Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

STOCHASTIC APPROXIMATION FOR SIMULATION OPTIMIZATION UNDER INPUT UNCERTAINTY WITH STREAMING DATA

Eunhye Song
Uday V. Shanbhag

Department of Industrial and Manufacturing Engineering
Pennsylvania State University
Leonhard Building
University Park, PA 16802, USA

ABSTRACT

We consider a simulation optimization problem whose objective function is defined as the expectation of a simulation output based on a continuous decision variable, where the parameters of the simulation input distributions are estimated based on independent and identically distributed streaming data from a real-world system. Finite-sample error in the input parameter estimates causes input uncertainty in the simulation output, which decreases as the data size increases. By viewing the problem through the lens of misspecified stochastic optimization, we develop a stochastic approximation (SA) framework to solve a sequence of problems defined by the sequence of input parameter estimates to increasing levels of exactness. Under suitable assumptions, we observe that the error in the SA solution diminishes to zero in expectation and propose a SA sampling scheme so that the resulting solution iterates converge to the optimal solution under the real-world input distribution at the best possible rate.

1 INTRODUCTION

Simulation optimization (SO) is widely employed for the resolution of decision-making problems that arise in a complex real-world stochastic systems when the system behavior can be implemented into a simulator. However, an analytical expression of the performance measure is often difficult to obtain. This randomness in the real-world system is modeled by input probability distributions in the simulator, which generates random variates as inputs and the resulting simulation output becomes a function of the inputs. When data collected from the real-world system is available, these input models may be statistically estimated from real-world observations. Thus, given any finite number of observations, the input models are afflicted with error, which becomes the source of input uncertainty in the simulation output.

Accounting for input uncertainty in SO introduces an additional challenge because the estimation error in the input models renders the SO problem *misspecified*. Consequently, even if we were to optimally solve a problem parametrized by an error corrupted input model, the resulting solution is subject to the risk of being suboptimal when implemented in the real-world system; Song and Nelson (2019) refer to this as input model risk.

Until recently, much of SO literature has focused on the setting where real-world input data are limited while additional data collection is either infeasible or expensive. In this context, computing the real-world optimum with high statistical confidence is challenging as shown by Corlu and Biller (2013) and Song et al. (2015) in the context of subset selection and ranking and selection (R&S), respectively. Alternatively, Fan et al. (2013) and Gao et al. (2017) study the distributionally robust R&S paradigm when the uncertainty about the input models can be represented by an ambiguity set of candidate input distributions. Because this formulation does not require the real-world input distribution to be known, the robust optimal solution can be found with a stringent statistical guarantee; however, it often tends to be too conservative when implemented under the real-world randomness.

We focus on a different problem where additional input data are indeed available. More specifically, we assume that the real-world input process generates i.i.d. observations over time, which we refer to as *streaming data*. In this setting, the estimation error of the input models and thus input uncertainty in simulation output decreases over time. Also, we may define a sequence of SO problems parameterized by the sequence of estimated input models. Then, the question is how to efficiently solve this sequence of problems such that the resulting sequence of solutions returned by the algorithm converges to the optimal solution under the real-world input distributions at the fastest rate. Concentrating on SO problems with continuous decision variables, we design a stochastic approximation (SA) scheme to study this problem. To the best of our knowledge, Wu and Zhou (2019) is the only other work in SO literature that studies the impact of input uncertainty to SO with possibility of additional data collection. However, their focus is on a discrete solution space. For the present, we assume that the system behavior is captured by a possibly expectation-valued function $\mathbb{E}[F(x,\theta^*,\omega)]$ and θ^* denotes the parametrization of the underlying distribution, which is available only in the limit.

Similar problems have been studied in the optimization literature under the broad umbrella of *misspecified optimization problems*, where the problem of interest is parametrized by a vector that can only be learned through data, often only available in a streaming form. For instance, the objective of the optimization problem may be parametrized by $\theta^* \in \mathbb{R}^m$ where θ^* is only available in the limit through a learning procedure. Instances of such problems arise in a range of settings, including portfolio optimization which relies on access to the covariance matrices and mean vectors, production management problems that rely on a priori knowledge of machine efficiencies, etc. Crucial to this model is the requirement that the learning of θ^* is unaffected by the optimization of $\mathbb{E}[F(x,\theta^*,\omega)]$. Research on misspecified optimization and game-theoretic problems appears to have originated with the research by Szidarovszky (2004), Bischi et al. (2007), and Bischi et al. (2008). More recently, there has been an effort to addressing both deterministic problems by Ahmadi and Shanbhag (2014) while stochastic problems have been examined by Jiang and Shanbhag (2016) and Jiang and Shanbhag (2013). In fact, similar questions have also been examined in the context of misspecified Nash games (see Jiang et al. (2018)) and misspecified Markov Decision Processes (MDPs) (see Jiang and Shanbhag (2015)). In particular, Jiang and Shanbhag (2016) considered a framework where the learning problem is cast as a stochastic convex optimization problem while the misspecified optimization problem is a parametrized stochastic convex optimization problem. By taking a single stochastic gradient step in the optimization space (i.e. x) and in the learning space (i.e. θ), the sequence of iterates $\{x_k\}$ converges to a minimizer of $\mathbb{E}[F(x,\theta^*,\omega)]$ in an almost sure sense. Furthermore, under some conditions, there is no decay in the rate of convergence.

In this paper, we revisit the far broader question of contending with input model uncertainty in the context of simulation optimization by viewing the problem through the lens of misspecified optimization. In contrast with the framework presented by Jiang and Shanbhag (2016), given a sequence $\{\theta_k\} \to \theta^*$, each step in the proposed optimization process computes an inexact solution x_k of $\mathbb{E}[F(x,\theta_k,\omega)]$. This requires a sequence of gradient steps in x and we show that under suitable requirements, the overall sub-optimality error diminishes at the Monte-Carlo rate of $\mathscr{O}(1/\sqrt{N_k})$, which governs the convergence rate of $\{\theta_k\}$.

The remainder of the paper is organized as follows: Section 2 provides a mathematical formulation of the SO problem under input uncertainty when streaming data are available. Section 3 introduces our stochastic approximation (SA) scheme and analyze its convergence properties. Preliminary numerics are provided in Section 4 and we conclude in Section 5.

2 PROBLEM STATEMENT

Suppose we observe i.i.d. streaming data from a real-world input distribution $G(\theta^*)$. We assume that the distribution family G is known while the true value of θ^* is unknown and may be estimated from the data via maximum likelihood estimation (MLE). Let us denote the sequences of incremental and cumulative number of real-world observations by $\{n_k\}_{k\geq 1}$ and $\{N_k\}_{k\geq 1}$, respectively. In other words, $N_k = \sum_{i=1}^{k} n_i$ while at any epoch k, the vector θ_k is estimated from N_k observations, where it is assumed that $\theta_k \in \Theta$, a compact set in $\mathbb{R}^d$. Additionally, we assume $\{n_k\}_{k\geq 1}$ is determined by the data streaming process which we cannot control. Under some regularity conditions for $G(\theta^*)$ (Newey and McFadden 1994),

$$\theta_k \xrightarrow[a.s]{k\to\infty} \theta^* \quad \text{and} \quad \mathbb{E}[\|\theta_k - \theta^*\|] = \mathscr{O}(N_k^{-1/2}). \tag{1}$$

Our objective is to solve the following SO problem using a SA scheme:

$$\text{Opt}(\theta^*): \quad \min_{x\in X} \quad f(x,\theta^*) \triangleq \mathbb{E}[F(x,\theta^*,\xi(\omega))],$$

where $X \subseteq \mathbb{R}^n$ is a closed, convex, and bounded set, $F : X \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}$ is a real-valued function that generates the simulation output, $\xi : \Omega \to \mathbb{R}^d$ denotes a random variate generated from estimated input distribution $G(\theta_k)$. Throughout, we assume the associated probability space is denoted by $(\Omega, \mathscr{F}, \mathbb{P})$ while $\mathbb{E}[\cdot]$ denotes the expectation operator. We denote the optimal solution to $\text{Opt}(\theta^*)$ by x^*.

At any finite epoch k, $\text{Opt}(\theta^*)$ is not available, but its plug-in version $\text{Opt}(\theta_k)$ is indeed available. We denote the optimal solution to the latter by x_k^*. In general, $x_k^* \neq x^*$ unless $f(\cdot,\cdot)$ has a very simple structure. When SA is applied to solve $\text{Opt}(\theta_k)$, one can only find its approximate solution $x_k (\neq x_k^*)$ in a finite number of SA iterations. As the data streaming process continues, an important question is how much SA effort to expend on solving each $\text{Opt}(\theta_k)$ so that the resulting sequence of solutions, denoted by $\{x_k\}$, converges to x^* at the best rate. A naive approach to tackle this problem is to solve $\text{Opt}(\theta_K)$ for some large K to obtain an approximate solution x_K. Unfortunately, such an approach provides, at best, an approximate x_K^*. An alternative is to restart the algorithm at each k and utilize a large number of SA steps to approximate x_k^* with high accuracy. This approach is wasteful since each SA step requires running (potentially computationally expensive) simulations to estimate the gradient of $f(x,\theta_k)$ as discussed in Section 3.1.

We present an efficient SA scheme which parsimoniously allocates simulation effort at each k while still achieving the best convergence rate of x_k. Crucial to our framework is the requirement that the data generating process is unaffected by the optimization of x (which is clearly reliant on θ). Before proceeding, we discuss some classical instances where such problems arise.

(I) *Queueing systems.* Designing a queueing system that balances service quality and operation costs arises frequently in operations research. Consider an optimization problem that requires minimizing the overall expected cost while controlling the service rate of the system. Faster service requires higher operational cost, whereas slower service leads to longer waiting time and may ultimately cause customers to abandon the queue, incurring lost sales cost. The optimal service rate balances both costs and depends on the arrival rate θ^* of the customers, which needs to be estimated by observing real-world arrival process. We provide an example of a simplified M/M/1/c queueing system in Section 4 as a demonstration.

(II) *Misspecified portfolio optimization.* Consider the problem of choosing a portfolio of assets from a set indexed by $i \in \{1,\ldots,n\}$ while minimizing a suitably defined mean-risk criterion. Specifically, if μ^* denotes the mean return while Σ^* denotes the covariance matrix, then the misspecified portfolio optimization problem can be cast as follows.

$$\begin{aligned} \min_{x\geq 0} \quad & x^T\mu^* + \frac{1}{2}x^T\Sigma^* x \\ \text{subject to} \quad & \sum_{i=1}^{n} x_i \leq 1. \end{aligned}$$

This can be viewed as a special case of $(\text{Opt}(\theta^*))$ where $f(x,\theta^*) \triangleq x^T\mu^* + \frac{1}{2}x^T\Sigma^* x$ and (μ^*,Σ^*) requires solving a stochastic learning problem.

3 ALGORITHM STATEMENT AND ANALYSIS

In this section, we formally describe our SA framework in Section 3.1 and analyze its performance in Section 3.2.

3.1 Algorithm Statement

Consider the following SA framework for $k \geq 1$ starting with initial point $x_0 \in \mathbb{R}^n$. At step k, given an estimate θ_k, we consider taking M_k stochastic gradient steps in x. Note that this in sharp contrast with Jiang and Shanbhag (2016) where a single stochastic gradient step was taken both in x and θ. If $\Pi_X(u)$ denotes the Euclidean projection of u onto X, then the sequences $\{x_k\}$ and $\{x_{k,j}\}$ are generated as follows:

$$\begin{aligned} x_{k,1} &:= x_{k-1} \\ x_{k,j+1} &:= \Pi_X\left[x_{k,j} - \gamma_{k,j}\nabla_x F(x_{k,j},\theta_k,\omega_{j,k})\right], \qquad j = 1,\ldots,M_k - 1, \\ x_k &:= \Pi_X\left[x_{k,M_k} - \gamma_{k,M_k}\nabla_x F(x_{k,M_k},\theta_k,\omega_{M_k,k})\right]. \end{aligned} \tag{2}$$

Here, $\nabla_x F(x,\theta,\omega)$ denotes the estimator of $\nabla_x f(x,\theta)$ obtained from simulations while $\gamma_{k,j}$ denotes the SA step size for the kth problem at the jth inner step. The "cost" of a SA step captures the cost of computing $\nabla_x F(x,\theta,\omega)$. Notice that we "warm-start" solving the kth problem by taking x_{k-1} as the initial solution.

3.2 Analysis

Throughout the remainder of this paper, we make the following assumption on the function f.

Assumption 1 (Properties of f, X, and Θ) Suppose X is a closed, convex, nonempty, and bounded set in $\mathbb{R}^n$ and Θ is a compact set in $\mathbb{R}^d$. Suppose $f(x,\theta)$ is a μ-strongly convex and continuously differentiable function in x for every $\theta \in \Theta$. Furthermore, $\|\nabla_x f(x,\theta) - \nabla_x f(y,\theta)\| \leq L\|x-y\|$ for all $\theta \in \Theta$. In addition, we assume that $\|\nabla_x f(x,\theta)\| \leq C_x$ and $\|\nabla_\theta f(x,\theta)\| \leq C_\theta$ for all $x \in X$ and $\theta \in \Theta$.

We assume the existence of an oracle that can produce unbiased (parametrized) gradients with suitable properties.

Assumption 2 Given a θ, there exists a first-order oracle that produces a sampled gradient $\nabla_x F(x,\theta,\omega)$. If $w = \nabla_x F(x,\theta,\omega) - \nabla_x f(x,\theta)$, then $\mathbb{E}[w] = 0$ and $\mathbb{E}[\|w\|^2] \leq \nu^2$ for all x,θ. Moreover, $\mathbb{E}[\|\nabla_x F(x,\theta,\omega)\|^2] \leq C^2$.

The following lemma shows convergence in mean when the step size for the SA scheme is set as $\gamma_{k,j} = \frac{1}{\mu j}$ for all $k \geq 1$.

Lemma 1 Suppose Assumptions 1 and 2 hold. Consider the sequences $\{x_k\}$ and $\{\theta_k\}$ generated by (2), given x_0. Then for any k, we have that the following holds.

$$\mathbb{E}[f(x_k,\theta^*) - f(x^*,\theta^*)] \leq \left(2C_\theta + \tfrac{C_x L}{\mu}\right)\mathbb{E}[\|\theta_k - \theta^*\|] + \frac{L\max\{\frac{C^2}{\mu^2}, \mathbb{E}[\|x_{k-1} - x_k^*\|^2]\}}{M_k}. \tag{3}$$

Proof. We begin by casting the sub-optimality gap in $f(x,\theta^*)$ as the sum of the following four terms.

$$\begin{aligned} |f(x_k,\theta^*) - f(x^*,\theta^*)| \leq{} & \underbrace{|f(x_k,\theta^*) - f(x_k,\theta_k)|}_{\text{Term a}} + \underbrace{|f(x_k,\theta_k) - f(x_k^*,\theta_k)|}_{\text{Term b}} \\ & + \underbrace{|f(x_k^*,\theta_k) - f(x^*,\theta_k)|}_{\text{Term c}} + \underbrace{|f(x^*,\theta_k) - f(x^*,\theta^*)|}_{\text{Term d}}. \end{aligned} \tag{4}$$

Consider Term a and let $\tilde{\theta} \in [\theta_k, \theta^*]$. Since $f(x,\theta)$ is differentiable in θ at any x, we have the following series of equalities and inequalities by invoking the mean-value theorem and Cauchy-Schwarz inequality for $\tilde{\theta} \in [\theta_k, \theta^*]$.

$$\begin{aligned}\text{Term a} \triangleq \mid f(x_k,\theta^*) - f(x_k,\theta_k) \mid &= \mid \nabla_\theta f(x_k,\tilde{\theta})^T(\theta_k - \theta^*) \mid \\ &\leq \|\nabla_\theta f(x_k,\tilde{\theta})\| \|(\theta_k - \theta^*)\| \leq C_\theta \|\theta_k - \theta^*\|.\end{aligned}$$

Term d can be similarly bounded as shown next.

$$\text{Term d} \triangleq |f(x^*,\theta_k) - f(x^*,\theta^*)| \leq C_\theta \|\theta_k - \theta^*\|.$$

Next, we examine term b. Recall that a stochastic gradient scheme with M_k stochastic gradient steps is employed for computing x_k, which is an approximation of x_k^* and denotes the unique minimizer of $f(x,\theta_k)$. Since $f(x,\theta)$ is μ-strongly convex in x, $L-$smooth for every θ with $\|\nabla_x f(x,\theta)\| \leq C_x$ for all x,θ, it follows from the rate statement for stochastic approximation schemes for strongly convex problems that the following holds for any k.

$$\mathbb{E}[f(x_k,\theta_k) - f(x_k^*,\theta_k)] \leq \frac{L \max\{\frac{C^2}{\mu^2}, \mathbb{E}[\|x_{k-1} - x_k^*\|^2]\}}{M_k}.$$

It can be seen that the following two inequalities hold for x_k^* and x^*.

$$(\nabla_x f(x^*,\theta^*) - \nabla_x f(x_k^*,\theta^*))^T(x^* - x_k^*) \geq \mu \|x_k^* - x^*\|^2 \tag{5}$$

$$\nabla_x f(x^*,\theta^*)^T(x_k^* - x^*) \geq 0. \tag{6}$$

Adding (5) and (6), we obtain that

$$\begin{aligned}\mu\|x_k^* - x^*\|^2 &\leq \nabla_x f(x_k^*,\theta^*)^T(x_k^* - x^*) \\ &= (\nabla_x f(x_k^*,\theta^*) - \nabla_x f(x_k^*,\theta_k))^T(x_k^* - x^*) + \underbrace{\nabla_x f(x_k^*,\theta_k)^T(x_k^* - x^*)}_{\leq 0} \\ &\leq \|\nabla_x f(x_k^*,\theta^*) - \nabla_x f(x_k^*,\theta_k)\| \|x_k^* - x^*\| \\ &\leq L_\theta \|\theta_k - \theta^*\| \|x_k^* - x^*\|, \end{aligned} \tag{7}$$

where the first inequality follows from noting that $x_k^* = \text{argmin}_{x \in X} f(x,\theta_k)$, implying that $\nabla_x f(x_k^*,\theta_k)^T(x^* - x_k^*) \geq 0$. This implies that $\|x_k^* - x^*\| \leq \frac{L_\theta}{\mu}\|\theta_k - \theta^*\|$. By invoking the mean-value theorem, we note that that term c can be bounded as follows where $\tilde{x} \in [x_k^*, x^*]$.

$$\begin{aligned}f(x_k^*,\theta_k) - f(x^*,\theta_k) = \nabla_x f(\tilde{x},\theta_k)^T(x_k^* - x^*) \implies |f(x_k^*,\theta_k) - f(x^*,\theta_k)| &\leq \|\nabla_x f(\tilde{x},\theta_k)\| \|x_k^* - x^*\| \\ &\leq C_x \tfrac{L_\theta}{\mu} \|\theta_k - \theta^*\|.\end{aligned}$$

By taking expectations on both sides on (4) and by recalling the bounds on each of the terms, we obtain the following relationship.

$$\begin{aligned}&\mathbb{E}[|f(x_k,\theta^*) - f(x^*,\theta^*)|] = \mathbb{E}[f(x_k,\theta^*) - f(x^*,\theta^*)] \\ &\leq C_\theta \mathbb{E}[\|\theta_k - \theta^*\|] + \frac{L \max\{\frac{C^2}{\mu^2}, \mathbb{E}[\|x_{k-1} - x_k^*\|^2]\}}{M_k + 1} + C_x \tfrac{L_\theta}{\mu} \mathbb{E}[\|\theta_k - \theta^*\|] + C_\theta \mathbb{E}[\|\theta_k - \theta^*\|] \\ &= \left(2C_\theta + \tfrac{C_x L_\theta}{\mu}\right) \mathbb{E}[\|\theta_k - \theta^*\|] + \frac{L \max\{\frac{C^2}{\mu^2}, \mathbb{E}[\|x_{k-1} - x_k^*\|^2]\}}{M_k + 1}.\end{aligned}$$

∎

Thus, for the step size sequence $\gamma_{k,j} = \frac{1}{\mu j}$, if $M_k \geq \sqrt{N_k}$, then the leading term of the right-hand side of (3) becomes $\mathcal{O}(N_k^{-1/2})$ from (1). However, such a step size sequence is chosen as if the SA iterations at each k are completely independent from prior iterations. In other words, Lemma 1 holds even if $x_{k,1}$ is chosen randomly at each k instead of warm-starting from x_{k-1}. We expect that a randomly selected x_{k-1} might prove less efficient than our scheme in (2) and one might expect $\{M_k\}$ growing much slower than $\{\sqrt{N_k}\}$ may still be sufficient to guarantee the convergence rate of $\mathcal{O}(N_k^{-1/2})$ in terms of expected sub-optimality. Indeed, we show in Proposition 1 that if we continue using the same sequence of step sizes such that $\gamma_{k,j} = 1/\mu(\sum_{i=1}^{k-1} M_i + j)$, then we may reduce the SA effort required to achieve the optimal rate. We first state the following lemma, which is needed for the proof of Proposition 1.

Lemma 2 Given the sequence of streaming data, suppose θ_k and θ_{k+1} are MLEs computed from N_k and $N_{k+1} = N_k + n_{k+1}$ observations. Under the same regularity conditions for (1), $\mathbb{E}[\|\theta_{k+1} - \theta_k\|^2] \leq \frac{R_k n_{k+1}}{N_k^2}$ for some finite $R_k \in \mathbb{R}^+$ for each $k \geq 1$, where R_k converges to a constant as $k \to \infty$.

Here, we omit the proof of Lemma 2 for a general MLE due to space limitations. Instead, we prove Lemma 2 for a simpler case where θ_k is a sample average of the N_k observations. For instance, if $G(\cdot)$ is an exponential distribution, such a θ_k is the MLE of the mean of the distribution.

Proof. Let $\{Z_i\}$ represents the sequence of i.i.d. observations with variance σ^2. For the simple case mentioned above, $\theta_k = \sum_{i=1}^{N_k} Z_i/N_k$ for any k and $\theta^* = \mathbb{E}[Z_i]$. Further, we define $S_{k,k+1} \equiv \sum_{i=N_k+1}^{N_{k+1}} Z_i$ so that $\theta_{k+1} = \frac{N_k\theta_k + S_{k,k+1}}{N_k + n_{k+1}}$. Thus, $\theta_{k+1} - \theta_k = \frac{S_{k,k+1}}{N_k+n_{k+1}} - \frac{n_{k+1}\theta_k}{N_k+n_{k+1}}$. Since θ_k is unbiased for all k,

$$\mathbb{E}[|\theta_{k+1} - \theta_k|^2] = \mathrm{Var}(\theta_{k+1} - \theta_k) = \frac{n_{k+1}\sigma^2}{(N_k + n_{k+1})^2} + \frac{n_{k+1}^2\sigma^2}{N_k(N_k + n_{k+1})^2}.$$

Thus, for any k, we can find $R_k < \infty$ such that $\mathbb{E}[|\theta_{k+1} - \theta_k|^2] = \frac{R_k n_{k+1}}{N_k^2}$. In this particular case, $R_k \to \sigma^2$. ∎

Note that Lemma 2 can be generalized to a more general class of estimators including generalized method of moment and least-squares estimators. Now, we proceed to show the main result of this paper.

Proposition 1 Suppose Assumptions 1 and 2 hold, and the sequence of streaming data arrives whose cumulative number of observations is $\{N_k\}_{k\geq 1}$.

(i) If $\{M_k\}_{k\geq 1}$ and $\alpha > 0$ are chosen such that the following holds for every $k \geq 0$

$$\left(1 + \sum_{i=1}^{k} M_i\right)^{1+\alpha} \leq \frac{N_k^2}{n_{k+1}}, \tag{8}$$

then the following holds for any k.

$$\mathbb{E}[|f(x_k,\theta^*) - f(x^*,\theta^*)|] \leq \left(2C_\theta + \frac{C_x L}{\mu}\right)\mathbb{E}[\|\theta_k - \theta^*\|] + 2L\frac{\max\{\frac{C^2}{\mu^2}, \mathbb{E}[\|x_0 - x_1^*\|^2]\} + \frac{L_\theta}{\mu}\sum_{i=1}^{k-1} R_i(1 + \sum_{\ell=1}^{i} M_\ell)^{-\alpha}}{1 + \sum_{i=1}^{k} M_i},$$

where $\{R_k\}_{k\geq 1}$ is a sequence of nonnegative real numbers that converges to a constant.

(ii) If in addition N_k satisfies the following requirement,

$$N_k^{\frac{1}{2}} \leq 1 + \sum_{i=1}^{k} M_i, \quad \sum_{i=1}^{k}\left(1 + \sum_{\ell=1}^{i} M_\ell\right)^{-\alpha} \xrightarrow{k\to\infty} (\text{constant}), \tag{9}$$

then $\mathbb{E}[|f(x_k,\theta^*) - f(x^*,\theta^*)|] = \mathcal{O}(N_k^{-1/2})$.

Proof. (i). From the proof of Lemma 1, the only difference lies in deriving a modified rate statement for Term b, which is the only part affected by using an alternate set of step-length sequences. When $k=1$, the step sizes are identical to those used in Lemma 1. Thus, we have

$$\mathbb{E}[\|x_1 - x_1^*\|^2] \le \frac{\max\{\frac{C^2}{\mu^2}, \mathbb{E}[\|x_0 - x_1^*\|^2]\}}{M_1+1}.$$

For the first step of the $k=2$ problem, the following bound is obtained from triangle inequality.

$$\mathbb{E}[\|x_{2,1} - x_2^*\|^2] \le 2\underbrace{\mathbb{E}[\|x_{2,1} - x_1^*\|^2]}_{=\mathbb{E}[\|x_1 - x_1^*\|^2]} + 2\mathbb{E}[\|x_2^* - x_1^*\|^2] \le \frac{2\max\{\frac{C^2}{\mu^2}, \mathbb{E}[\|x_0 - x_1^*\|^2]\}}{M_1+1} + 2\mathbb{E}[\|x_2^* - x_1^*\|^2].$$

As shown in (7), we may claim that $\mathbb{E}[\|x_2^* - x_1^*\|^2] \le \frac{L_\theta}{\mu}\mathbb{E}[\|\theta_2 - \theta_1\|^2]$. From Lemma 2, we have $\mathbb{E}[\|\theta_{k+1} - \theta_k\|^2] \le \frac{R_k n_{k+1}}{N_k^2}$ for some converging sequence $\{R_k\} \in \mathbb{R}^+$. Therefore,

$$\begin{aligned}
\mathbb{E}[\|x_{2,1} - x_2^*\|^2] &\le \frac{2\max\{\frac{C^2}{\mu^2}, \mathbb{E}[\|x_0 - x_1^*\|^2]\}}{M_1+1} + \frac{2L_\theta}{\mu}\frac{R_1 n_2}{N_1^2} \\
&\overset{(8)}{\le} \frac{2\max\{\frac{C^2}{\mu^2}, \mathbb{E}[\|x_0 - x_1^*\|^2]\}}{M_1+1} + \frac{2L_\theta}{\mu}\frac{R_1}{(M_1+1)^{1+\alpha}} \\
&= \frac{2\max\{\frac{C^2}{\mu^2}, \mathbb{E}[\|x_0 - x_1^*\|^2]\} + \frac{2L_\theta}{\mu}R_1(M_1+1)^{-\alpha}}{M_1+1}
\end{aligned}$$

Next, we show that for $1 \le j \le M_2 - 1$,

$$\mathbb{E}[\|x_{2,j+1} - x_2^*\|^2] \le \frac{2\max\{\frac{C^2}{\mu^2}, \mathbb{E}[\|x_0 - x_1^*\|^2]\} + \frac{2L_\theta}{\mu}R_1(1+M_1)^{-\alpha}}{M_1+j+1}. \tag{10}$$

For $1 \le j \le M_2 - 1$, by directly applying the proof from Shapiro et al. (2009), Chapter 5.9.1

$$\mathbb{E}[\|x_{2,j+1} - x_2^*\|^2] \le (1 - 2\gamma_{2,j}\mu)\mathbb{E}[\|x_{2,j} - x_2^*\|^2] + \gamma_{2,j}^2 C^2.$$

Setting $\gamma_{2,j} = 1/\mu(M_1+j)$, we obtain for $j=1$

$$\begin{aligned}
\mathbb{E}[\|x_{2,2} - x_2^*\|^2] &\le (1 - 2/(M_1+1))\mathbb{E}[\|x_{2,1} - x_2^*\|^2] + C^2/\mu^2(M_1+1)^2 \\
&\overset{(10)}{\le} \left(1 - \frac{2}{(M_1+1)}\right)\frac{2\max\{\frac{C^2}{\mu^2}, \mathbb{E}[\|x_0 - x_1^*\|^2]\} + \frac{2L_\theta}{\mu}R_1(M_1+1)^{-\alpha}}{M_1+1} + \frac{C^2}{\mu^2(M_1+1)^2}.
\end{aligned} \tag{11}$$

We now consider two cases.
Case i. Suppose $\mathbb{E}[\|x_0 - x_1^*\|^2] \ge C^2/\mu^2$. Then we have that the following holds.

$$\begin{aligned}
(11) &\le \left(1 - \frac{2}{(M_1+1)}\right)\frac{2\mathbb{E}[\|x_0 - x_1^*\|^2] + \frac{2L_\theta}{\mu}R_1(M_1+1)^{-\alpha}}{M_1+1} + \frac{\mathbb{E}[\|x_0 - x_1^*\|^2]}{(M_1+1)^2} \\
&\le \left(2\mathbb{E}[\|x_0 - x_1^*\|^2] + \frac{2L_\theta}{\mu}R_1(M_1+1)^{-\alpha}\right)\underbrace{\left(\frac{1}{M_1+1} - \frac{1}{(M_1+1)^2}\right)}_{\le \frac{1}{M_1+1} - \frac{1}{(M_1+1)(M_1+2)}} \le \frac{2\mathbb{E}[\|x_0 - x_1^*\|^2] + \frac{2L_\theta}{\mu}R_1(M_1+1)^{-\alpha}}{M_1+2}.
\end{aligned}$$

Case ii. Similarly, if $\mathbb{E}[\|x_0 - x_1^*\|^2] < C^2/\mu^2$. Then the following holds.

$$
\begin{aligned}
(11) &\le \left(1 - \frac{2}{(M_1+1)}\right) \frac{\frac{2C^2}{\mu^2} + \frac{2L_\theta}{\mu} R_1 (M_1+1)^{-\alpha}}{M_1+1} + \frac{\frac{C^2}{\mu^2}}{(M_1+1)^2} \\
&\le \left(\tfrac{2C^2}{\mu^2} + \tfrac{2L_\theta}{\mu} R_1 (M_1+1)^{-\alpha}\right) \left(\frac{1}{M_1+1} - \frac{1}{(M_1+1)^2}\right) \le \frac{\frac{2C^2}{\mu^2} + \frac{2L_\theta}{\mu} R_1 (M_1+1)^{-\alpha}}{M_1+2}.
\end{aligned}
$$

Combining the two cases above, (10) holds for $j=1$. For $2 \le j \le M_j - 1$, we assume that (10) holds. Taking the last step of the $k=2$ problem, it follows that

$$
\mathbb{E}[\|x_2 - x_2^*\|^2] \le \frac{2\max\{\frac{C^2}{\mu^2}, \mathbb{E}[\|x_0 - x_1^*\|^2]\} + \frac{2L_\theta}{\mu} R_1 (M_1+1)^{-\alpha}}{M_1 + M_2 + 1}.
$$

Proceeding inductively, we have that the required relation holds for $k \ge 2$.

$$
\mathbb{E}[\|x_k - x_k^*\|^2] \le \frac{2\max\{\frac{C^2}{\mu^2}, \mathbb{E}[\|x_0 - x_1^*\|^2]\} + \frac{2L_\theta}{\mu} \sum_{i=1}^{k-1} R_i (1 + \sum_{\ell=1}^{i} M_\ell)^{-\alpha}}{1 + \sum_{i=1}^{k} M_i}.
$$

This result provides an upper bound for Term b in (4). The bounds on other terms are identical to those shown in the proof of Lemma 1. In the end, we have

$$
\begin{aligned}
\mathbb{E}[f(x_k,\theta^*) - f(x^*,\theta^*)] &\le \left(2C_\theta + \tfrac{C_x L}{\mu}\right) \mathbb{E}[\|\theta_k - \theta^*\|] \\
&\quad + 2L \frac{\max\{\frac{C^2}{\mu^2}, \mathbb{E}[\|x_0 - x_1^*\|^2]\} + \frac{L_\theta}{\mu} \sum_{i=1}^{k-1} R_i (1 + \sum_{\ell=1}^{i} M_\ell)^{-\alpha}}{1 + \sum_{i=1}^{k} M_i}.
\end{aligned}
$$

(ii). If α and $\{M_k\}_{k\ge1}$ are chosen to satisfy (9), then the first term of the right-hand side of the inequality above becomes the leading term in $O(N_k^{-1/2})$. ∎

Remark 1 For given $\{N_k\}$, a natural concern is whether there exists a feasible choice of sequence $\{M_k\}$ that satisfy the required relationships. Here we provide an example of $\alpha, \{N_k\}$, and $\{M_k\}$ that do satisfy the conditions in Proposition 1. Suppose $n_k = k$ for all $k \ge 2$ and $n_1 = N_1 \ge 4$, where in reality, N_1 should be large enough so that we have enough degrees of freedom to compute MLE θ_1. Consequently, $N_k = N_1 + \sum_{i=2}^{k} i = N_1 - 1 + k(k+1)/2$. Let $\alpha = 1+\delta$ where δ is a small positive scalar and we choose $\{M_k\}_{k\ge1}$ to satisfy $\sum_{i=1}^{k} M_i = \lceil\sqrt{N_k}\rceil - 1$. Note that this sampling rule allows for $M_i = 0$ for some i, which implies that we skip solving the ith problem. Then,

$$
\sum_{i=1}^{k-1}\left(1 + \sum_{\ell=1}^{i} M_\ell\right)^{-\alpha} \le \sum_{i=1}^{k-1} N_i^{-\alpha/2} = \sum_{i=1}^{k-1}\left(N_1 - 1 + \frac{i(i+1)}{2}\right)^{-(1+\delta)/2}
$$

converges to a constant because $\delta > 0$. Moreover,

$$
\sum_{i=1}^{k} M_i = \lceil\sqrt{N_k}\rceil - 1 \implies (1 + \textstyle\sum_{i=1}^{k} M_i)^{1+\alpha} = \left(\lceil\sqrt{N_k}\rceil\right)^{1+\alpha} \ge N_k^{\frac{1+\alpha}{2}}.
$$

Furthermore, $(1+\sum_{i=1}^{k} M_i)^{1+\alpha} \le (\sqrt{N_k}+1)^{1+\alpha} \le N_k^2/(k+1)$, where the last inequality holds for all $k \ge 1$ by selecting α to be sufficiently close to 1, given $N_1 \ge 4$.

Figure 1 provides a plot of $N_k^{\frac{1+\alpha}{2}}, (1+\sum_{i=1}^{k} M_i)^{1+\alpha}$, and N_k^2/n_{k+1} in log-scale for $k = 1,2,\ldots,20$ when $N_1 = 10$ and $\alpha = 1.1$. We observe that $(1+\sum_{i=1}^{k} M_i)^{1+\alpha}$ is strictly bounded between $N_k^{\frac{1+\alpha}{2}}$ and N_k^2/n_{k+1} for all k. The resulting $\{M_k\}_{k\ge1}$ is $\{3,0,0,1,0,1,1,0,\ldots\}$ for this example. Notice that we do not take SA steps for some k.

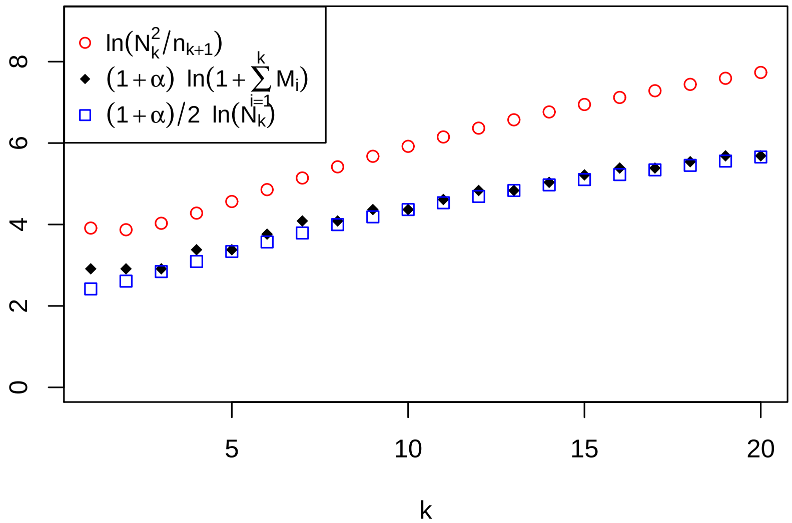


Figure 1: A plot of logarithmic values of $N_k^{\frac{1+\alpha}{2}}, (1+\sum_{i=1}^k M_i)^{1+\alpha}$, and N_k^2/n_{k+1} for $1 \le k \le 20$ when $\alpha = 1.1, N_1 = 10, n_k = k$ for $k \ge 2$ and M_k is chosen to satisfy $\sum_{i=1}^k M_i = \lceil \sqrt{N_k} \rceil - 1$ for $k \ge 1$.

4 EMPIRICAL PERFORMANCE

In this section, we examine the empirical performance of the proposed SA scheme analyzed in Proposition 1 by applying it to a simple SO problem whose optimal solution can be found numerically. Consider a service system that may be modeled as a $M/M/1/c$ queue with system capacity $c = 20$ where the arrival rate is $\theta^* = 0.7$. We assume the latter is unknown to us and needs to be estimated by observing the arrival process of the system. In particular, suppose the interarrival times are collected periodically and available in the form of streaming data. We further assume the sequence of sample sizes $\{n_k\}$ for this arrival process is the same as in Figure 1. Our goal is to control the service rate x to minimize the expected cost of operation:

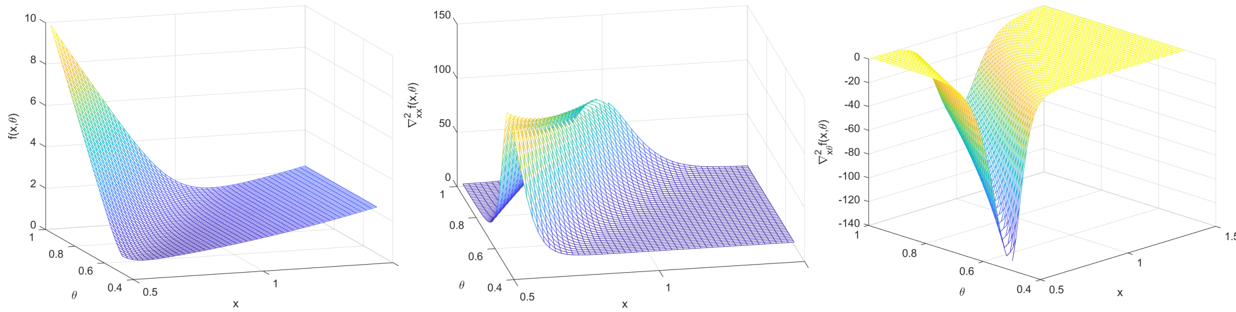
$$(\text{cost of operation}) = \beta(\text{number of rejected customers per unit time}) + (\text{service cost per unit time}),$$

where β is the cost of rejecting a customer due to the system capacity constraint. In particular, we choose $\beta = 20$ for our experiment and assume $(\text{service cost per unit time}) = x^2$, i.e., the cost grows quadratically as the service rate increases. Thus,

$$f(x,\theta) = \mathbb{E}[(\text{cost of operation})] = \beta\theta \Pr\{\text{balking}\} + x^2.$$

From the steady-state analysis of a $M/M/1/c$ queue, the balking probability can be derived as follows (see Medhi (2003))

$$\Pr\{\text{balking}\} = \begin{cases} \dfrac{1}{c+1}, & \text{if } x = \theta, \\ \dfrac{\theta^c(x-\theta)}{x^{c+1}-\theta^{c+1}}, & \text{otherwise.} \end{cases} \tag{12}$$



(a) $f(x,\theta)$ (b) $\nabla^2_{xx} f(x,\theta)$ (c) $\nabla^2_{x\theta} f(x,\theta)$

Figure 2: The expected cost function, $f(x,\theta)$, for the $M/M/1/c$ queueing example.

Note that (12) is continuous at $x=\theta$. Figure 2(a) displays the plot of $f(x,\theta)$ for $0.5 \le x \le 1.5$ and $0.5 \le \theta \le 1$. To confirm convexity of $f(x,\theta)$ in x, we derive $\nabla_x f(x,\theta)$ and $\nabla^2_{xx} f(x,\theta)$ for $x \ne \theta$ as follows.

$$\nabla_x f(x,\theta) = \beta\theta^{c+1}\left(\frac{1}{x^{c+1}-\theta^{c+1}} - \frac{(c+1)(x-\theta)x^c}{(x^{c+1}-\theta^{c+1})^2}\right) + 2x,$$

$$\nabla^2_{xx} f(x,\theta) = \beta\theta^{c+1}\left((x-\theta)\left(\frac{2(c+1)^2x^{2c}}{(x^{c+1}-\theta^{c+1})^3} - \frac{c(c+1)x^{c-1}}{(x^{c+1}-\theta^{c+1})^2}\right) - \frac{2(c+1)x^c}{(x^{c+1}-\theta^{c+1})^2}\right) + 2.$$

Additionally, it can be shown that $\lim_{x\to\theta^-}\nabla_x f(x,\theta) = \lim_{x\to\theta^+}\nabla_x f(x,\theta)$ and $\lim_{x\to\theta^-}\nabla^2_{xx} f(x,\theta) = \lim_{x\to\theta^+}\nabla^2_{xx} f(x,\theta)$. Figure 2(b) shows that $\nabla^2_{xx} f(x,\theta) \ge 2$, which confirms the first part of Assumption 1 holds for this example. Given $\theta^* = 0.7$, it can be found numerically that $f(x,\theta^*) = 0$ has a unique optimum at $x^* = 0.8386$. Moreover,

$$\nabla^2_{x\theta} f(x,\theta) = \beta(c+1)\left\{\frac{\theta^c}{x^{c+1}-\theta^{c+1}} + \frac{\theta^{c+1}(x^c+\theta^c)-(c+1)\theta^c(x-\theta)x^c}{(x^{c+1}-\theta^{c+1})^2} - \frac{2(c+1)\theta^{2c+1}(x-\theta)x^c}{(x^{c+1}-\theta^{c+1})^3}\right\},$$

which is displayed in Figure 2(c). Thus, by choosing Θ to be a compact set of $\theta \ge 0$, the second part of Assumption 1 can be satisfied.

Although $f(x,\theta)$ is known for this example, we did not take advantage of this knowledge when applying the SA scheme and only used it for the purpose of evaluation. Instead, we implemented a discrete-event simulator that estimates the cost function by simulating the system until 200 customers leave the system. To avoid the issue of initial bias, we sample the number of customers in the system at the beginning of the simulation from the steady-state number-in-system distribution of the $M/M/1/c$ queue. A similar trick is used in Song and Nelson (2015). To obtain the gradient estimator, $\nabla_x F(x_{k,j},\theta_k,\omega_{j,k})$, in (2), we apply the finite-difference method (Fu 2006) by running 10 replications at $x_{k,j}$ and $x_{k,j}+\delta$, respectively, where $\delta = 0.02$. We acknowledge that this introduces bias in the gradient estimation, a concern which will be revisited in future work. For the SA sample sizes, $\{M_k\}$, we used the same sequence presented in Figure 1.

Table 1 shows the results from 30 runs of our algorithm on this test problem. All runs were made for $k = 500$ iterations and the cumulative real-world sample sizes (N_k) and the cumulative number of SA steps ($\sum_{i=1}^k M_i$) are presented for $k = 50, 100, \ldots, 500$. To evaluate the performance, we estimated $\mathbb{E}[|f(x_k,\theta^*) - f(x^*,\theta^*)|]$ from the 30 runs with their standard errors displayed in parentheses. As a benchmark, we also present $\mathbb{E}[|f(x_k^*,\theta^*) - f(x^*,\theta^*)|]$ as the "best possible" sequence of error. Recall that x_k^* is the optimal solution for $\text{Opt}(\theta_k)$, which is obtained by numerically solving $\nabla_x f(x,\theta_k) = 0$ for x. Clearly, this is the best error possible, even if one expends infinite effort to solve each $\text{Opt}(\theta_k)$. Notice that our proposed algorithm has comparable error to that of the best possible case for $k \ge 200$, while the corresponding SA sample size grows slowly as k increases.

Table 1: Comparison between the proposed SA scheme with the best-possible progress.

k	50	100	150	200	250	300	350	400	450	500
N_k	1,284	5,059	11,334	20,109	31,384	45,159	61,434	80,209	101,484	125,259
Proposed SA										
$\sum_{i=1}^{k} M_i$	35	71	106	141	177	212	247	283	318	353
$\mathbb{E}[\lvert f(x_k,\theta^*)-f(x^*,\theta^*)\rvert]$	0.364	0.1086	0.0295	0.0087	0.0062	0.0057	0.0046	0.0039	0.0038	0.0036
	(0.0165)	(0.0115)	(0.0079)	(0.0049)	(0.0027)	(0.0012)	(0.0005)	(0.0002)	(0.0002)	(0.0002)
Best possible										
$\mathbb{E}[\lvert f(x_k^*,\theta^*)-f(x^*,\theta^*)\rvert]$	0.0311	0.0181	0.0126	0.0103	0.0077	0.005	0.0045	0.0043	0.0032	0.0032
	(0.0037)	(0.0023)	(0.0018)	(0.0014)	(0.0008)	(0.0006)	(0.0005)	(0.0005)	(0.0005)	(0.0004)

5 CONCLUSIONS

We consider the development of a stochastic approximation framework for contending with an SO problem when the parameters of the underlying distribution are unknown but may be learnt through a streaming data process. This is an instance of an SO problem afflicted by input uncertainty. Given a sequence of parameters that is converging to the true parameter associated with the input distribution, we present a framework in which an increasingly accurate solution of the simulation optimization problem is computed via an SA framework. We proceed to show that the produced sequence of estimators of the solution to the SO problem converges to their true counterpart in an expected-value sense and at the canonical Monte Carlo rate. In particular, our analysis provides requirements on the sample-size requirements in the SA scheme to ensure the achievement of the optimal rate. In future work, we will relax some of the assumptions made in this paper including that an unbiased estimator of the gradient is available.

REFERENCES

Ahmadi, H., and U. V. Shanbhag. 2014. "Data-driven First-order Methods for Misspecified Convex Optimization Problems: Global Convergence and Rate Estimates". In *Proceedings of the 53rd IEEE Conference on Decision and Control*, 4228–4233. December 15th–17th, Los Angeles, California.

Bischi, G. I., A. K. Naimzada, and L. Sbragia. 2007. "Oligopoly Games with Local Monopolistic Approximation". *Journal of Economic Behavior & Organization* 62(3):371–388.

Bischi, G. I., L. Sbragia, and F. Szidarovszky. 2008. "Learning the Demand Function in a Repeated Cournot Oligopoly Game". *International Journal of Systems Science* 39(4):403–419.

Corlu, C., and B. Biller. 2013. "A Subset Selection Procedure under Input Parameter Uncertainty". In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 463–473. Piscataway, New Jersey: Institute of Electrial and Electronics Engineers Inc.

Fan, W., L. J. Hong, and X. Zhang. 2013. "Robust Selection of the Best". In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 868–876. Piscataway, New Jersey: Institute of Electrial and Electronics Engineers Inc.

Fu, M. C. 2006. "Chapter 19 Gradient Estimation". In *Simulation*, edited by S. G. Henderson and B. L. Nelson, Volume 13 of *Handbooks in Operations Research and Management Science*, 575–616. Oxford, UK: Elsevier.

Gao, S., H. Xiao, E. Zhou, and W. Chen. 2017. "Robust Ranking and Selection with Optimal Computing Budget Allocation". *Automatica* 81:30–36.

Jiang, H., and U. V. Shanbhag. 2013. "On the Solution of Stochastic Optimization Problems in Imperfect Information Regimes". In *Proceedings of the 2013 Winter Simulations Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 821–832. Piscataway, New Jersey: Institute of Electrial and Electronics Engineers Inc.

Jiang, H., and U. V. Shanbhag. 2015. "Data-driven Schemes for Resolving Misspecified MDPs: Asymptotics and Error Analysis". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 3801–3812. Piscataway, New Jersey: Institute of Electrial and Electronics Engineers Inc.

Jiang, H., and U. V. Shanbhag. 2016. "On the Solution of Stochastic Optimization and Variational Problems in Imperfect Information Regimes". *SIAM Journal on Optimization* 26(4):2394–2429.

Jiang, H., U. V. Shanbhag, and S. P. Meyn. 2018. "Distributed Computation of Equilibria in Misspecified Convex Stochastic Nash Games". *IEEE Transactions on Automatic Control* 63(2):360–371.

Medhi, J. 2003. "CHAPTER 3 Birth-and-Death Queueing Systems: Exponential Models". In *Stochastic Models in Queueing Theory* (Second Edition ed.)., 65–164. Burlington, Massachusetts: Academic Press.

Newey, W. K., and D. McFadden. 1994. “Chapter 36 Large Sample Estimation and Hypothesis Testing”. Volume 4 of *Handbook of Econometrics*, 2111–2245. North Holland: Elsevier.

Shapiro, A., D. Dentcheva, and A. Ruszczyński. 2009. *Lectures on Stochastic Programming*. Philadelphia, Pennsylvania: Society for Industrial and Applied Mathematics.

Song, E., and B. L. Nelson. 2015. “Quickly Assessing Contributions to Input Uncertainty”. *IIE Transactions* 47(9):893–909.

Song, E., and B. L. Nelson. 2019. “Input-–Output Uncertainty Comparisons for Discrete Optimization via Simulation”. *Operations Research* 67(2):562–576.

Song, E., B. L. Nelson, and L. J. Hong. 2015. “Input Uncertainty and Indifference-zone Ranking & Selection”. In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 414–424. Piscataway, New Jersey: Institute of Electrial and Electronics Engineers Inc.

Szidarovszky, F. 2004. “Global Stability Analysis of a Special Learning Process in Dynamic Oligopolies”. *Journal of Economic Research* 9(2):175–190.

Wu, D., and E. Zhou. 2019. “Ranking and Selection under Input Uncertainty: Fixed Confidence and Fixed Budget”. https://arxiv.org/abs/1708.08526, accessed 16$^{\text{th}}$ September.

AUTHOR BIOGRAPHIES

EUNHYE SONG is Harold and Inge Marcus Early Career Assistant Professor in the Department of Industrial and Manufacturing Engineering at the Penn State University. Her research interests include simulation design of experiments, simulation uncertainty and risk quantification, optimization via simulation under input model risk and large-scale discrete optimization via simulation. Her email address is eus358@psu.edu. Her website is http://eunhyesong.info.

UDAY V. SHANBHAG holds the Gary and Sheila Bello Chair in the Harold and Inge Marcus Department of Industrial and Manufacturing Engineering at Pennsylvania State University, prior to which he was an assistant and (tenured) associate professor at the University of Illinois at Urbana-Champaign (2006-2012). He holds a Ph.D. in Management science and Engineering from Stanford University (2006). His research interests lies in stochastic and nonlinear optimization, nonsmooth analysis and variational inequality problems. His email address is udaybag@psu.edu. His website is http://www.personal.psu.edu/vvs3/.