Jafarnia-Jahromi** is a PhD candidate in Electrical Engineering Department at University of Southern California (USC), Los Angeles, CA, USA. He received his MS in Computer Science, MA in Applied Mathematics, both from USC. Prior to that, he graduated from Sharif University of Technology with a dual BS in Electrical Engineering and Computer Science. His research interests include reinforcement learning, stochastic control, machine learning, and online learning.



**Rahul Jain** received a B.Tech. degree from IIT Kanpur, Kanpur, India, an MS degree in ECE from Rice University, an M.A. degree in statistics and the PhD degree in EECS from the University of California, Berkeley, Berkeley, CA. He is the K.C. Dahlberg Early Career Chair and Associate Professor of ECE, CS\* and ISE\* (\*by courtesy) with the University of Southern California, Los Angeles, CA, USA. His research interests include reinforcement learning, stochastic control, statistical learning, stochastic networks, and game theory, and power systems, transportation, and healthcare on the applications side. Dr. Jain has received numerous awards including the NSF CAREER award, the ONR Young Investigator award, an IBM Faculty award, the James H. Zumberge Faculty Research and Innovation Award, and is currently a US Fulbright Scholar. He is an associate editor for IEEE Transactions on Automatic Control, IEEE Transaction on the Control of Networked Systems and Queueing Systems: Theory and Applications, and also serves on the IEEE Control Systems Society's Conference Editorial Board.