Rare-event Analysis and Simulation of Queues with Time-varying Rates

Ad Ridder

School of Business and Economics, Vrije University Amsterdam, Netherlands
ad.ridder@vu.nl

March 30, 2022

1 Introduction

In this study we are interested in rare-event probabilities in Markovian queues with time-varying arrival rates (nonhomogeneous Poisson arrivals) and time-varying service rates. Time-varying queues as they are called, act as important models for service systems such as call centers, for communcation systems such as wireless networks, for road intersection control in traffic management, for hospital admissions in health care logistics (Schwarz et al. (2016)). All these systems should operate properly under many situations, meaning that unwanted events, such as breakdown, failure, congestion, or overflow, happen rarely. In order to quantify and analyse the occurence of a disaster, we investigate here a simple, specific case. Namely we consider an $M_t/M_t/1$ queue which is stable on average, but that reaches a high level in a short time interval, representing the unwanted event.

The program that we pursue, is to analyse the behaviour of the queue during this interval, in other words "how does the rare event occur?", and secondly, to compute efficiently the probability of occurrence. This will be executed in three steps. First, we apply fluid scaling (Mandelbaum and Massey (1995)) to observe the regular behaviour, also called most likely behaviour, of the queue which should indicate that indeed the unwanted event is rare, meaning an exponential decaying probability of occurrence. Next, large deviations arguments from Dupuis and Ellis (1995); Shwartz and Weiss (1995) give us the most likely behaviour of the queue during the interval to the rare event. Finally, for estimating the rare-event probabilities, we propose an importance sampling simulation algorithm which is efficient in the sense that it has subexponential complexity (Juneja and Shahabuddin (2006)).

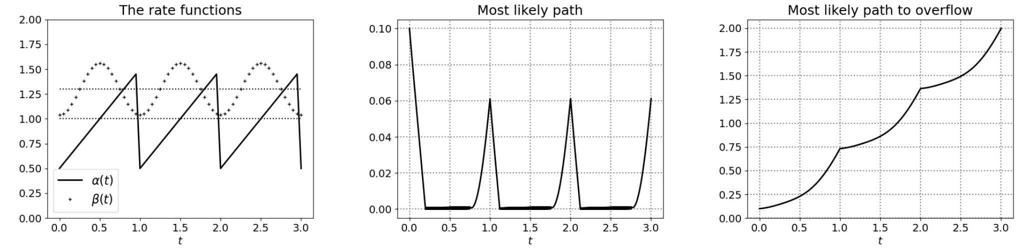

Figure 1: The rate functions (left), the most likely sample path (middle), and the optimal path to overflow (right).

2 Problem Statement

We shall describe the model and the problem in more detail. The arrival process at the $M_t/M_t/1$ queue is a nonhomogeneous Poisson process with rate function $\alpha(t)$, and the exponential server works at time-dependent rate function $\beta(t)$. For instance, these rate functions are periodic, as illustrated in Figure 1. Scaling time and space gives the scaled queueing processes $\{X^{(n)}(t) : t \geq 0\}, n = 1, 2, \ldots$ with rate functions $n\alpha(t)$, and $n\beta(t)$, and where the queue content is divided by n. Letting $n \to \infty$, these processes converge almost surely to a deterministic function $x(t) = \overline{x}+\int_0^t \alpha(s)\, ds-\int_0^t \beta(s) I\{x(s) > 0\}\, ds$, called the typical behaviour or most likely path, see Figure 1. It shows that when the queue follows this regular behaviour, it will never reach high levels. The rare event is that the scaled queue reaches from an initial state $\overline{x} > 0$, an overflow level $\overline{y} \gg \overline{x}$ at a finite, fixed horizon T. Its probability, $\ell_n = \mathbb{P}\big(X^{(n)}(T) \geq \overline{y} \,|\, X^{(n)}(0) = \overline{x}\big)$ satisfies a large deviation asymptotics, implying that ℓ_n is a rare event probability; i.e., $\lim_{n\to\infty} \frac{1}{n} \log \ell_n = -J$ for some $J > 0$.

For reaching high overflow levels $\overline{y}$ at small horizon T, the queue should build up rapidly, contrary to its regular behaviour, as illustrated in Figure 1. The rate functions are biased by an exponential tilt to become $\alpha^*(t) = \alpha(t)e^{\theta}$, and $\beta^*(t) = \beta(t)e^{-\theta}$, where the tilting factor $\theta > 0$, is computed in the same way as for the $M/M/1$ queue, which is thoroughly analysed in Shwartz and Weiss (1995). Then we obtain for small overflow horizon T, that the rare event happened most likely because the scaled queue builds up along the path $x(t) = \overline{x}+\int_0^t \alpha^*(s)\, ds-\int_0^t \beta^*(s) I\{x(s) > 0\}\, ds$, on the interval $[0, T]$.

For computing numerically the rare event probabilities ℓ_n, we have implemented an importance sampling algorithm that simulates the $M_t/M_t/1$ queue having the biased rate functions $n\alpha^*(t)$ and $n\beta^*(t)$. For accurate estimates we execute so many runs until the standard error of the estimator is about 2.5% of the estimated value. When we would implement this requirement for standard Monte Carlo simulations, the required sample sizes would grow exponentially (in n); see Figure 2. However for the implemented importance sampling algorithm the sample sizes grow subexponentially, which leads to the conjecture that the importance sampling estimator is asymptotically efficient. The importance sampling estimates are verified (i) by running standard Monte Carlo simulations for small overflow levels, and (ii) by applying numerical procedures for computing transient probabilities in finite nonhomogeneous Markov chains (Ingolfsson et al. (2007)). Figure

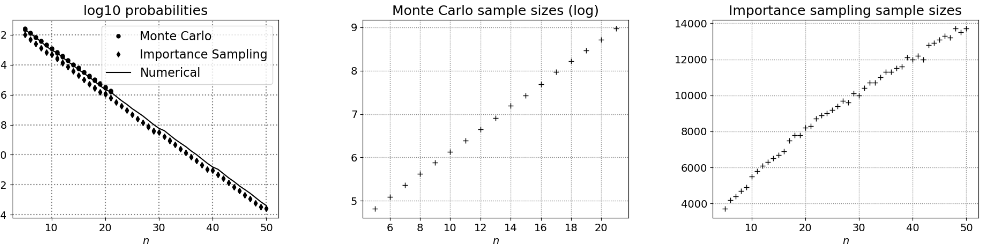

Figure 2: Left: numerical approximations, and confidence intervals for the overflow probabilities by simulation. Middle and right: required sample sizes for standard Monte Carlo and for the importance sampling.

2 gives a visual overview of results.

3 Discussion

When we inspect the numerical results of Figure 2 (left), and when we assume that the Monte Carlo simulations give the true values, we observe that the numerical approximations slighlty underestimate these, possibly due to state space truncation and numerical roundoffs. Secondly, the importance sampling gives even further underestimates. This is a well-known issue of importance sampling, which could be caused by small likelihood ratios when there is overbiasing (Smith (2001)), or by missing part of the rare event set (Glasserman and Wang (1997)). Moreover, difficulties arise for obtaining efficient importance sampling algorithms for processes with boundaries, which is illustrated by the counter examples in Asmussen et al. (2002); Glasserman and Kou (1995). These difficulties will arise when the rare event would involve large horizons T. However, for small horizons the boundary plays no role, but this needs further investigation and thorough analysis. Next, future work will be to consider rare events in more complex and realistic time-varying queues.

References

Asmussen, A., P. Fuckerieder, M. Jobmann, and H.-P. Schwefel (2002). Large deviations and fast simulation in the presence of boundaries. *Stochastic Processes and their Applications 102*, 1–23.

Dupuis, P. and R. Ellis (1995). Large deviations analysis of queueing systems. In F. Kelly and R. Williams (Eds.), *Stochastic Networks*, Volume 71 of *The IMA Volumes in Mathematics and its Applications*, pp. 347–365. New York: Springer.

Glasserman, P. and S.-G. Kou (1995). Analysis of an importance sampling estimator for tandem queues. *ACM Transactions on Modeling and Computer Simulation 5*(1), 22–42.

Glasserman, P. and Y. Wang (1997). Counterexamples in importance sampling for large deviations probabilities. *The Annals of Applied Probability 7*(3), 731–746.

Ingolfsson, A., E. Akhmetshina, S. Budge, Y. Li, and X. Wu (2007). A survey and experimental comparison of service-level-approximation methods for nonstationary M(t)/M/s(t) queueing systems with exhaustive discipline. *INFORMS Journal on Computing 19*(2), 201–214.

Juneja, S. and P. Shahabuddin (2006). Rare-event simulation techniques: An introduction and recent advances. In S. Henderson and B. Nelson (Eds.), *Simulation*, Volume 13 of *Handbook in Operations Research and Management Science*, Chapter 11, pp. 291 – 350. Elsevier.

Mandelbaum, A. and W. Massey (1995). Strong approximations for time-dependent queues. *Mathematics of Operations Research 20*(1), 33–64.

Schwarz, J., G. Selinka, and R. Stolletz (2016). Performance analysis of time-dependent queueing systems: Survey and classification. *Omega 63*, 170–189.

Shwartz, A. and A. Weiss (1995). *Large Deviations for Performance Analysis, Queues, Communication, and Computing*. New York: Chapman & Hall.

Smith, P. (2001). Underestimation of rare event probabilities in importance sampling simulations. *Simulation 76*(3), 140–150.