DDA 6104: Final Project Proposal
Efficient Steady State Tail Quantile Estimation for Non-preemptive Priority Queues

Kang Gong
School of Data Science
The Chinese University of Hong Kong, Shenzhen
kanggong@link.cuhk.edu.cn

Ziyue Lin
School of Data Science
The Chinese University of Hong Kong, Shenzhen
120090372@link.cuhk.edu.cn

1 Motivation

In the ever-evolving landscape of computing and telecommunication, the effective management and delivery of services are of paramount importance. Service-level agreements (SLA) play a pivotal role in ensuring that service providers meet the expectations and requirements of their clients. These agreements delineate the terms, standards, and commitments regarding service quality, availability, and performance. Therefore, the accurate and efficient management of SL is essential for both service providers and their clients, ensuring that the digital infrastructure supporting our interconnected world operates seamlessly.

In computing and telecommunication industries, a commonly used performance metric in SLA takes the form a guarantee on the tail quantile of a job's response time, also known as sojourn time, at steady state, i.e.,

$$Q(p) = \inf\{\gamma : P(\text{response time} > \gamma) < 1 - p\},$$

where in practice p is chosen to be closed to 1, like $p = 0.999$ (Harchol-Balter, 2021), so as to guarantee a reliable service quality. However, a reliable service quality basically means that it is rare to observe a violation that a job's response time exceeds γ, which means a naïve Monte-Carlo simulation is extremely inefficient. Worse yet, different jobs' response times are positively correlated to each other. Therefore, an efficient simulation algorithm is needed.

2 Related Works

2.1 Rare Event Simulation of Queueing System

The rare event in a queueing system, like the response time violation, is often addressed by importance sampling method, see Blanchet and Mandjes (2009) and the references therein. A sequence of works are focused on single-class systems, including $G/G/1$ queue (Blanchet et al., 2007), fluid network (Chang et al., 1994), and Jackson network (Parekh and Walrand, 1989; Dupuis and Wang, 2009). Setayeshgar and Wang (2013) extend importance sampling based rare event simulation to a preemptive server with 2 priority classes. In these works, the importance distributions are generally derived from large deviation principle as well as differential game approaches, which heavily depends on the specific system dynamics. For this reason, these works could hardly be extended to systems with multiple customer classes and general service protocols.

To the best of our knowledge, Guang et al. (2022) firstly propose a novel importance sampling algorithm on tail quantile estimation for the non-preemptive multi-class priority queue. They use cross-entropy method (Rubinstein and Kroese, 2004) to optimize the importance distribution. To estimate the steady-state distribution and facilitate output analysis, Guang et al. (2022) and Blanchet et al. (2007) apply the regenerative simulation framework (Asmussen and Glynn, 2007). However, in order to intensify the rare events, the importance distribution inevitably enlarges the regenerative cycles defined in their models, which in turn deteriorates the algorithm's efficiency.

2.2 Off-policy Evaluation via Importance Sampling

Another stream of literature about steady state importance sampling lives in the off-policy estimation of long-horizon average reward in the field of reinforcement learning (RL). Off-policy policy evaluation (Sutton and Barto, 2018) is to estimate the expected reward of a given target policy from the data obtained by an agent running a different behavior policy in the environment. The reason why data is obtained by a different policy is that, in many real application domains, deploying a new policy could be highly costly or dangerous, such as medical interventions (Murphy et al., 2001), econometrics (Hirano et al., 2003), education (Mandel et al., 2014), and advertisement (Bottou et al., 2013).

Importance sampling is one of most state-of-art off-policy estimation methods, see (Liu, 2001; Xie et al., 2019) and the references therein. Most existing importance sampling based estimators are based on the product of the importance ratios of many steps in the trajectory (Liu et al., 2020), and thus variances in individual steps would accumulate multiplicatively. It is well known that these estimators , if well-defined, suffer from high variance when estimating long-horizon average reward, which is called curse of horizon (Liu et al., 2018).

Liu et al. (2018) address the problem of long-horizon off-policy policy evaluation by applying importance weighting on the state space rather than the trajectory space. The core idea is to employ importance sampling on the average visitation distribution of single state-action pairs, which could substantially reduce estimation variance compared to dealing with the distribution of entire trajectories. They use a mini-max loss function for estimating the true stationary density ratio, which yields a closed-form representation when combined with a reproducing kernel Hilbert space (RKHS). Chen et al. (2020) extend off-policy policy evaluation when the trajectory data are generated by multiple behavior policies, where they proposed estimated mixture policy (EMP) to reduce the variances of evaluation as well as offered an useful induction bias for estimating the state-action stationary distribution correction.

Given these promising developments in addressing the long-horizon off-policy evaluation problem, we are prompted to explore the possibility of leveraging these successful approaches to tackle the formidable challenge of tail-quantile estimation for non-preemptive priority queues within the telecommunication industry. This research proposal aims to investigate the applicability of methods in off-policy reinforcement learning to improve the efficiency and accuracy of tail-quantile estimation, while also simplifying the associated computational complexity. In doing so, we hope to contribute to the advancement of robust estimation techniques in the telecommunications sector and enhance the reliability of SLA agreements.

3 Feasible Research Plan

Let $F_R(\gamma)$ be the cdf of steady state sojourn time. Intuitively, we would obtain a good estimation of $Q(p)$ if the estimation of $F_R(\gamma)$ is accurate for γ in a neighborhood of $Q(p)$. In Section 3.1, a simple queueing model is introduced. In Section 3.2, a novel steady state importance sampling algorithm is presented to estimate $F_R(\gamma)$ for the simple model. In Section 3.3, we would introduce the possible extensions of our work.

3.1 Model Setting

Let us consider a single server with one job class for simplicity. Each job arrives according to a Poisson process with rate $\lambda \in \mathbb{R}_+$ and the server has i.i.d. exponential service times with rate $\mu > \lambda$. Following the setting in real-life systems, we assume that the server is non-preemptive and jobs are served in a first-come-first-serve manner. The arrival process and service times are assumed to be mutually independent.

In order to analyze this queueing system's service level performance, we particularly focus on each job's total sojourn time, consisting of its waiting time and service time. Following the famous Lindley's recursion, the sojourn time R_{t+1} for the $(t+1)$st job in the system would be written as

$$R_{t+1} = (R_t - A_{t+1})^+ + S_{t+1}, \tag{1}$$

where A_{t+1} and S_{t+1} are $(t+1)$st job's inter-arrival time and service time with pdf f_A and f_S respectively, and $R_0 = 0$. Let R be a random variable, with pdf f_R, having the stationary distribution of (R_t). Then, the ccdf of steady state sojourn time could be expressed as

$$\mathbf{P}(R > \gamma) = \int_{\mathbb{R}} \mathbf{1}\{r > \gamma\} f_R(r) dr.$$

It would be unlikely to observe $\mathbf{1}\{R > \gamma\}$ when γ is large, hence an efficient algorithm is required.

3.2 Algorithm Design

Let us consider an alternative single server system, driven by inter-arrival times $(\tilde{A}_t)$ and service times $(\tilde{S}_t)$ with some different pdfs $f_{\tilde{A}}$ and $f_{\tilde{S}}$ respectively. Intuitive, in this system, the sojourn times $(\tilde{R}_t)$ are more likely to exceed γ compared with the original systems. Let $\tilde{R}$ be a random variable, with pdf $f_{\tilde{R}}$, having the stationary distribution of $(\tilde{R}_t)$. Then, the ccdf of steady state sojourn time for the original system could be expressed as

$$\mathbf{P}(R > \gamma) = \int_{\mathbb{R}} \mathbf{1}\{r > \gamma\} \frac{f_R(r)}{f_{\tilde{R}}(r)} \cdot f_{\tilde{R}}(r) dr = \mathbf{E}_{\tilde{R} \sim f_{\tilde{R}}} \left[\mathbf{1}\{\tilde{R} > \gamma\} \frac{f_R(\tilde{R})}{f_{\tilde{R}}(\tilde{R})} \right].$$

However, the stationary distributions of sojourn times for both systems are unknown in general. Therefore, Theorem 1 below is needed for the estimation of stationary density ratio $f_R(r)/f_{\tilde{R}}(r)$.

Theorem 1. *Suppose $\tilde{A}, \tilde{S}, \tilde{R}$ are independently sample from $f_{\tilde{A}}$, $f_{\tilde{S}}$, $f_{\tilde{R}}$ respectively. Let $\tilde{R}' = (\tilde{R} - \tilde{A})^+ + \tilde{S}$, which also follows the distribution $f_{\tilde{R}}$. Then, a function $w(r)$ equals $f_R(r)/f_{\tilde{R}}(r)$ (up to a constant factor) if and only if it satisfies*

$$\begin{aligned} &\mathbf{E}[\Delta(w; \tilde{A}, \tilde{S}, \tilde{R}) \mid \tilde{R}'] = 0, \quad \text{for all possible } \tilde{R}', \\ &\text{with} \quad \Delta(w; \tilde{A}, \tilde{S}, \tilde{R}) \equiv w(\tilde{R}) \cdot \frac{f_A(\tilde{A}) f_S(\tilde{S})}{f_{\tilde{A}}(\tilde{A}) f_{\tilde{S}}(\tilde{S})} - w(\tilde{R}'). \end{aligned} \tag{2}$$

The tricky part in in (2) is that the conditional expectation is taken time reversed. It is unlikely to observe multiple difference pairs of $(\tilde{A}, \tilde{S}, \tilde{R})$ transiting to the same $\tilde{R}'$, and thus the conditional expectation is hard to estimated. Following the ideas in Liu et al. (2018), this problem can be addressed by introducing a discriminator function and constructing a min-max loss function. Specifically, the loss function is defined as

$$L(w, f) \equiv \mathbf{E}[\Delta(w; \tilde{A}, \tilde{S}, \tilde{R}) f(\tilde{R}')].$$

Following Theorem 1, $w \propto f_R/f_{\tilde{R}}$ if and only if $L(w, f) = 0$ for any function f. From a conservative perspective, we can then estimate $f_R/f_{\tilde{R}}$ by solving the following min-max problem:

$$\min_w \; D(w) \equiv \max_{f \in \mathcal{F}} L(w/z_w, f)^2,$$

where $z_w = \mathbf{E}[w(\tilde{R})]$ is the normalization term to avoid the trivial solution $w \equiv 0$. $\mathcal{F}$ is a set of discriminator functions, which should be rich enough to identify w. A good example would be neural networks. Alternatively, Theorem 2 below shows another choice of $\mathcal{F}$, which enables a closed form representation of $D(w)$.

Theorem 2. *Assume $\mathcal{H}$ is a reproducing kernel Hilbert space (RKHS) of functions with a positive definite kernel $k(r, \bar{r})$. Let $\mathcal{F} \equiv \{f \in \mathcal{H} : ||f||_{\mathcal{H}} \leq 1\}$ be the unit ball of $\mathcal{H}$. We have*

$$\max_{f \in \mathcal{F}} L(w, f)^2 = \mathbf{E}[\Delta(w; \tilde{A}_1, \tilde{S}_1, \tilde{R}_1) \Delta(w; \tilde{A}_2, \tilde{S}_2, \tilde{R}_2) k(\tilde{R}_1', \tilde{R}_2')],$$

where $(\tilde{A}_i, \tilde{S}_i, \tilde{R}_i) \overset{iid}{\sim} f_{\tilde{A}} \times f_{\tilde{S}} \times f_{\tilde{R}}$.

Inspired by Theorem 2, we summarize our steady state importance sampling algorithm in Algorithm

Algorithm 1 Steady State Importance Sampling

Input: Simulation data $\mathcal{D} = \{\tilde{A}_t, \tilde{S}_t, \tilde{R}_{t-1}, \tilde{R}_t\}_{t=1}^T$ from alternative queueing system.
Initiate the density ratio $w(r) = w_\theta(r)$ to be a neural network parameterized by θ.
for iteration $= 1, 2, ...$ **do**
Randomly choose a batch $\mathcal{M}$ of from the data $\mathcal{D}$, i.e., $\mathcal{M} \subset \{1, ..., T\}$
Update parameter $\theta \leftarrow \theta - \epsilon \nabla_\theta \hat{D}(w_\theta / z_{w_\theta})$, where

$$\hat{D}(w_\theta) = \frac{1}{|\mathcal{M}|} \sum_{i,j \in \mathcal{M}} \Delta(w_\theta; \tilde{A}_i, \tilde{S}_i, \tilde{R}_{i-1}) \Delta(w_\theta; \tilde{A}_j, \tilde{S}_j, \tilde{R}_j) k(\tilde{R}'_i, \tilde{R}'_j),$$

and z_{w_θ} is a normalization constant, i.e., $z_{w_\theta} = \frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} w_\theta(\tilde{R}_i)$.
end for
Output: Estimate the tail probability of steady state sojourn time for the original system by

$$\hat{\mathbf{P}}_T(R > \gamma) = \frac{\sum_{t=1}^n w_\theta(\tilde{R}_i) \mathbf{1}\{\tilde{R}_i > \gamma\}}{\sum_{t=1}^n w_\theta(\tilde{R}_i)}.$$

3.3 Extensions

In this section, we would list the following potential extensions that we plan to work on.

- **Multi-class Priority Queues:** From the perspective of real life applications, it is natural to analyze the service level performance for a single server queue with jobs of multiple priority classes. We need to estimate the tail probability of steady state sojourn time for each priority class. However, now the state space is getting larger, and the stationary density ratio would be difficult to compute.
- **Data-driven:** In order to estimate the stationary density ratio via (2), we assume the distributions of inter-arrival and service times are known for both original and alternative queueing systems. Instead, ccdf of steady state sojourn time for the original system could also be expressed as

$$\mathbf{P}(R > \gamma) = \mathbf{E}_{\tilde{A}, \tilde{S}, \tilde{R} \sim f_{\tilde{A}, \tilde{S}, \tilde{R}}} \left[\mathbf{1}\{\tilde{R} > \gamma\} \frac{f_{A,S,R}(\tilde{A}, \tilde{S}, \tilde{R})}{f_{\tilde{A}, \tilde{S}, \tilde{R}}(\tilde{A}, \tilde{S}, \tilde{R})} \right].$$

  Thus, we can estimate the joint stationary density ratio without knowing the exact distributions of the inter-arrival and service times.
- **Normalization:** We use $z_{w_\theta} = \frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} w_\theta(\tilde{R}_i)$ in Algorithm 1 to avoid the trivial estimation $w \equiv 0$. This method would be inefficient, and we may set a regularization term when solving the min-max problem, c.f. Zhang et al. (2020).
- **Choice of Importance Measure:** The cross-entropy method in (Rubinstein and Kroese, 2004) maybe helpful for choosing better importance distributions of inter-arrival and service times in the alternative queueing system.

References

Asmussen, S. and Glynn, P. W. (2007). *Stochastic simulation: algorithms and analysis*, volume 57. Springer.

Blanchet, J., Glynn, P., and Liu, J. C. (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 Mandjes, M. (2009). Rare event simulation for queues. *Rare Event Simulation Using Monte Carlo Methods*, pages 87–124.

Bottou, L., Peters, J., Quiñonero-Candela, J., Charles, D. X., Chickering, D. M., Portugaly, E., Ray, D., Simard, P., and Snelson, E. (2013). Counterfactual reasoning and learning systems: The example of computational advertising. *Journal of Machine Learning Research*, 14(11).

Chang, C.-S., Heidelberger, P., Juneja, S., and Shahabuddin, P. (1994). Effective bandwidth and fast simulation of atm intree networks. *Performance Evaluation*, 20(1-3):45–65.

Chen, X., Wang, L., Hang, Y., Ge, H., and Zha, H. (2020). Infinite-horizon off-policy policy evaluation with multiple behavior policies. *International Conference on Learning Representations*.

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

Guang, J., Hong, G., Chen, X., Peng, X., Chen, L., Bai, B., and Zhang, G. (2022). Tail quantile estimation for non-preemptive priority queues. In *2022 Winter Simulation Conference (WSC)*, pages 85–96. IEEE.

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

Hirano, K., Imbens, G. W., and Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4):1161–1189.

Liu, J. S. (2001). *Monte Carlo strategies in scientific computing*, volume 75. Springer.

Liu, Q., Li, L., Tang, Z., and Zhou, D. (2018). Breaking the curse of horizon: Infinite-horizon off-policy estimation. *Advances in neural information processing systems*, 31.

Liu, Y., Bacon, P.-L., and Brunskill, E. (2020). Understanding the curse of horizon in off-policy evaluation via conditional importance sampling. In *International Conference on Machine Learning*, pages 6184–6193. PMLR.

Mandel, T., Liu, Y.-E., Levine, S., Brunskill, E., and Popovic, Z. (2014). Offline policy evaluation across representations with applications to educational games. In *AAMAS*, volume 1077.

Murphy, S. A., van der Laan, M. J., Robins, J. M., and Group, C. P. P. R. (2001). Marginal mean models for dynamic regimes. *Journal of the American Statistical Association*, 96(456):1410–1423.

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

Rubinstein, R. Y. and Kroese, D. P. (2004). *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning*, volume 133. Springer.

Setayeshgar, L. and Wang, H. (2013). Efficient importance sampling schemes for a feed-forward network. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 23(4):1–19.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Xie, T., Ma, Y., and Wang, Y.-X. (2019). Towards optimal off-policy evaluation for reinforcement learning with marginalized importance sampling. *Advances in Neural Information Processing Systems*, 32.

Zhang, R., Dai, B., Li, L., and Schuurmans, D. (2020). Gendice: Generalized offline estimation of stationary values. *arXiv preprint arXiv:2002.09072*.