Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

OPTIMIZATION OF QUEUEING SYSTEMS USING STREAMING SIMULATION

Robert J. Lambert[1], James A. Grant[2], Rob Shone[3], and Roberto Szechtman[4]

[1]STOR-i Centre for Doctoral Training, Lancaster University, Lancaster, LA1 4YW, UK
[2]School of Mathematical Sciences, Lancaster University, Lancaster, LA1 4YF, UK
[3]Department of Management Science, Lancaster University, Lancaster, LA1 4YX, UK.
[4]Operations Research Department, Naval Postgraduate School, Monterey, CA 93943, USA

ABSTRACT

We consider the problem of adaptively determining the optimal number of servers in an $M/G/c$ queueing system in which the unknown arrival rate must be estimated using data that arrive sequentially over a series of observation periods. We propose a stochastic simulation-based approach that uses iteratively updated parameters within a greedy decision-making policy, with the selected number of servers minimising a Monte Carlo estimate of a chosen objective function. Under minimal assumptions, we derive a central limit theorem for the Monte Carlo estimator and derive an asymptotic bound on the probability of incorrect selection of the policy. We also demonstrate the empirical performance of the policy in a finite-time numerical experiment.

1 INTRODUCTION

Simulation optimization is a popular tool that can be used as a method for effectively optimising complex stochastic systems (He and Song 2024). This is typically considered to be an offline procedure, where model input parameters are estimated from some fixed collection of observations and a chosen policy is used to find the optimal system conditions. Recently, a body of work has emerged focusing on tackling simulation optimization problems in an online setting. Specifically, it considers problems where system observations are not fixed and instead are obtained sequentially, also known as *streaming data*. Such problems are of interest for two main reasons. Firstly, when solving these problems we require a different set of analytical and methodological tools to those used in the classical simulation optimization literature, often taking inspiration from fields such as machine learning and online decision-making. Secondly, with recent technological advancements, streaming data is more prevalent across a wide range of domains, facilitating the need for adaptive optimization procedures for complex and evolving stochastic systems.

The first instance of extending simulation optimization procedures to account for streaming observations was within the context of Ranking and Selection (R&S) problems. Wu and Zhou (2019) propose an extension of the Optimal Computing Budget Allocation (OCBA) algorithm that accounts for diminishing input uncertainty as a consequence of the sequential accumulation of observations. Wu, Wang, and Zhou (2024) also extend existing fixed-confidence R&S procedures to account for streaming input data. They propose extensions of Sequential Elimination algorithms that allow for the aggregation of simulation output across periods with changing underlying input parameters.

Song and Shanbhag (2019) consider a multi-period simulation optimization problem, specifically focusing on optimising over a continuous decision space. During each period, they receive a new batch of i.i.d. system observations and collate them with all those acquired over previous rounds. They then iteratively update unknown system parameters, before implementing a warm-started stochastic gradient descent algorithm to minimize some expected cost function. This work is extended in He, Shanbhag, and Song (2024), where additional variants of the original algorithm are proposed and assumptions, such as access to an unbiased gradient estimator, are relaxed.

In this paper, we consider optimizing the number of servers within a multi-server queueing system with streaming observations. This is a problem with a discrete decision space, similar to that considered by Wu and Zhou (2019) and Wu, Wang, and Zhou (2024). However, rather than adopting an R&S approach, we develop a policy and evaluate its performance using a stochastic process framework, similar to that of Gibbons, Grant, and Szechtman (2023). We consider the motivating problem of an $M/G/c$ queueing system which is both widely applicable and analytically challenging, due to the general service time distribution. We contribute to the streaming simulation literature by providing a detailed theoretical analysis for a generalisable class of objective functions, allowing for a broader range of applications.

The remainder of the paper is structured as follows. In Section 2 we formally introduce the sequential optimization problem. Section 3 outlines the proposed decision-making policy that utilizes streaming simulation data. In Section 4 we evaluate the asymptotic performance of the policy, deriving a central limit theorem for the simulated Monte Carlo estimate, as well as a vanishing bound on the asymptotic probability of incorrect selection. In Section 5, the finite-time empirical performance of the policy is evaluated in a numerical experiment.

2 PROBLEM STATEMENT

Consider a general $M/G/c$ queueing system over the time interval $[0,T]$, where $T < \infty$. The system is initialized as empty and customers arrive according to a Poisson process with constant arrival rate $\eta \in \mathbb{R}_+$. Service times follow some light-tailed distribution parameterized by $\theta \in \mathbb{R}^d$, for $d \in \mathbb{N}$, and the system employs $c \in \mathbb{N}$ servers. Let the random variable $X(c;\eta,\theta) \geq 0$ denote some measure of the system's transient behavior on $[0,T]$, e.g., mean queue length, mean customer waiting time, or the proportion of customers whose waiting times exceed some pre-determined threshold. Let α denote the expected value of X, defined as

$$\alpha(c;\eta,\theta) := \mathbb{E}[X(c;\eta,\theta)]. \tag{1}$$

We focus on determining the optimal number of servers to employ within an $M/G/c$ queueing system with finite, unknown arrival rate, $\lambda > 0$. The way we infer the value of λ is by observing i.i.d. realizations of the arrival process over a number of discrete observation periods, $n \in \mathbb{N}$, each of length T. We assume that we have knowledge of the service distribution parameters θ and let $\mathscr{C} \subset \mathbb{N}$ denote the finite set of possible numbers of servers to employ within the system. Each period, we must use the accumulated data within a policy that determines the optimal number of servers to implement within the system. We let $c_n \in \mathscr{C}$ denote the number of servers chosen by our policy during period n.

We consider an expected objective function that is a sum of two costs. The first corresponds to the cost incurred by implementing an inefficient system and is given by the expected performance measure stated in (1) evaluated at $\eta = \lambda$. The second is the cost associated with implementing additional servers within the system and can be expressed as some deterministic function $\beta : \mathscr{C} \mapsto \mathbb{R}$. We let f denote the expected objective for our problem,

$$f(c;\lambda,\theta) := \alpha(c;\lambda,\theta) + \beta(c). \tag{2}$$

Our goal is to sequentially choose server allocations c_n over decision periods n so as to minimize the cumulative expected loss incurred across all periods. Obtaining a closed form expression for the expected performance measure α is often not possible when considering the transient behavior of an $M/G/c$ queue. In Section 3, we propose a decision-making policy that approximates α using a Monte Carlo estimator.

3 DECISION-MAKING POLICY

Let $\{N(t;\lambda); t \in [0,T]\}$ denote the homogeneous Poisson arrival process for the $M/G/c$ system being optimized. In addition, let $\{N_i(t;\lambda); t \in [0,T]\}_{i=1}^n$ denote the i.i.d. realizations of this process observed during periods $i = 1,\ldots,n$. By considering the cumulative number of arrivals to the system, we can approximate the arrival rate parameter with the MLE of a Poisson process,

$$\bar{\lambda}_n = \frac{1}{n}\sum_{i=1}^{n} N_i(T;\lambda)/T. \tag{3}$$

We propose a decision-making policy that carries out repeated simulation to approximate the expected objective function. Let $m_n \in \mathbb{N}$ denote the period n simulation budget allocation for each $c \in \mathscr{C}$. In absence of the true parameter λ, we simulate i.i.d. realizations of X for each $c \in \mathscr{C}$ using the estimate $\bar{\lambda}_n$, which we denote $X_j(c;\bar{\lambda}_n)$ for $j = 1,\ldots,m_n$. We use a standard Monte Carlo estimator in our analysis. While variance reduction techniques may improve finite-time efficiency, they do not affect asymptotic behavior and yield similar results up to constant factors. Since our focus is on asymptotic properties, we prioritize analytical tractability over variance reduction in this work.

As we assume that the service distribution is known, we omit service distribution parameters $\theta \in \mathbb{R}^d$ from the above and all further notation. We define the nth period Monte Carlo estimator of $\alpha(c;\bar{\lambda}_n)$ as follows,

$$\bar{X}_n(c;\bar{\lambda}_n) := \frac{1}{m_n}\sum_{j=1}^{m_n} X_j(c;\bar{\lambda}_n). \tag{4}$$

The estimate given in expression (4) allows us to obtain the following approximation of the expected objective (2),

$$\bar{f}_n(c;\bar{\lambda}_n) = \bar{X}_n(c;\bar{\lambda}_n) + \beta(c). \tag{5}$$

We propose a decision-making policy that greedily selects the number of servers that minimizes the approximated objective

$$c_n = \operatorname*{argmin}_{c\in\mathscr{C}} \bar{f}_n(c;\bar{\lambda}_n) \tag{6}$$

where any ties are broken arbitrarily.

Algorithm 1 gives the full pseudocode for the greedy decision-making policy.

Algorithm 1: Greedy decision-making policy

for $n \in \mathbb{N}$ **do**
- Observe number of arrivals, $N_n(T;\lambda)$
- Update arrival rate estimate, $\bar{\lambda}_n = \frac{1}{n}\sum_{i=1}^{n} N_i(T;\lambda)/T$
- **for** $c \in \mathscr{C}$ **do**
  - **for** $j = 1,\ldots,m_n$ **do**
    - Given $\bar{\lambda}_n$, simulate stochastic performance measure, $X_j(c;\bar{\lambda}_n)$
  - **end**
  - Obtain Monte Carlo estimate, $\bar{X}_n(c;\bar{\lambda}_n) = \frac{1}{m_n}\sum_{j=1}^{m_n} X_j(c;\bar{\lambda}_n)$
  - Approximate objective, $\bar{f}_n(c;\bar{\lambda}_n) = \bar{X}_n(c;\bar{\lambda}_n) + \beta(c)$
- **end**
- Greedily select minimizer $c_n = \operatorname*{argmin}_{c\in\mathscr{C}} \bar{f}_n(c;\bar{\lambda}_n)$

end

For clarity and conciseness, we omit explicit dependence on $c \in \mathscr{C}$ when it is clear from context. Additionally, a subscript n indicates dependence on the parameter $\bar{\lambda}_n$.

4 ASYMPTOTIC RESULTS

In Section 4.1 we derive a central limit theorem (CLT) for the Monte Carlo estimate defined in (4). In Section 4.2 we show that the greedy policy is asymptotically optimal by bounding the limiting probability of incorrect selection.

4.1 Central Limit Theorem

In the following analysis, we define the conditional expected performance measure α_n as follows

$$\alpha_n(c) := \mathbb{E}[X(c;\eta) \mid \eta = \bar{\lambda}_n], \quad \text{for all } n \in \mathbb{N}. \tag{7}$$

We make the following assumption about (7),

Assumption 1 For all $c \in \mathscr{C}$, and $n \in \mathbb{N}$,

$$\alpha_n(c) < \infty, \text{ w.p. } 1.$$

Let $N_j(T;\bar{\lambda}_n)$ denote the total number of arrivals during simulation replication $j \in \{1,\dots,m_n\}$ during period n. We make the following assumption about the boundedness of $X_j(c;\bar{\lambda}_n)$.

Assumption 2 There exists finite $q \in \mathbb{N}$ such that for all $c \in \mathscr{C}$ and $n \in \mathbb{N}$, $X_j(c;\bar{\lambda}_n) \geq 0$ is bounded by the order q monomial of $N_j(T;\bar{\lambda}_n)$, i.e., for all $j \in \{1,\dots,m_n\}$

$$X_j(c;\bar{\lambda}_n) \leq N_j(T;\bar{\lambda}_n)^q, \text{ w.p. } 1.$$

System efficiency is largely driven by the number of customers present. It is therefore natural and appropriate to bound performance measures, X, with functions of the arrival process, N. In particular, for systems with light-tailed service distributions, Assumption 2 holds for many common forms of X, including those proposed in Section 2.

The following lemma shows that the finite moments of the centred random variable $X_j(c;\bar{\lambda}_n) - \alpha_n(c)$ are finite and is necessary for the derivation of the CLT.

Lemma 1 For finite integer $p \geq 1$, and $X_j(c;\bar{\lambda}_n)$ satisfying Assumptions 1 and 2,

$$\mathbb{E}[(X_j(c;\bar{\lambda}_n) - \alpha_n(c))^p] < \infty, \quad \text{for all } j \in \{1,\dots,m_n\}.$$

Proof of Lemma 1. Without loss of generality, $\mathbb{E}[X_j(c;\bar{\lambda}_n)] = \mathbb{E}[X_1(c;\bar{\lambda}_n)]$ due to the i.i.d. property of the simulation replications. By the binomial expansion and Assumption 2,

$$\begin{aligned}\mathbb{E}[(X_1(c;\bar{\lambda}_n) - \alpha_n(c))^p] \leq \mathbb{E}[(X_1(c;\bar{\lambda}_n) + \alpha_n(c))^p] &= \sum_{k=0}^{p} \binom{p}{k} \alpha_n(c)^{p-k} \mathbb{E}[(X_1(c;\bar{\lambda}_n)^k] \\ &\leq \sum_{k=0}^{p} \binom{p}{k} \alpha_n(c)^{p-k} \mathbb{E}[N_1(T;\bar{\lambda}_n)^{qk}]. \end{aligned} \tag{8}$$

By the total law of expectation,

$$\mathbb{E}[N_1(T;\bar{\lambda}_n)^{qk}] = \mathbb{E}[\mathbb{E}[N_1(T;\eta)^{qk} \mid \eta = \bar{\lambda}_n]]. \tag{9}$$

Using the properties of the Poisson arrival process, $N_1(T;\eta) \sim \text{Pois}(\eta T)$. For integer $qk \geq 0$, the interior expected value $\mathbb{E}[N_1(T;\eta)^{qk} \mid \eta = \bar{\lambda}_n]$ corresponds to the degree qk Touchard polynomial in $\bar{\lambda}_n T$,

$$S_{qk}(\bar{\lambda}_n T) = \sum_{r=1}^{qk} (\bar{\lambda}_n T)^r \left\{ {qk \atop r} \right\},$$

where the values in $\{\}$ denote the Stirling numbers of the second kind.

As each term in $S_{qk}(\bar{\lambda}_n T)$ is a finite power of $\bar{\lambda}_n = \frac{1}{n}\sum_{i=1}^{n} N_i(T;\lambda)/T$, where system arrivals $N_i(T;\lambda) \sim$ Poisson(λT), it follows that (9) is finite. By Assumption 1, $\alpha_n(c) < \infty$ and by definition, q, p are both finite. Therefore, expression (8) is a linear combination of finite terms and is itself finite. □

The final assumption required to derive the CLT concerns the asymptotic behavior of the simulation budget allocation.

Assumption 3 As the number of observation periods $n \to \infty$, the number of simulation replications m_n allocated to each $c \in \mathscr{C}$ satisfies $m_n \to \infty$.

In the Monte Carlo estimator for α, $\bar{X}_n$ there are two sources of uncertainty. These correspond to the parameter uncertainty in $\bar{\lambda}_n$ and the simulation uncertainty due to carrying out a finite number of simulation replications, m_n. As $n \to \infty$, by the Strong Law of Large numbers, $\bar{\lambda}_n \xrightarrow{a.s.} \lambda$ and the associated input uncertainty vanishes in the limit. To ensure that the total variance vanishes, it is necessary to assume that $m_n \to \infty$ also.

In Theorem 1 we state the CLT for $\bar{X}_n(c;\bar{\lambda}_n)$, letting $\xrightarrow{d}$ denote convergence in distribution.

Theorem 1 The Monte Carlo estimate $\bar{X}_n(c;\bar{\lambda}_n)$ satisfies

$$\sqrt{m_n}(\bar{X}_n(c;\bar{\lambda}_n) - \alpha_n(c)) \xrightarrow{d} \sigma(c)N(0,1), \quad \text{as } n \to \infty, \tag{10}$$

where $\sigma^2(c) = \lim_{n\to\infty} \mathrm{Var}(X_1(c;\bar{\lambda}_n))$.

Proof of Theorem 1. To derive the desired CLT in terms of $\bar{X}_n(c;\bar{\lambda}_n)$, we first consider the standardized random variable $\tilde{X}_j(c;\bar{\lambda}_n)$ for each $c \in \mathscr{C}$, defined as follows

$$\tilde{X}_j(c;\bar{\lambda}_n) = \frac{1}{\sqrt{m_n}}\left(X_j(c;\bar{\lambda}_n) - \alpha_n(c)\right), \tag{11}$$

with $\mathbb{E}[\tilde{X}_j(c;\bar{\lambda}_n)] = 0$, for all $n \in \mathbb{N}$ and $j \in \{1,\dots,m_n\}$.

In the following, (i) and (ii) are conditions for the Lindeberg-Feller CLT to hold for the sum of standardized random variables $\tilde{X}_j(c;\bar{\lambda}_n)$ (Lindeberg 1922; Durrett 2019). We state these conditions and prove that they hold individually.

(i) $\sum_{j=1}^{m_n} \mathbb{E}\left[\tilde{X}_j(c;\bar{\lambda}_n)^2\right] \to \sigma(c)^2, \quad \text{as } n \to \infty.$

As the simulation realizations for any $c \in \mathscr{C}$ are i.i.d. within a given period $n \in \mathbb{N}$, we have

$$\sum_{j=1}^{m_n} \mathbb{E}\left[\tilde{X}_j(c;\bar{\lambda}_n)^2\right] = \mathbb{E}\left[(X_1(c;\bar{\lambda}_n) - \alpha_n(c))^2\right].$$

As $X_j(c;\bar{\lambda}_n) \geq 0$ a.s., it follows that $\mathbb{E}\left[\tilde{X}_j(c;\bar{\lambda}_n)^2\right] \geq 0$, meaning the LHS is a summation of non-negative terms. The RHS is independent of m_n and by Lemma 1 is finite and bounded. By Assumption 3, the series $\sum_{j=1}^{m_n} \mathbb{E}\left[\tilde{X}_j(c;\bar{\lambda}_n)^2\right]$ as $n \to \infty$ converges to some constant, which we denote $\sigma(c)^2 > 0$. Thus condition (i) holds.

(ii) For all $\varepsilon > 0$, $\lim_{n\to\infty} \sum_{j=1}^{m_n} \mathbb{E}\left[\tilde{X}_j(c;\bar{\lambda}_n)^2 \mathbb{1}\left\{|\tilde{X}_j(c;\bar{\lambda}_n)| > \varepsilon\right\}\right] = 0.$

Within period $n \in \mathbb{N}$, the repeated simulations for any $c \in \mathscr{C}$ are i.i.d., giving

$$\begin{aligned}\sum_{j=1}^{m_n} \mathbb{E}\left[\tilde{X}_j(\bar{\lambda}_n)^2 \mathbb{1}\left\{|\tilde{X}_j(\bar{\lambda}_n)| > \varepsilon\right\}\right] &= m_n \mathbb{E}\left[\left(\frac{X_1(\bar{\lambda}_n) - \alpha_n}{\sqrt{m_n}}\right)^2 \mathbb{1}\left\{\left|\frac{X_1(\bar{\lambda}_n) - \alpha_n}{\sqrt{m_n}}\right| > \varepsilon\right\}\right] \\ &= \mathbb{E}\left[\left(X_1(\bar{\lambda}_n) - \alpha_n\right)^2 \mathbb{1}\left\{|X_1(\bar{\lambda}_n) - \alpha_n| > \varepsilon\sqrt{m_n}\right\}\right].\end{aligned}$$

By the Cauchy-Schwarz inequality,

$$\left(\mathbb{E}\left[\left(X_1(\bar{\lambda}_n)-\alpha_n\right)^2 \mathbb{1}\left\{|X_1(\bar{\lambda}_n)-\alpha_n|>\varepsilon\sqrt{m_n}\right\}\right]\right)^2 \leq \mathbb{E}\left[\left(X_1(\bar{\lambda}_n)-\alpha_n\right)^4\right]\mathbb{P}\left(|X_1(\bar{\lambda}_n)-\alpha_n|>\varepsilon\sqrt{m_n}\right).$$

By Lemma 1, we have $\mathbb{E}[(X_1(\bar{\lambda}_n)-\alpha_n)^4]<\infty$. By Chebyshev's inequality,

$$\mathbb{P}\left(|X_1(\bar{\lambda}_n)-\alpha_n|>\varepsilon\sqrt{m_n}\right)\leq \frac{\mathbb{E}[(X_1(\bar{\lambda}_n)-\alpha_n)^2]}{\varepsilon^2 m_n}.$$

As shown in condition (i), by Lemma 1, the numerator in the bound is finite. Therefore, as $m_n\to\infty$, the bound tends to zero, as required.

As conditions (i) and (ii) hold, the Lindeberg Feller CLT applies to $\tilde{X}_j(c,\bar{\lambda}_n)$ and rearrangement results in (10). □

Let $\alpha'(c;\eta)$ denote the partial derivative of (1) with respect to the arrival rate parameter η,

$$\alpha'(c;\eta)=\frac{\partial\mathbb{E}[X(c;\eta)]}{\partial\eta}. \tag{12}$$

In the following analysis, we define $\alpha'(c;\lambda)$ and $\alpha'(c;\bar{\lambda}_n)$ to denote the partial derivative (12) evaluated at $\eta=\lambda$ and $\bar{\lambda}_n$ respectively. We make the following assumption about the partial derivative (12).

Assumption 4 For expected performance measure $\alpha(c;\eta)$, its partial derivative $\alpha'(c;\eta)$ exists for all $\eta>0$, and $\alpha'(c;\lambda)\neq 0$.

The streaming budget n is beyond the control of the analyst, so there are three possible cases, depending on the simulation budget m_n in relation to n. Intuitively, m_n should be of order at least n to induce a convergence rate of order $n^{-1/2}$, while if $m_n=o(n)$ then the Monte Carlo simulation error dominates the error originating from the streaming data. These ideas are formalized in the next result.

Theorem 2 Consider a problem instance satisfying Assumptions 1 — 4.

(i) If m_n is of order less than n, $m_n/n\to 0$ as $n\to\infty$, then,

$$\sqrt{m_n}(\bar{X}_n(c)-\alpha(c))\xrightarrow{d}\sigma(c)N(0,1),\quad \text{as } n\to\infty,$$

where $N(0,1)$ is a normally distributed random variable with mean zero and unit variance.

(ii) If m_n is of order bigger than n, $m_n/n\to\infty$ as $n\to\infty$, then,

$$\sqrt{n}(\bar{X}_n(c)-\alpha(c))\xrightarrow{d}\sqrt{\lambda/T}|\alpha'(c;\lambda)|N(0,1),\quad \text{as } n\to\infty.$$

(iii) If m_n is of order n, $m_n/n\to a$ as $n\to\infty$, for $a\in(0,\infty)$, then,

$$\sqrt{n}(\bar{X}_n(c)-\alpha(c))\xrightarrow{d}\sqrt{(\lambda/T)(\alpha'(c;\lambda))^2+a^{-1}\sigma^2(c)}N(0,1),\quad \text{as } n\to\infty.$$

Proof. By the Delta method CLT, since $\mathrm{Var}(\bar{\lambda}_n)=\lambda/(nT)$,

$$\sqrt{n}(\alpha_n(c)-\alpha(c))\xrightarrow{d}\sqrt{\lambda/T}|\alpha'(c;\lambda)|N(0,1),\quad \text{as } n\to\infty. \tag{13}$$

Therefore,

(i) Case $m_n/n \to 0$ as $n \to \infty$:

$$\sqrt{m_n}(\bar{X}_n(c) - \alpha(c)) = \underbrace{\sqrt{m_n}(\bar{X}_n(c) - \alpha_n(c))}_{\xrightarrow{d}\sigma(c)N(0,1) \text{ by Thm. 1}} + \underbrace{\sqrt{m_n/n}}_{\to 0} \underbrace{\sqrt{n}(\alpha_n(c) - \alpha(c))}_{\xrightarrow{d}\sqrt{\lambda/T}|\alpha'(c;\lambda)|N(0,1) \text{ by (13)}} .$$

The rightmost term in the RHS converges in distribution to 0 by Slutsky's Theorem. From there the result follows, also by Slutsky's Theorem.

(ii) Case $m_n/n \to \infty$ as $n \to \infty$:

$$\sqrt{n}(\bar{X}_n(c) - \alpha(c)) = \underbrace{\sqrt{n/m_n}}_{\to 0} \underbrace{\sqrt{m_n}(\bar{X}_n(c) - \alpha_n(c))}_{\xrightarrow{d}\sigma(c)N(0,1) \text{ by Thm. 1}} + \underbrace{\sqrt{n}(\alpha_n(c) - \alpha(c))}_{\xrightarrow{d}\sqrt{\lambda/T}|\alpha'(c;\lambda)|N(0,1) \text{ by (13)}} .$$

The proof argument is identical to the one of case (ii), and is thus omitted.

(iii) Case $m_n/n \to a$ as $n \to \infty$ (proof sketch due to space limitations): First of all, $\bar{X}_n(c) - \alpha_n(c)$ and $\bar{\lambda}_n$ are uncorrelated, and satisfy a multivariate Lindeberg-Feller CLT,

$$\sqrt{n}\begin{pmatrix} \bar{X}_n(c) - \alpha_n(c) \\ \bar{\lambda}_n - \lambda \end{pmatrix} \xrightarrow{d} N\left(0, \begin{pmatrix} \sigma^2(c)/a & 0 \\ 0 & \lambda/T \end{pmatrix}\right),$$

as $n \to \infty$. The multivariate Delta method then leads to the claimed result.

□

The choice of m_n is important for practitioners. When simulation is computationally cheap, selecting m_n with order greater than n can reduce variance constants. Conversely, for complex systems, the smallest m_n satisfying the theoretical convergence conditions is often preferable. The convergence rates for the different m_n regimes in Theorem 2 are validated in Section 5.

4.2 Probability of incorrect selection

In this section, we evaluate the performance of the algorithm by considering the limiting behavior of the probability of incorrect selection (PICS). Let $k = |\mathscr{C}|$ denote the cardinality of the finite decision set, and enumerate each of the elements as $c^{(1)}, \ldots, c^{(k)}$. Without loss of generality let $c^{(1)} = \text{argmin}_{c \in \mathscr{C}} f(c)$ denote the unique optimal number of servers. During period $n \in \mathbb{N}$, let $\mathbb{P}\left(c_n \neq c^{(1)}\right)$ denote the PICS. For each $i \in \{2, \ldots, k\}$, define the sub-optimality gap of action $c^{(i)}$ as follows

$$\Delta_i := f(c^{(i)}) - f(c^{(1)}) = \alpha(c^{(i)}; \lambda) + \beta(c^{(i)}) - \alpha(c^{(1)}; \lambda) - \beta(c^{(1)}),$$

with $\Delta_i > 0$ by the uniqueness of minimizer $c^{(1)}$.

Theorem 3 For a problem instance satisfying Assumptions 1 — 4, the nth period selected action of the greedy policy, c_n, converges in probability to the optimal action, $c^{(1)}$, i.e.,

$$\lim_{n \to \infty} \mathbb{P}\left(c_n \neq c^{(1)}\right) = 0.$$

Proof of Theorem 3. From the definition of c_n given in equation (6), the PICS, $\mathbb{P}(c_n \neq c^{(1)})$ corresponds to the probability that for any $i \neq 1$, $\bar{f}_n(c^{(i)}) < \bar{f}_n(c^{(1)})$.

By a union bound

$$\mathbb{P}\left(c_n \neq c^{(1)}\right) = \mathbb{P}\left(\bigcup_{i=2}^{k} \bar{f}_n(c^{(i)}) < \bar{f}_n(c^{(1)})\right) \leq (k-1) \max_{i \in \{2,\ldots,k\}} \mathbb{P}\left(\bar{f}_n(c^{(i)}) < \bar{f}_n(c^{(1)})\right).$$

Therefore, we must show for $i \neq 1$, $\mathbb{P}\left(\bar{f}_n(c^{(i)}) < \bar{f}_n(c^{(1)})\right) \to 0$.

Since

$$\begin{aligned}\{\bar{f}_n(c^{(1)}) - \bar{f}_n(c^{(i)}) > 0\} &= \{\bar{X}_n(c^{(1)}) + \beta(c^{(1)}) - \bar{X}_n(c^{(i)}) - \beta(c^{(i)}) > 0\} \\ &= \{\bar{X}_n(c^{(1)}) - \alpha(c^{(1)}) - \bar{X}_n(c^{(i)}) + \alpha(c^{(i)}) > \Delta_i\} \\ &\subseteq \{\bar{X}_n(c^{(1)}) - \alpha(c^{(1)}) > \Delta_i/2\} \cup \{\bar{X}_n(c^{(i)}) - \alpha(c^{(i)}) < -\Delta_i/2\},\end{aligned}$$

we obtain

$$\begin{aligned}\mathbb{P}\left(\bar{f}_n(c^{(i)}) < \bar{f}_n(c^{(1)})\right) &\leq \mathbb{P}\left(\bar{X}_n(c^{(1)}) - \alpha(c^{(1)}) > \Delta_i/2\right) + \mathbb{P}\left(\bar{X}_n(c^{(i)}) - \alpha(c^{(i)}) < -\Delta_i/2\right) \\ &\leq \mathbb{P}\left(\bar{X}_n(c^{(1)}) - \alpha_n(c^{(1)}) > \Delta_i/4\right) + \mathbb{P}\left(\bar{X}_n(c^{(i)}) - \alpha_n(c^{(i)}) < -\Delta_i/4\right) \\ &\quad + \mathbb{P}\left(\alpha_n(c^{(1)}) - \alpha(c^{(1)}) > \Delta_i/4\right) + \mathbb{P}\left(\alpha_n(c^{(i)}) - \alpha(c^{(i)}) < -\Delta_i/4\right).\end{aligned}$$

By Chebyshev's inequality,

$$\mathbb{P}\left(\bar{X}_n(c^{(1)}) - \alpha_n(c^{(1)}) > \Delta_i/4\right) + \mathbb{P}\left(\bar{X}_n(c^{(i)}) - \alpha_n(c^{(i)}) < -\Delta_i/4\right) \leq \frac{\mathrm{Var}(\bar{X}_n(c^{(1)}))}{(\Delta_i/4)^2} + \frac{\mathrm{Var}(\bar{X}_n(c^{(i)}))}{(\Delta_i/4)^2},$$

and condition (i) in the proof of Theorem 1 means that $\mathrm{Var}(\bar{X}_n(c^{(1)}))$ and $\mathrm{Var}(\bar{X}_n(c^{(i)}))$ are of order $O(1/m_n)$. Lastly, $\mathbb{P}(\alpha_n(c^{(1)}) - \alpha(c^{(1)}) > \Delta_i/4) + \mathbb{P}(\alpha_n(c^{(i)}) - \alpha(c^{(i)}) < -\Delta_i/4) \to 0$ exponentially fast in n by a large deviations contraction principle argument, since $\alpha(c;\eta)$ is continuous by Assumption 4. □

5 EMPIRICAL STUDY

In this section, we evaluate the empirical performance of Algorithm 1 on a representative problem instance. Let the expected performance measure, $\alpha(c;\lambda)$, be the expected mean length of the queue on the interval $[0,T]$. We verify that such a measure satisfies Assumption 1 as the expected mean queue length over a finite time horizon is finite almost surely for any finite arrival rate and number of servers. In addition, Assumption 2 is satisfied with $q = 1$ yielding a trivial upper bound. The following lemma, whose proof is given in Appendix A, ensures that Assumption 4 also holds.

Lemma 2 For $\alpha(c;\eta)$ defined as the expected mean length of the queue, $\alpha(c;\eta)$ satisfies Assumption 4.

Algorithm 1 is designed for queueing systems with generic service distributions. Here, we consider an $M/M/c$ problem instance as it allows us to obtain an accurate numerical approximation for $\alpha(c;\lambda)$ which we use as the ground truth in our numerical implementation. We discretize the interval $[0,T]$ into steps of size $\Delta t = 10^{-4}$. Given the system is empty at time 0, we use Euler's method to numerically solve the Kolmogorov forwards equations for an $M/M/c$ queue. Given these approximate transition probabilities, we compute the expected queue length for all discrete $t \in [0,T]$, which we then use to compute the expected mean length of the queue over the whole interval. Repeating this process for all $c \in \mathscr{C}$, we obtain the ground truth estimate for the expected objective given in (2).

To verify the theoretical results derived in Theorems 2 and 3, we evaluate algorithm performance using the empirical PICS. Accounting for the three regimes presented in Theorem 2, we consider $m_n = O(\sqrt{n})$, $m_n = O(n)$ and $m_n = O(n\log(n))$. In addition, we consider the expected cumulative regret of the policy as another means of evaluating algorithm performance.

Let $r_n = \mathbb{E}[f(c_n) - f(c^{(1)})]$ denote the expected instantaneous regret of the policy during period $n \in \mathbb{N}$. This corresponds to the expected loss incurred by choosing sub-optimal action $c_n \neq c^{(1)}$. Using the theoretical bounds on the PICS given in Theorem 3, a conservative bound on the cumulative regret for the sub-linear regime is given by

$$R_n := \sum_{j=1}^{n} r_j \leq \int_1^n r_j dj = O\left(\int_1^n j^{-1/2} dj\right) = O(n^{1/2}). \tag{14}$$

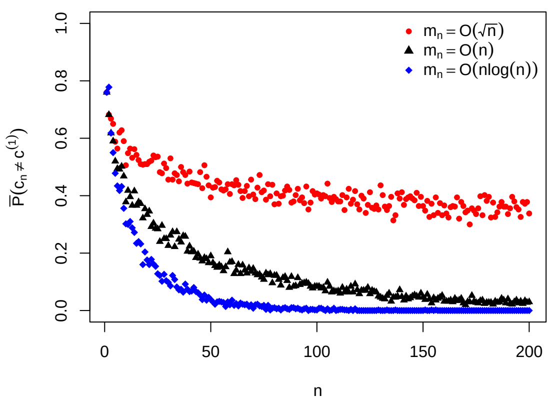
Similarly, for the linear and super-linear regimes, we have

$$R_n := \sum_{j=1}^{n} r_j \leq \int_1^n r_j dj = O\left(\int_1^n j^{-1} dj\right) = O(\log(n)). \tag{15}$$

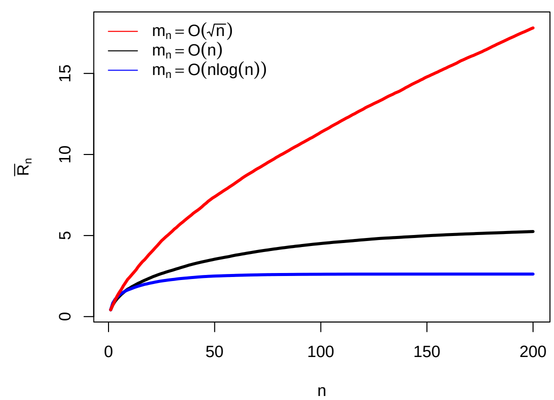
Within the numerical implementation, we show the empirical mean cumulative regret for each of the regimes satisfy the above bounds.

We construct a numerical experiment considering the following problem instance. Customers arrive to the system according to a homogeneous Poisson process with unknown rate $\lambda = 3$ on the interval $[0, T]$, with $T = 10$. Service times are exponentially distributed with mean $\mu = 1$ and we consider the decision set for the possible number of servers $\mathscr{C} = \{1, \ldots, 10\}$. Finally, we consider the cost function in the objective $\beta(c) = c/3$. We implement the greedy policy over $n = 200$ periods and carry out a total of 500 independent macro replications of the experiment to evaluate its performance.

Figure 1a shows the mean PICS, while Figure 1b shows the mean cumulative regret, both computed over 500 macro replications. In both plots, we consider the three regimes for m_n described in Theorem 2: red corresponds to a sub-linear regime with $m_n = O(\sqrt{n})$, black to a linear regime with $m_n = O(n)$ and blue to a super-linear regime with $m_n = O(n\log(n))$. For the cumulative regret plot we do not include confidence intervals as they are negligible.



(a) Mean PICS for the three regimes of the greedy policy over 500 macro replications.



(b) Mean cumulative regret for the three regimes of the greedy policy over 500 macro replications.

Figure 1

Within Figure 1a, we observe that the rate of decay for the mean PICS for each regime satisfies the bounds presented in Theorem 3. For the sub-linear regime, we observe that the PICS decays order $O(\max\{n^{-1}, m_n^{-1}\}) = O(n^{-1/2})$. Whilst for the linear and super-linear regimes, their decay is bounded above by $O(\max\{n^{-1}, m_n^{-1}\}) = O(n^{-1})$. The cumulative regret plots shown in Figure 1b also support the findings presented above. In addition, taking a log-log transform of the sub-linear regime cumulative regret shows that the $m_n = O(\sqrt{n})$ regime satisfies the theoretical bound given in (14). Equivalently, log-transforming the period n shows that the cumulative regret for linear and super-linear regimes satisfy the bound given in (15).

6 CONCLUSION

In this paper, we considered the problem of adaptively selecting the optimal number of servers to implement in an $M/G/c$ queueing system with streaming observational data. We demonstrated the effectiveness of a simple greedy decision-making policy that carries Monte Carlo estimation using repeated stochastic simulation. We derived a CLT for the Monte Carlo estimate with rates dependent on the simulation budget allocation. Using this, we derived a theoretical bound on the asymptotic PICS for the policy. These results were supported in a finite-time implementation of the policy. The empirical performance of the policy was evaluated both in terms of the PICS, as well as the expected mean cumulative regret.

The theoretical results within this paper lay the foundations for understanding the behavior and limitations of simulation-based decision-making algorithms in more-complex streaming observation settings. Several natural problem extensions remain. First, we assume a finite decision space, $\mathscr{C}$, whereas, in reality we often face problems with unbounded, countably infinite state spaces. Extending the analysis to these settings is a natural next step. Second, we consider a problem in which the data generation process was decision-independent. In reality, we often find that system performance depends highly on practitioner decisions. In the forthcoming work, Lambert et al. (2025), we build upon the analysis presented in this paper, extending the results to endogenous problem settings. Potential decision-dependent problem instances include, but are not limited to, settings with censored streaming observations (Gibbons, Grant, and Szechtman 2023), heterogeneous customer behavior (Inoue, Ravner, and Mandjes 2023), and decision-dependent arrival processes (Lambert et al. 2025). Within these settings a greedy policy may no longer be optimal, instead necessitating algorithms that achieve a balance between exploration and exploitation.

ACKNOWLEDGMENTS

This paper is based on work completed while Robert Lambert was part of the EPSRC funded STOR-i Centre for Doctoral Training (EP/S022252/1).

A APPENDIX

Differentiability of expected mean queue length

Proof of Lemma 2. Let the process $Q(t;\eta,c)$ denote the length at time $t<\infty$ of an $M/G/c$ queue with empty initial conditions and arrival rate $\eta>0$. Let $\alpha(c;\eta)$ denote the expected mean length of the queue over the interval $[0,T]$ and express its partial derivative with respect to η as follows

$$\frac{\partial\alpha(c;\eta)}{\partial\eta}=\frac{\partial}{\partial\eta}\mathbb{E}\left[\frac{1}{T}\int_0^T Q(t;\eta,c)dt\right]=\frac{1}{T}\int_0^T\frac{\partial\mathbb{E}\left[Q(t;\eta,c)\right]}{\partial\eta}dt. \tag{16}$$

In (16), as $Q(t;\eta,c)\geq 0$ w.p. 1, by Tonelli's theorem we exchange the limit and the expectation. By Leibniz's integral rule, for $\mathbb{E}[Q(t;\eta,c)]$ that is continuous in η we can differentiate under the integral. We first show that $\mathbb{E}\left[Q(t;\eta,c)\right]$ is convex in η, satisfying the continuity condition for Leibniz's rule, before proving that $\mathbb{E}\left[Q(t;\eta,c)\right]$ is differentiable with respect to η.

Let $\eta,\delta>0$ and consider an $M/G/c$ queueing system with arrival rate $\eta+\delta$ and empty initial conditions. By interpreting the Poisson arrival stream with rate $\eta+\delta$ as a superposition of two independent streams with rates η and δ, we can apply a coupling argument to show that processing these streams separately results in sub-additive congestion and queue lengths, yielding the inequality

$$\mathbb{E}[Q(t;\eta+\delta,c)]\geq\mathbb{E}[Q(t;\eta,c)]+\mathbb{E}[Q(t;\delta,c)]. \tag{17}$$

Trivially, we note that $\mathbb{E}[Q(t;0,c)]=0$ for all $t<\infty$, as such, rearranging (17) yields the following inequality

$$\mathbb{E}[Q(t;\eta+\delta,c)]-\mathbb{E}[Q(t;\eta,c)]\geq\mathbb{E}[Q(t;\delta,c)]-\mathbb{E}[Q(t;0,c)]. \tag{18}$$

Therefore, $\mathbb{E}[Q(t;\eta,c)]$ is convex and increasing in η and hence continuous. In addition, it is semi-differentiable, admitting left and right derivatives, denoted by $\frac{\partial^-}{\partial\eta}\mathbb{E}[Q(t;\eta,c)]$ and $\frac{\partial^+}{\partial\eta}\mathbb{E}[Q(t;\eta,c)]$ respectively, for all $\eta>0$. To establish differentiability we must show that $\frac{\partial^-}{\partial\eta}\mathbb{E}[Q(t;\eta,c)]=\frac{\partial^+}{\partial\eta}\mathbb{E}[Q(t;\eta,c)]$ for all $\eta>0$, which we demonstrate using the following sample path approach.

Let $k, \eta > 0$, and let $(a_1, \dots, a_{n_\omega})$ denote a sequence of arrival times sampled from a Poisson process with rate $\eta + k$, restricted to the interval $[0,t]$, where n_ω is the random number of arrivals observed by time t. Let $(s_1, \dots, s_{n_\omega})$ denote a sequence of i.i.d. service times drawn from the known service distribution. Let $\omega = \{(a_i, s_i)\}_{i=1}^{n_\omega}$ represent the set of customer arrival-service pairs for a single random trajectory, and let Ω denote the set of all such sets of tuples. Given $\omega \in \Omega$, we can construct a sample path for the queue length process of an $M/G/c$ queue, denoted by $\{q(\tau; \omega, c); \tau \in [0,t]\}$. We note that unlike Q, q is deterministic, i.e., given the same set of arrival-service pairs, ω, we obtain the same sample path on $[0,t]$.

To construct a sample path corresponding to an $M/G/c$ system with arrival rate $\eta + h$ for some $0 < h < \min\{k, \eta\}$, we apply a thinning procedure. Let $\omega_1 \subseteq \omega$ denote the set of arrival-service pairs (a_i, s_i) accepted, given each pair has independent probability of acceptance $(\eta + h)/(\eta + k)$ and let $\{q(\tau; \omega_1, c); \tau \in [0,t]\}$ denote the sample path constructed using ω_1. This thinning procedure can be repeated two further times, to construct paths corresponding to systems with arrival rates η and $\eta - h$ respectively. First, let $\omega_2 \subseteq \omega_1$ denote the arrival-service pairs obtained after thinning ω_1 with acceptance probability $\eta/(\eta + h)$, with corresponding sample path $\{q(\tau; \omega_2, c); \tau \in [0,t]\}$. Similarly, let $\omega_3 \subseteq \omega_2$ denote the arrival-service pairs obtained after thinning ω_2 with acceptance probability $(\eta - h)/\eta$, with corresponding sample path $\{q(\tau; \omega_3, c); \tau \in [0,t]\}$.

By the total law of expectation, we can express the right and left derivatives of the expected queue length at time t as

$$\frac{\partial^+}{\partial \eta} \mathbb{E}[Q(t; \eta, c)] = \lim_{h \to 0} \mathbb{E}_\omega \left[\frac{\mathbb{E}_{\omega_1, \omega_2}[q(t; \omega_1, c) - q(t; \omega_2, c) \mid \omega]}{h} \right], \tag{19}$$

$$\frac{\partial^-}{\partial \eta} \mathbb{E}[Q(t; \eta, c)] = \lim_{h \to 0} \mathbb{E}_\omega \left[\frac{\mathbb{E}_{\omega_2, \omega_3}[q(t; \omega_2, c) - q(t; \omega_3, c) \mid \omega]}{h} \right], \tag{20}$$

where, for instance, $\mathbb{E}_{\omega_1, \omega_2}[q(t; \omega_1, c) - q(t; \omega_2, c) \mid \omega]$ denotes the expected difference in the queue length at time t given initial arrival-service pairs $\omega \in \Omega$ and thinned subsets ω_1, ω_2. The conditional expectation is taken over all possible subsets $\omega_2 \subseteq \omega_1 \subseteq \omega$ generated via the thinning procedure.

Let us consider (19). Given $\omega \in \Omega$, we let the random variable $D_{1,2} \mid \omega$ denote the difference in the cardinalities of thinned subsets ω_1 and ω_2, i.e., $|\omega_1| - |\omega_2|$ given $|\omega| = n_\omega$. By the total law of probability, we have

$$\mathbb{E}_{\omega_1, \omega_2}[q(t; \omega_1, c) - q(t; \omega_2, c) \mid \omega] = \sum_{d=0}^{n_\omega} \mathbb{E}_{\omega_1, \omega_2}[q(t; \omega_1, c) - q(t; \omega_2, c) \mid D_{1,2} = d, \omega] \mathbb{P}(D_{1,2} = d \mid \omega). \tag{21}$$

As $\omega_2 \subseteq \omega_1 \subseteq \omega$, the event $D_{1,2} = d \mid \omega$ corresponds to the event where d of the customers accepted during the first thinning step are rejected during the second thinning step. As such, $D_{1,2} \mid \omega \sim \text{Binom}(n_\omega, p)$ where $p = \frac{\eta+h}{\eta+k}\left(1 - \frac{\eta}{\eta+h}\right) = \frac{h}{\eta+k}$. We can evaluate the conditional expectation in (21) as follows. For each $d \in \{0, \dots, n_\omega\}$, let $\mu(t \mid D_{1,2} = d, \omega)$ denote the mean change in the queue length at time t, i.e., the mean value of $q(t; \omega_1, c) - q(t; \omega_2, c)$ taken over all pairs (ω_1, ω_2) satisfying $D_{1,2} = d$. We note that this is obtained combinatorially considering the initial n_ω arrival-service pairs and is therefore independent of h.

Using the above, we obtain the following expression for the right derivative (19).

$$\lim_{h \to 0} \mathbb{E}_\omega \left[\frac{\mathbb{E}_{\omega_1, \omega_2}[q(t; \omega_1, c) - q(t; \omega_2, c) \mid \omega]}{h} \right] = \mathbb{E}_\omega \left[\lim_{h \to 0} \frac{\sum_{d=0}^{n_\omega} \mu(t \mid D_{1,2} = d, \omega) \mathbb{P}(D_{1,2} = d \mid \omega).}{h} \right] \tag{22}$$

$$= \mathbb{E}_\omega \left[\mu(t \mid D_{1,2} = 1, \omega) \frac{n_\omega}{\eta + k} \right]. \tag{23}$$

In line (22) we exchange limit and expectation using a dominated convergence argument, as $\mathbb{E}_{\omega_1, \omega_2}[q(t; \omega_1, c) - q(t; \omega_2, c) \mid \omega] \le n_\omega$ and $\mathbb{E}_\omega[n_\omega] = (\eta + k)t < \infty$ for finite t. In line (23), we note that for $d = 0$,

$\mu(t \mid D_{1,2} = d, \omega) = 0$ by definition. In addition, for $p = h/(\eta + k)$, $\mathbb{P}(D_{1,2} = d \mid \omega) = o(h)$ for $d > 1$. This means that the summation term corresponding to $D_{1,2} = 1$ is the only one that does not vanish in the limit. We construct a similar argument for the left derivative, given in (20). The set-up is identical to that of the right derivative, however, we let $D_{2,3} \mid \omega$ denote the difference in the cardinalities of thinned subsets ω_2 and ω_3. As $\omega_3 \subseteq \omega_2 \subseteq \omega_1 \subseteq \omega$, this corresponds to the event that d of the customers accepted during the first two thinning steps were rejected during the third. We have $D_{2,3} \mid \omega \sim \text{Binom}(n_\omega, q)$, where $q = \frac{\eta+h}{\eta+k}\frac{\eta}{\eta+h}\left(1 - \frac{\eta-h}{\eta}\right) = \frac{h}{\eta+k}$. As $q = p$ and $\mu(t \mid D_{2,3} = d, \omega)$ is calculated combinatorially in the same way as $\mu(t \mid D_{1,2} = d, \omega)$, the remainder of the proof is identical. Therefore, as the left and right partial derivatives match, the partial derivative (16) with respect to η, exists and is positive for all $\eta > 0$, hence satisfying Assumption 4.

□

REFERENCES

Durrett, R. 2019. *Probability: Theory and Examples*, Volume 49. Cambridge University Press.

Gibbons, C., J. A. Grant, and R. Szechtman. 2023. "Sequential Simulation Optimization with Censoring: An Application to Bike Sharing Systems". In *2023 Winter Simulation Conference (WSC)*, 3624–3635. IEEE.

He, L., U. V. Shanbhag, and E. Song. 2024. "Stochastic Approximation for Multi-period Simulation Optimization with Streaming Input Data". *ACM Transactions on Modeling and Computer Simulation* 34(2):1–27.

He, L., and E. Song. 2024. "Introductory Tutorial: Simulation Optimization Under Input Uncertainty". In *2024 Winter Simulation Conference (WSC)*, 1338–1352. IEEE.

Inoue, Y., L. Ravner, and M. Mandjes. 2023. "Estimating customer impatience in a service system with unobserved balking". *Stochastic Systems* 13(2):181–210.

Lambert, R. J., J. A. Grant, R. Shone, and R. Szechtman. 2025. "Optimization of stochastic systems with decision-dependent arrival processes using streaming simulation". *Working Paper*.

Lindeberg, J. W. 1922. "Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung". *Mathematische Zeitschrift* 15(1):211–225.

Song, E., and U. V. Shanbhag. 2019. "Stochastic approximation for simulation optimization under input uncertainty with streaming data". In *2019 Winter Simulation Conference (WSC)*, 3597–3608. IEEE.

Wu, D., Y. Wang, and E. Zhou. 2024. "Data-driven ranking and selection under input uncertainty". *Operations Research* 72(2):781–795.

Wu, D., and E. Zhou. 2019. "Fixed confidence ranking and selection under input uncertainty". In *2019 Winter Simulation Conference (WSC)*, 3717–3727. IEEE.

AUTHOR BIOGRAPHIES

ROBERT J LAMBERT is a Ph.D. student of the Statistics and Operational Research with Industry Centre for Doctoral Training at Lancaster University. His research interests include applied probability, simulation and online learning. His email address is r.lambert1@lancaster.ac.uk.

JAMES A GRANT is a Lecturer (equiv. Assistant Professor) in the School of Mathematical Sciences at Lancaster University. His primary research interests are in reinforcement learning and multi-armed bandits. His email address is j.grant@lancaster.ac.uk and homepage is https://www.lancaster.ac.uk/staff/grantj/.

ROB SHONE is a Lecturer (equiv. Assistant Professor) in the Department of Management Science at Lancaster University. His research interests include stochastic modelling, queueing systems and reinforcement learning. His email address is r.shone@lancaster.ac.uk.

ROBERTO SZECHTMAN received his Ph.D. from Stanford University. Currently, he is Professor in the Operations Research Department at the Naval Postgraduate School. His research interests include applied probability and military operations research. His email address is rszechtm@nps.edu.