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.

WHEN MACHINE LEARNING MEETS IMPORTANCE SAMPLING: A MORE EFFICIENT RARE EVENT ESTIMATION APPROACH

Ruoning Zhao[1], and Xinyun Chen[1]

[1]School of Data Science, the Chinese University of Hong Kong, Shenzhen, GUANGDONG, CHINA

ABSTRACT

Driven by applications in telecommunication networks, we explore the simulation task of estimating rare event probabilities for tandem queues in their steady state. Existing literature has recognized that importance sampling methods can be inefficient, due to the exploding variance of the path-dependent likelihood functions. To mitigate this, we introduce a new importance sampling approach that utilizes a marginal likelihood ratio on the stationary distribution, effectively avoiding the issue of excessive variance. In addition, we design a machine learning algorithm to estimate this marginal likelihood ratio using importance sampling data. Numerical experiments indicate that our algorithm outperforms the classic importance sampling methods.

1 INTRODUCTION

In computing and telecommunication industries, service level agreements (SLA) in commercial contracts are aimed to ensure that the promised quality of service (QoS) is met under a given network design, see Cisco (2025) for an example of commercial SLA. A commonly used QoS performance metric in SLA takes the form of a guarantee on the tail probability of certain queueing function of the network in steady state (Milner and Olsen 2008), i.e.,

$$\mathbb{P}(X_\infty > \gamma) \leq p,$$

where the random variable X_∞ could be the steady-state total number of jobs in the system or a job's sojourn time, and p is a small number. Intuitively, this type of SLA aims to maintain the network congestion within an acceptable threshold with a high probability in the long run. Since obtaining an analytic expression for the stationary distribution is often not feasible for general stochastic network models, the industry generally relies on simulation techniques to numerically compute the probability $\mathbb{P}(X_\infty > \gamma)$, thereby assessing the SLA.

Since the value of p is typically very small, like $p = 10^{-5}$ (Harchol-Balter 2021), calculating the probability $\mathbb{P}(X_\infty > \gamma)$ essentially becomes a problem of rare event simulation. A common approach to improve the efficiency of rare event simulation is by importance sampling (IS). Nonetheless, estimating the stationary distribution X_∞ adds an additional level of complexity to the design of the IS algorithms. Earlier studies (Glasserman and Kou 1995; De Boer 2006) have demonstrated that, for such simulation tasks, finding an importance distribution that could effectively reduce the variance is quite challenging, even in the very basic case of two-station tandem queueing network.

In this paper, we highlight that these challenges largely originate from the path-dependent characteristics of the classic IS algorithms, which use the regenerative structures of queueing processes to deal with the stationary distribution. To tackle this issue, we introduce a novel IS algorithm, demonstrated through the two-station tandem queueing network example. This algorithm applies IS directly on the "marginal" stationary distribution of individual samples in the state space, rather than on the regenerative cycle sample paths. Given that the marginal likelihood ratio of stationary distributions is in general unknown for queueing networks, we use the off-policy evaluation method from reinforcement learning literature to approximate the stationary likelihood ratio from simulation data. Broadly speaking, our approach integrates traditional

IS with machine learning techniques. The numerical findings indicate that our algorithm outperforms the traditional regenerative IS method, achieving lower mean squared errors. Furthermore, the advantage of our algorithm is relatively robust with respect to the choice of importance distribution.

The rest of the paper is organized as follows. We provide a brief literature review on related topics in Section 2. In Section 3, we introduce the estimation objective in the tandem queues and discuss the drawbacks of the state-of-the-art IS method. In Section 4, we first explain the key components in algorithm design and then present the complete algorithm. Numerical results are reported in Section 5. Section 6 concludes the paper with insights into future research directions.

2 RELATED WORKS

The rare event simulation in queueing systems, like the response time violations, is typically addressed by IS methods, see Blanchet and Mandjes (2009) and the references therein. Parekh and Walrand (1989) is probably one of the first to propose IS methods with state-independent choices of importance distributions for single-station and tandem queues. Glasserman and Kou (1995) and De Boer (2006) further analyze the variance reduction performance of the Parekh-Walrand method, and prove that the state-independent IS method in certain tandem queues is not asymptotically optimal. As a result, there has been a growing body of research related to designing IS methods with state-dependent choices of importance distributions for queueing systems, including single-class $G/G/1$ queue (Blanchet et al. 2007), single-class Jackson network (Dupuis and Wang 2009), two-class $M/M/1$ queue (Setayeshgar and Wang 2011), and multi-class open/closed network (Dupuis and Wang 2007). In these works, the importance distributions are generally derived from large deviation principles as well as differential game approaches, which depend heavily on the specific system dynamics and are difficult to solve.

Another stream of related literature lives in the off-policy estimation of long-horizon average reward in the field of reinforcement learning (RL), see Xie et al. (2019), Uehara et al. (2022), and the references therein. The Most common set of off-policy estimation methods is derived from IS estimators. The likelihood weights are based on the product of the importance ratios of many steps in a trajectory (Liu et al. 2020), and thus variances in individual steps can accumulate multiplicatively. Liu et al. (2018) shows that applying importance weighting on the state space rather than the trajectory space can substantially reduce estimation variance. Another set of off-policy estimation methods first fits a parametric model to learn the environment dynamics using data collected under the behavior policy, and then use this model to simulate trajectories under the evaluation policy, see Fonteneau et al. (2013) and Chow et al. (2015). Kallus and Uehara (2022) proposes an estimation method that learns the environment dynamics and stationary likelihood ratios simultaneously, and shows that their estimator remains efficient with long time horizon. Most off-policy estimation methods are focused on average reward, while our goal is to evaluate the tail probability.

3 PROBLEM SETUP

In the remaining part of the paper, we will illustrate the development of our proposed IS algorithm and evaluate its performance using a two-station tandem queueing network as our primary example. We emphasize here that extending our algorithm to other Markovian queueing networks is straightforward.

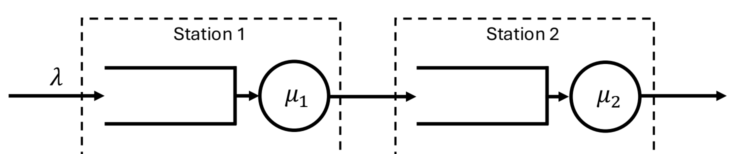

Figure 1: Structure of a two-station tandem queueing network.

3.1 Model & Objective

We consider a two-station tandem queueing network, whose structure is shown in Figure 1. Each job arrives according to a Poisson process with rate $\lambda \in \mathbb{R}_+$. Each station $i \in \{1,2\}$ has one single server which has i.i.d. exponential service times with rate $\mu_i \in \mathbb{R}_+$. Since our objective is to evaluate the QoS at steady state, we assume that the queueing network is stable by setting the arrival rate below the service rate in each station, i.e.,

$$\lambda < \mu_1, \quad \lambda < \mu_2. \tag{1}$$

We also assume that the server in station 1 has a larger service rate, i.e., $\mu_1 > \mu_2$, since finding a reliable importance distribution in this scenario is quite challenging as shown in Glasserman and Kou (1995) and De Boer (2006). Jobs are served in a first-come-first-serve manner. The arrival process and service times are assumed to be mutually independent. The system is empty initially.

To evaluate this queueing system's service level performance, we particularly focus on the total number of jobs in the entire network. In particular, the queueing dynamic with respect to the number X_t^i of jobs in station i at time t could be written as

$$X_t^1 = A_t - D_t^1,$$
$$X_t^2 = D_t^1 - D_t^2,$$

where A_t is the Poisson arrival process and D_t^i is the number of jobs leaving station i up to time t.

The process $\{(X_t^1, X_t^2)\}$ forms a continuous time Markov chain (CTMC). Under the stability assumption (1), this CTMC admits a unique stationary distribution π. Let $X = (X^1, X^2)$ be a random vector following the distribution π. Our objective is to estimate a classic QoS metric in queue systems: the steady-state queue length overflow probability (Whitt 1993),

$$p_\gamma \equiv \mathbb{P}(X^1 + X^2 \geq \gamma).$$

The event $\{X^1 + X^2 \geq \gamma\}$ becomes increasingly rare as the threshold γ grows large, necessitating an efficient estimation algorithm. Before formally introducing our algorithm, we first review the state-of-the-art IS method for stationary queue length overflow probabilities in tandem queueing networks.

3.2 Regenerative Importance Sampling Method

The state-of-the-art IS estimation for stationary tail probabilities relies on a regenerative simulation, see Asmussen and Glynn (2007) and Guang et al. (2022). For simplicity, we work with the uniformized discrete time Markov chains (DTMC) derived from the original tandem queueing network. We slightly abuse the notation by letting $\{(X_t^1, X_t^2)\}$ represent the state of this uniformized DTMC. The transition kernel (probability) P of this uniformized DTMC is summarized in the following equation,

$$P(X_{t+1} \mid X_t) = \begin{cases} \lambda/(\lambda+\mu_1+\mu_2), & \text{if } X_{t+1}^1 = X_t^1 + 1,\ X_{t+1}^2 = X_t^2, \\ \mu_1/(\lambda+\mu_1+\mu_2), & \text{if } X_{t+1}^1 = X_t^1 - 1,\ X_{t+1}^2 = X_t^2 + 1, \\ \mu_2/(\lambda+\mu_1+\mu_2), & \text{if } X_{t+1}^1 = X_t^1,\ X_{t+1}^2 = X_t^2 - 1, \\ 1 - \sum_{\mathbf{x} \neq X_t} P(X_{t+1} = \mathbf{x} \mid X_t), & \text{if } X_{t+1} = X_t. \end{cases}$$

This uniformized DTMC shares the same stationary distribution π as the original tandem queueing network (Ross 2014). The dynamics of this DTMC $\{(X_t^1, X_t^2)\}$ can be viewed as a regenerative process, where the system regenerates whenever a job finds the system is empty upon departure. Let α denote the cycle length in units of time for a regenerative cycle. By the renewal reward theorem (Crane and Lemoine 1977), the stationary tail probability can be expressed as

$$\mathbb{P}(X^1 + X^2 \geq \gamma) = \frac{\mathbb{E}\left[\sum_{t=1}^{\alpha} \mathbf{1}\{X_t^1 + X_t^2 \geq \gamma\}\right]}{\mathbb{E}[\alpha]}, \tag{2}$$

where the numerator is the expected number of the target events observed in a regenerative cycle, and the denominator is the the expected length of a regenerative cycle. Thus, estimating this probability reduces to separately estimating the denominator and the numerator.

We first describe the IS estimation of the numerator in (2) using data from an alternative system. In each regenerative cycle, the alternative system first follows the dynamic of the uniformized DTMC derived from a modified tandem queueing network with exponential arrival rate $\tilde{\lambda}$ and exponential service rates $\tilde{\mu}_i$. This modified system is intentionally unstable with

$$\tilde{\lambda} \geq \tilde{\mu}_1 \text{ or } \tilde{\lambda} \geq \tilde{\mu}_2,$$

so that the target event $\{X_t^1 + X_t^2 \geq \gamma\}$ can be observed frequently. Once the total number of jobs in the system reaches γ at time $\tilde{\tau}$, the dynamic of the system is switched back to the uniformized DTMC derived from the original tandem queueing network. Let $\tilde{X}_t^i$ denote the number of jobs in station i at time t in the alternative system, and $\tilde{\alpha}$ be the regenerative cycle length. Then, we have the following equality,

$$\mathbb{E}\left[\sum_{t=1}^{\alpha} \mathbf{1}\{X_t^1 + X_t^2 \geq \gamma\}\right] = \mathbb{E}\left[\left(\sum_{t=1}^{\tilde{\alpha}} \mathbf{1}\{\tilde{X}_t^1 + \tilde{X}_t^2 \geq \gamma\}\right) \cdot \left(\prod_{t=1}^{\tilde{\tau}} \frac{P(\tilde{X}_t \mid \tilde{X}_{t-1})}{Q(\tilde{X}_t \mid \tilde{X}_{t-1})}\right)\right],$$

where P and Q are the transition kernels of the uniformized DTMC derived from the original and alternative queueing networks.

Since the denominator in (2) does not involve a rare event, it is directly estimated using data from the original system. Let us independently generate m_1 and m_2 regenerative cycles of data from the alternative and original systems respectively. Using an extra subscript i to denote the ith cycle, the state-of-the-art IS estimator is given by

$$\hat{\mathbf{P}}(X^1 + X^2 \geq \gamma) = \frac{m_1^{-1} \sum_{i=1}^{m_1} \left[\left(\sum_{t=1}^{\tilde{\alpha}_i} \mathbf{1}\{\tilde{X}_{t,i}^1 + \tilde{X}_{t,i}^2 \geq \gamma\}\right) \cdot \left(\prod_{t=1}^{\tilde{\tau}_i} \frac{P(\tilde{X}_{t,i} | \tilde{X}_{t-1,i})}{Q(\tilde{X}_{t,i} | \tilde{X}_{t-1,i})}\right)\right]}{m_2^{-1} \sum_{i=1}^{m_2} \alpha_i}.$$

The performance of this IS estimator hinges on the choice of parameters $(\tilde{\lambda}, \tilde{\mu}_1, \tilde{\mu}_2)$ for the alternative system. De Boer (2006) proves that letting $(\tilde{\lambda}, \tilde{\mu}_1, \tilde{\mu}_2) = (\mu_2, \mu_1, \lambda)$ is the only possible choice so that the IS estimator can use the orderwisely smallest amount of data to achieve the same level of variance, which is the so-called *asymptotically efficient* estimator. However, as implied in Glasserman and Kou (1995), this IS estimator with any choice of parameters $(\tilde{\lambda}, \tilde{\mu}_1, \tilde{\mu}_2)$ is not asymptotically efficient when the original system parameters satisfy

$$\frac{\mu_2(\mu_1 + \mu_2)}{(\lambda + \mu_1)^2} > 1. \tag{3}$$

The IS estimator suffers from the excessive variance introduced by the path-dependent likelihood ratio $\prod_{t=1}^{\tilde{\tau}_i} \frac{P(\tilde{X}_{t,i} | \tilde{X}_{t-1,i})}{Q(\tilde{X}_{t,i} | \tilde{X}_{t-1,i})}$. The main reason is that there are a wide range of sample paths in a regenerative cycle that can reach the target event $\{X_t^1 + X_t^2 \geq \gamma\}$, see Figure 2, and using a certain alternative system with parameters $(\tilde{\lambda}, \tilde{\mu}_1, \tilde{\mu}_2)$ may not reduce the variance contributed by the path-dependent likelihood ratio on every sample path. For instance, using the alternative system with parameters $(\tilde{\lambda}, \tilde{\mu}_1, \tilde{\mu}_2) = (\mu_2, \mu_1, \lambda)$ could reduce the variance contributed by the path-dependent likelihood ratio on the blue path in Figure 2, however, the likelihood ratio on the red path will then introduce an explosive amount of variance to the IS estimator (Glasserman and Kou 1995). This limitation motivates us to propose a robust IS method that can achieve variance reduction with a wider choice of alternative systems.

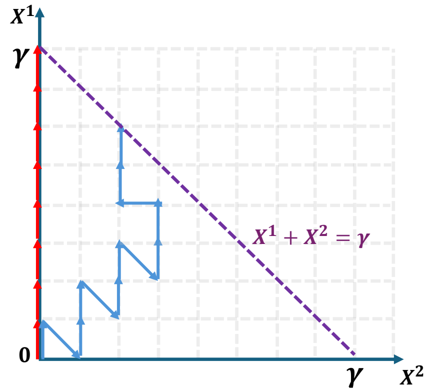

Figure 2: Two representative sample paths reaching the target event $\{X^1 + X^2 \geq \gamma\}$ from the empty state.

4 IMPORTANCE SAMPLING WITH MARGINAL STATIONARY DISTRIBUTION

Inspired by Liu et al. (2018), we intend to apply the IS method directly on the "marginal" stationary distribution of each sample data. Specifically, let us consider an alternative two-station tandem queueing network with exponential arrival rate $\tilde{\lambda}$ and exponential service rates $\tilde{\mu}_i$, where the parameters satisfy the stability condition

$$\tilde{\lambda} < \tilde{\mu}_1, \quad \tilde{\lambda} < \tilde{\mu}_2.$$

This queueing network has a stationary distribution $\tilde{\pi}$. Let $\tilde{X} = (\tilde{X}^1, \tilde{X}^2)$ be a random vector following the distribution $\tilde{\pi}$. The steady-state queue length overflow probability for the original queueing network can then be expressed as

$$\begin{aligned}
\mathbb{P}(X^1 + X^2 \geq \gamma) &= \sum_{x_1, x_2} \mathbf{1}\{x_1 + x_2 \geq \gamma\} \cdot \pi(x_1, x_2), \\
&= \sum_{x_1, x_2} \mathbf{1}\{x_1 + x_2 \geq \gamma\} \frac{\pi(x_1, x_2)}{\tilde{\pi}(x_1, x_2)} \cdot \tilde{\pi}(x_1, x_2), \\
&= \mathbb{E}_{\tilde{X} \sim \tilde{\pi}} \left[\mathbf{1}\{\tilde{X}^1 + \tilde{X}^2 \geq \gamma\} \frac{\pi(\tilde{X})}{\tilde{\pi}(\tilde{X})} \right].
\end{aligned}$$

It is noted that $\mathbf{1}\{\tilde{X}^1 + \tilde{X}^2 \geq \gamma\} \frac{\pi(\tilde{X})}{\tilde{\pi}(\tilde{X})}$ is a promising choice of the IS estimator, as it avoids the excessive variance introduced by the likelihood ratio that depends on the whole sample path. However, a critical challenge occurs: the stationary distributions π and $\tilde{\pi}$ are typically unknown for complex queueing networks. Therefore, a vital step in our proposed algorithm is to learn the stationary likelihood ratio $\pi/\tilde{\pi}$, which is illustrated in detail in the follow section.

4.1 Learning Stationary Likelihood Ratio $\pi/\tilde{\pi}$

Let P and Q be the transition kernels of DTMC derived from the original and alternative tandem queueing networks. Suppose X and $\tilde{X}$ are sampled from π and $\tilde{\pi}$, respectively, and $\tilde{X}'$ is sampled from $Q(\cdot \mid \tilde{X})$, which also follows the distribution $\tilde{\pi}$. By the definition of Markov chain's transition kernel and stationary distribution, we can derive the following equality for the expectation of $f(X)$ for any function f:

$$\mathbb{E}_{X \sim \pi}[f(X)] = \mathbb{E}\left[f(\tilde{X}') \cdot \frac{\pi(\tilde{X}')}{\tilde{\pi}(\tilde{X}')} \right] = \mathbb{E}\left[f(\tilde{X}') \cdot \frac{\pi(\tilde{X})}{\tilde{\pi}(\tilde{X})} \cdot \frac{P(\tilde{X}'|\tilde{X})}{Q(\tilde{X}'|\tilde{X})} \right].$$

This observation can be leveraged to obtain the following key property of the stationary likelihood ratio $\pi/\tilde{\pi}$.

Proposition 1 (Liu et al. 2018) A function $w(x)$ equals $\pi(x)/\tilde{\pi}(x)$ (up to a constant factor) if and only if it satisfies

$$\mathbb{E}[f(\tilde{X}')\Delta(w;\tilde{X},\tilde{X}')] = 0, \quad \text{for any function } f,$$
$$\text{with} \quad \Delta(w;\tilde{X},\tilde{X}') \equiv w(\tilde{X}) \cdot \frac{P(\tilde{X}' \mid \tilde{X})}{Q(\tilde{X}' \mid \tilde{X})} - w(\tilde{X}'). \tag{4}$$

Following the approach in Liu et al. (2018), we can estimate $\pi/\tilde{\pi}$ by solving the following min-max problem:

$$\min_{w \neq 0} L(w) \equiv \max_{f \in \mathscr{F}} \mathbb{E}[f(\tilde{X}')\Delta(w;\tilde{X},\tilde{X}')]^2,$$

where the condition $w \neq 0$ is added to avoid the trivial solution $w \equiv 0$, and $\mathscr{F}$ is a set of test functions that should be rich enough to identify w. Let $\mathscr{H}$ be a reproducing kernel Hilbert space (RKHS) of functions with a positive definite kernel $k(r,\bar{r})$. A good choice of $\mathscr{F}$ is the unit ball of the RKHS $\mathscr{H}$, i.e., $\mathscr{F} \equiv \{f \in \mathscr{H} : ||f||_{\mathscr{H}} \leq 1\}$, as it yields a closed form representation of the loss function $L(w)$:

$$L(w) = \mathbb{E}[\Delta(w;\tilde{X}_a,\tilde{X}'_a)\Delta(w;\tilde{X}_b,\tilde{X}'_b)k(\tilde{X}'_a,\tilde{X}'_b)].$$

Here, $(\tilde{X}_a,\tilde{X}'_a)$ and $(\tilde{X}_b,\tilde{X}'_b)$ are independent transition pairs, where $\tilde{X}_i$ and $\tilde{X}'_i$ follow $\tilde{\pi}$ and $Q(\cdot \mid \tilde{X}_i)$ respectively, for $i \in \{a,b\}$. See Berlinet and Thomas-Agnan (2011) for additional background on RKHS.

4.2 A Machine Learning Based Importance Sampling Algorithm

According to Proposition 1, a function w is proportional to the stationary likelihood ratio $\pi/\tilde{\pi}$ if and only if the loss function $L(w) = 0$. Following this idea, we parameterize $w(x) = w_\theta(x)$ as a neural network. This neural network is then trained by minimizing the loss function $L(w_\theta)$ using a stochastic gradient method such as Adam.

In practice, we design a robust loss function $L(w)$ for a better performance of our algorithm, namely,

$$L(w) = \mathbb{E}\left[\mathbb{E}[\Delta(w;\tilde{X}_a,\tilde{X}'_a)\Delta(w;\tilde{X}_b,\tilde{X}'_b)k(\tilde{X}'_a,\tilde{X}'_b) \mid k = k_i]\right] + \frac{\alpha}{2}(\mathbb{E}[w(\tilde{X}_a)] - 1)^2.$$

where $(\tilde{X}_a,\tilde{X}'_a)$ and $(\tilde{X}_b,\tilde{X}'_b)$ are independent transition pairs, $\tilde{X}_i$ and $\tilde{X}'_i$ follow $\tilde{\pi}$ and $Q(\cdot \mid \tilde{X}_i)$ respectively, for $i \in \{a,b\}$, and the regularization parameter $\alpha > 0$. In detail, we enlarge the set of test function by assuming that the test function has an equal possibility to lay in a range of different reproducing kernel Hilbert spaces with kernel functions k_i. The extra regularization term is added in $L(w)$ to prevent the neural network w_θ from converging to the trivial estimation $w_\theta \equiv 0$.

In conclusion, we summarize our Machine Learning based steady state Importance Sampling algorithm (MLIS) in Algorithm 1.

Algorithm 1 Machine Learning Based Importance Sampling (MLIS)

Input: Simulation data $\mathscr{D} = \{X_t\}_{t=1}^{T}$ of the uniformized DTMC derived from the alternative system. Kernel function set $\{k_i(\mathbf{x},\mathbf{y})\}$. Regularization parameter α. Learning rate ε. Batch size b_D and b_k.

Initiate the density ratio $w(r) = w_\theta(r)$ to be a neural network parameterized by θ.

for iteration $= 1, 2, \ldots, N$ **do**

Randomly choose a batch $\mathscr{M}_D$ of size b_D from the data $\mathscr{D}$, i.e., $\mathscr{M} \subset \{1, \ldots, T-1\}$.

Randomly choose a batch $\mathscr{M}_k$ of size b_k from the kernel function set $\{k_i(\mathbf{x},\mathbf{y})\}$.

Compute the sample average loss function

$$\hat{L}(w_\theta) = \frac{1}{b_D^2 b_k} \sum_{i,j \in \mathscr{M}_D} \sum_{u \in \mathscr{M}_k} \Delta(w_\theta; X_i, X_{i+1}) \Delta(w_\theta; X_j, X_{j+1}) k_u(X_{i+1}, X_{j+1}) + \frac{\alpha}{2} \left(\frac{1}{b_D} \sum_{i \in \mathscr{M}_D} w_\theta(X_i) - 1 \right)^2 .$$

Update the parameter θ by $\theta \leftarrow \theta - \varepsilon \nabla_\theta \hat{L}(w_\theta)$.

end for

Output: Estimate the tail probability of the original system by

$$\hat{\mathbf{P}}_{MLIS}(X^1 + X^2 \geq \gamma) = \frac{\sum_{t=1}^{T} w_\theta(X_t) \cdot \mathbf{1}\{X_t^1 + X_t^2 \geq \gamma\}}{\sum_{t=1}^{T} w_\theta(X_t)}.$$

5 NUMERICAL EXPERIMENTS

In this section, we would compare the performance of our method MLIS summaried in Algorithm 1 with different benchmark methods for a two-station tandem queueing network with parameters (λ, μ_1, μ_2) satisfying the inequality in (3). It is well known that the queue length overflow probability at steady state for a tandem queueing network has an explicit expression, i.e.,

$$\mathbb{P}(X^1 + X^2 \geq \gamma) = \frac{(1-\rho_1)\rho_2^{\gamma+1} - (1-\rho_2)\rho_1^{\gamma+1}}{\rho_2 - \rho_1}, \tag{5}$$

where the load parameter $\rho_i = \lambda/\mu_i$ for $i \in \{1, 2\}$.

5.1 Benchmarks and Performance Metrics

The first benchmark method is the marginal importance sampling method (MIS). Let us consider an alternative two-station tandem queueing network with exponential arrival rate $\tilde{\lambda}$ and exponential service rates $\tilde{\mu}_i$, where the parameters satisfy the stability condition $\tilde{\lambda} < \tilde{\mu}_2 \leq \tilde{\mu}_1$ and queueing network has a stationary distribution $\tilde{\pi}$. Let us generate a sample path $\{\tilde{X}_t\}$ of the uniformized DTMC derived from this alternative system, then the MIS estimator is computed by the following equation, i.e.,

$$\hat{\mathbf{P}}_{MIS}(X^1 + X^2 \geq \gamma) = \frac{\sum_{t=1}^{T} \frac{\pi}{\tilde{\pi}}(\tilde{X}_t) \cdot \mathbf{1}\{\tilde{X}_t^1 + \tilde{X}_t^2 \geq \gamma\}}{\sum_{t=1}^{T} \frac{\pi}{\tilde{\pi}}(\tilde{X}_t)}.$$

The key distinction between MIS estimator and our MLIS estimator is that MIS directly uses the stationary likelihood ratio $\pi/\tilde{\pi}$, whereas MLIS learns this ratio via a neural network. While MIS is expected to outperform our algorithm, we hope the performance gap, that is the price of learning the stationary likelihood ratio $\pi/\tilde{\pi}$, would be small.

The other benchmark method is the regenerative importance sampling method (RIS) discussed in Section 3.2, where the RIS estimator is computed by the following equation, i.e.,

$$\hat{\mathbf{P}}_{RIS}(X^1+X^2\geq\gamma)=\frac{m_1^{-1}\sum_{i=1}^{m_1}\left[\left(\sum_{t=1}^{\tilde{\alpha}_i}\mathbf{1}\{\tilde{X}_{t,i}^1+\tilde{X}_{t,i}^2\geq\gamma\}\right)\cdot\left(\prod_{t=1}^{\tilde{\tau}_i}\frac{P(\tilde{X}_{t,i}|\tilde{X}_{t-1,i})}{Q(\tilde{X}_{t,i}|\tilde{X}_{t-1,i})}\right)\right]}{m_2^{-1}\sum_{i=1}^{m_2}\alpha_i}.$$

Specifically, the numerator is computed by m_1 regenerative cycles of data from the alternative system with parameters $(\tilde{\lambda},\tilde{\mu}_1,\tilde{\mu}_2)=(\mu_2,\mu_1,\lambda)$, and the denominator is computed by m_2 regenerative cycles of data from the original system. For a fair comparison, we let the number of data sample generated for computing a RIS estimator be equal to (or very closed to) that generated for computing an MIS estimator, namely

$$\sum_{i=1}^{m_1}\tilde{\alpha}_i+\sum_{i=1}^{m_2}\alpha_i\approx T.$$

The mean squared error (MSE) for each IS method is estimated based on 500 independent rounds of simulation, that is

$$MSE_\gamma=\frac{1}{500}\sum_{i=1}^{500}\left(\hat{\mathbf{P}}_{a,i}(X^1+X^2\geq\gamma)-\mathbb{P}(X^1+X^2\geq\gamma)\right)^2,$$

where $\hat{\mathbf{P}}_{a,i}(X^1+X^2\geq\gamma)$ is the IS estimator for some method a in the i-th round of simulation. We evaluate the performance of different IS methods in term of their relative mean squared errors (rMSE), namely

$$rMSE_\gamma=\frac{\sqrt{MSE_\gamma}}{\mathbb{P}(X^1+X^2\geq\gamma)}.$$

5.2 Implementation Details for MLIS

We hereby reveal the implementation details for our method MLIS. Our algorithm uses sample data of the uniformized DTMC derived from the alternative tandem queueing network with parameters $(\tilde{\lambda},\tilde{\mu}_1,\tilde{\mu}_2)$, which satisfy the stability condition $\tilde{\lambda}<\tilde{\mu}_2\leq\tilde{\mu}_1$.

The performance of our algorithm is evaluated with different choices of alternative systems. For each alternative system, a fully connected neural network w_θ is constructed to approximate the stationary likelihood ratio $\pi/\tilde{\pi}$ of queue length distributions in the original and alternative systems. Each neural network consists of two hidden layers, each with $1,024$ neurons activated by the ReLU function, and the output layer is activated by the softplus function, i.e., $\text{softplus}(x)=\log(1+\exp(x))$, to ensure a nonnegative result. The learning rate ε is set to 10^{-5}.

Four types of kernel functions are used to construct the loss function for training the neural network. Namely, for $\sigma>0$ and $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$,

$$\begin{aligned}
\text{Gaussian:}\quad & k_{GS}(\mathbf{x},\mathbf{y})=\exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}\right),\\
\text{Laplacian:}\quad & k_{LP}(\mathbf{x},\mathbf{y})=\exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|}{\sigma}\right),\\
\text{Inverse multiquadratic:}\quad & k_{IM}(\mathbf{x},\mathbf{y})=\left(1+\frac{||\mathbf{x}-\mathbf{y}||^2}{\sigma^2}\right)^{-1/2},\\
\text{Linear:}\quad & k_{LN}(\mathbf{x},\mathbf{y})=\mathbf{x}^T\mathbf{y}.
\end{aligned}$$

The parameter σ is chosen to be the median of Euclidean distances computed using a batch of sample data. We set the data batch size $b_D=3,000$, kernel batch size $b_k=4$.

5.3 Performance Comparison

The simulation is implemented in Python and runs on a 2021 version MacBook Pro with an 8-core Apple M1 pro chip. The threshold parameter γ is set to be large so that the target event, as we can compute its probability by (5), is indeed a rare event. The numerical results and computation times are reported in Table 1. It is clear that our algorithm MLIS with different choices of alternative systems all achieve lower relative MSE than the state-of-the-art RIS method. MLIS maintains a low level of relative MSE even when the target event $\{X^1+X^2 \geq \gamma\}$ becomes rarer. Conversely, the relative MSE for the state-of-the-art RIS method increases dramatically. This result is consistent with our previous discussion in Section 3.2 since a wider range of sample paths in a regenerative cycle can reach the target event when the threshold parameter γ increases. Moreover, MLIS with different choices of alternative systems maintains relative MSE within the same order of magnitude as the MIS method, which indicates that our algorithm estimates the true stationary likelihood ratio $\pi/\tilde{\pi}$ accurately.

Table 1: Performance of different IS estimation methods in a two-station tandem queueing network with parameter $(\lambda,\mu_1,\mu_2) = (1/10, 23/50, 11/25)$. The relative MSE is estimated based on 500 independent rounds of simulation. Time horizon $T = 100,000$.

Target Probability	Method	$\tilde{\lambda}$	$\tilde{\mu}_1$	$\tilde{\mu}_2$	rMSE	Running Time (s)
$\mathbb{P}(X^1+X^2 \geq 16) = 4.891\times 10^{-10}$	MLIS	3/11	4/11	4/11	0.325	611
	MIS	3/11	4/11	4/11	0.153	417
	MLIS	5/17	6/17	6/17	0.278	598
	MIS	5/17	6/17	6/17	0.168	432
	MLIS	7/23	8/23	8/23	0.191	603
	MIS	7/23	8/23	8/23	0.145	409
	RIS	11/25	23/50	1/10	0.573	698
$\mathbb{P}(X^1+X^2 \geq 18) = 2.712\times 10^{-11}$	MLIS	3/11	4/11	4/11	0.311	609
	MIS	3/11	4/11	4/11	0.195	441
	MLIS	5/17	6/17	6/17	0.236	612
	MIS	5/17	6/17	6/17	0.187	428
	MLIS	7/23	8/23	8/23	0.227	607
	MIS	7/23	8/23	8/23	0.158	425
	RIS	11/25	23/50	1/10	1.859	701
$\mathbb{P}(X^1+X^2 \geq 20) = 1.489\times 10^{-12}$	MLIS	3/11	4/11	4/11	0.467	597
	MIS	3/11	4/11	4/11	0.259	437
	MLIS	5/17	6/17	6/17	0.464	603
	MIS	5/17	6/17	6/17	0.202	431
	MLIS	7/23	8/23	8/23	0.453	610
	MIS	7/23	8/23	8/23	0.171	447
	RIS	11/25	23/50	1/10	3.241	702

5.4 Influence of Kernel Functions

In this section, we evaluate the performance of our method MLIS with a total of 15 different combinatorial choices of the four commonly used kernel functions mentioned in Section 5.2. We set the data batch size $b_D = 3,000$ and the kernel batch size b_k to equal the number of selected kernels in each experiment.

The numerical results and training times are reported in Table 2. A key observation is that the linear kernel plays an important role in improving the performance of our algorithm, and any set of kernel functions without linear kernel cannot help our algorithm learn the correct stationary likelihood ratio $\pi/\tilde{\pi}$. Moreover, although using more kernels can only marginally reduce our estimator's relative mean squared error, it significantly improves the efficiency of learning the likelihood ratio.

Table 2: Performance of MLIS with different choices of kernel sets in a two-station tandem queueing network with parameter $(\lambda,\mu_1,\mu_2) = (1/10, 23/50, 11/25)$. The alternative tandem queueing network is chosen with parameter $(\tilde{\lambda},\tilde{\mu}_1,\tilde{\mu}_2) = (7/23, 8/23, 8/23)$. The relative MSE is estimated based on 500 independent rounds of simulation. Time horizon $T = 100,000$.

Chosen Kernels	rMSE for Estimating $\mathbb{P}(X^1+X^2 \geq 16)$	rMSE for Estimating $\mathbb{P}(X^1+X^2 \geq 18)$	rMSE for Estimating $\mathbb{P}(X^1+X^2 \geq 20)$	Training Time (s)
k_{LN}	1.235	2.811	5.678	2314
k_{LN}, k_{GS}	0.254	0.288	0.777	2502
k_{LN}, k_{LP}	0.324	0.253	0.677	2030
k_{LN}, k_{IM}	0.226	0.327	0.773	1901
k_{LN}, k_{GS}, k_{LP}	0.265	0.295	0.825	1561
k_{LN}, k_{GS}, k_{IM}	0.245	0.337	0.904	1493
k_{LN}, k_{LP}, k_{IM}	0.247	0.342	0.562	1447
k_{LN}, k_{GS}, k_{LP}, k_{IM}	0.191	0.227	0.453	982
MIS	0.145	0.158	0.171	-

6 CONCLUSION

In this paper, we propose a novel algorithm to estimate the tail probability of the stationary distribution of queuing networks, combining importance sampling with machine learning techniques. In detail, our algorithm applies importance sampling directly on the stationary distributions, to avoid the excessive variance encountered by the classic path-dependent method, and leverages machine learning techniques to approximate the likelihood ratio corresponding to the stationary distributions. Numerical experiments demonstrate that our algorithm, across a reasonable wide range of importance distributions, consistently outperforms the benchmark methods.

There are several interesting directions for future exploration. First, despite the numerical findings showing that our algorithm's performance is stable across a reasonably wide range of importance distributions, it remains worthwhile to determine, through either analytical or numerical means, the optimal set of importance distributions to further reduce the variance. Moreover, our numerical results suggest that the linear kernel has a significant impact on enhancing our algorithm's performance. We hypothesize that the linear kernel effectively captures the tail structure of the true stationary likelihood ratio $\pi/\tilde{\pi}$, but a formal mathematical justification remains necessary. One interesting direction of future study is to derive theoretic performance guarantees, e.g. variance bound or convergence rate, and their dependence on the chosen kernels. Another valuable direction, from the practical aspect, is to extend MLIS across a wide range of queueing networks, especially to large scale systems.

ACKNOWLEDGMENTS

The authors acknowledge support from NSFC 72171205 and RCXY 20210609103124047.

REFERENCES

Asmussen, S., and P. W. Glynn. 2007. *Stochastic simulation: algorithms and analysis*. New York: Springer.

Berlinet, A., and C. Thomas-Agnan. 2011. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. New York: Springer.

Blanchet, J., P. Glynn, and J. C. Liu. 2007. "Fluid heuristics, Lyapunov bounds and efficient importance sampling for a heavy-tailed G/G/1 queue". *Queueing Systems* 57:99–113.

Blanchet, J., and M. Mandjes. 2009. "Rare Event Simulation for Queues". In *Rare Event Simulation using Monte Carlo Methods*, edited by G. Rubino and B. Tuffin, 87–124. Chichester: John Wiley & Sons, Ltd.

Chow, Y., M. Petrik, and M. Ghavamzadeh. 2015. "Robust policy optimization with baseline guarantees". *arXiv preprint arXiv:1506.04514*.

Cisco 2025. "Cisco Nexus 9000 Series NX-OS Quality of Service Configuration Guide". https://www.cisco.com/c/en/us/td/docs/dcn/nx-os/nexus9000/105x/configuration/qos/cisco-nexus-9000-series-nx-os-quality-of-service-configuration-guide-105x.html, accessed 9th June 2025.

Crane, A. A., and J. Lemoine. 1977. *An Introduction to the Regenerative Method for Simulation Analysis*. Berlin: Springer-Verlag.

De Boer, P.-T. 2006. "Analysis of state-independent importance-sampling measures for the two-node tandem queue". *ACM Trans. Model. Comput. Simul.* 16(3):225–250.

Dupuis, P., and H. Wang. 2007. "Subsolutions of an Isaacs Equation and Efficient Schemes for Importance Sampling". *Mathematics of Operations Research* 32(3):723–757.

Dupuis, P., and H. Wang. 2009. "Importance sampling for Jackson networks". *Queueing Systems* 62(1-2):113–157.

Fonteneau, R., S. A. Murphy, L. Wehenkel, and D. Ernst. 2013. "Batch mode reinforcement learning based on the synthesis of artificial trajectories". *Annals of Operations Research* 208(1):383–416.

Glasserman, P., and S.-G. Kou. 1995, January. "Analysis of an importance sampling estimator for tandem queues". *ACM Trans. Model. Comput. Simul.* 5(1):22–42.

Guang, J., G. Hong, X. Chen, X. Peng, L. Chen, B. Bai *et al*. 2022. "Tail Quantile Estimation for Non-Preemptive Priority Queues". In *2022 Winter Simulation Conference (WSC)*, 85–96 https://doi.org/10.1109/WSC57314.2022.10015324.

Harchol-Balter, M. 2021. "Open problems in queueing theory inspired by datacenter computing". *Queueing Systems* 97(1-2):3–37.

Kallus, N., and M. Uehara. 2022. "Efficiently Breaking the Curse of Horizon in Off-Policy Evaluation with Double Reinforcement Learning". *Operations Research* 70(6):3282–3302.

Liu, Q., L. Li, Z. Tang, and D. Zhou. 2018. "Breaking the curse of horizon: infinite-horizon off-policy estimation". In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, edited by S. Bengio and H. M. Wallach, 5361–5371. Red Hook: Curran Associates Inc.

Liu, Y., P.-L. Bacon, and E. Brunskill. 2020. "Understanding the curse of horizon in off-policy evaluation via conditional importance sampling". In *Proceedings of the 37th International Conference on Machine Learning*. July 13th-18th, 6184-6193.

Milner, J. M., and T. L. Olsen. 2008. "Service-Level Agreements in Call Centers: Perils and Prescriptions". *Management Science* 54(2):238–252.

Parekh, S., and J. Walrand. 1989. "A quick simulation method for excessive backlogs in networks of queues". *IEEE Transactions on Automatic Control* 34(1):54–66.

Ross, S. M. 2014. *Introduction to probability models*. Burlington: Academic press.

Setayeshgar, L., and H. Wang. 2011. "Large deviations for a feed-forward network". *Advances in Applied Probability* 43(2):545–571.

Uehara, M., C. Shi, and N. Kallus. 2022. "A review of off-policy evaluation in reinforcement learning". *arXiv preprint arXiv:2212.06355*.

Whitt, W. 1993. "Tail probabilities with statistical multiplexing and effective bandwidths in multi-class queues". *Telecommunication Systems* 2(1):71–107.

Xie, T., Y. Ma, and Y.-X. Wang. 2019. "Towards Optimal Off-Policy Evaluation for Reinforcement Learning with Marginalized Importance Sampling". In *Advances in Neural Information Processing Systems*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, 9668–9678. Red Hook: Curran Associates, Inc.

AUTHOR BIOGRAPHIES

RUONING ZHAO is a Ph.D. student in the School of Data Science at The Chinese University of Hong Kong, Shenzhen. He earned his bachelor degree in the School of Science and Engineering at The Chinese University of Hong Kong, Shenzhen. His research interests include stochastic simulation and queueing theory. His email address is 115010144@link.cuhk.edu.cn.

XINYUN CHEN is an Associate Professor in the School of Data Science at The Chinese University of Hong Kong, Shenzhen. She earned her Ph.D. in Operations Research at Columbia University. Her research interests include applied probability, stochastic simulation, queueing theory and reinforcement learning. Her email address is chenxinyun@cuhk.edu.cn and her website is https://myweb.cuhk.edu.cn/chenxinyun.