This article was downloaded by: [128.12.122.217] On: 23 May 2022, At: 11:15
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA

Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Technical Note—Approximating Systems Fed by Poisson Processes with Rapidly Changing Arrival Rates

Zeyu Zheng , Harsha Honnappa , Peter W. Glynn

To cite this article:
Zeyu Zheng , Harsha Honnappa , Peter W. Glynn (2021) Technical Note—Approximating Systems Fed by Poisson Processes with Rapidly Changing Arrival Rates. Operations Research 69(5):1566-1574. https://doi.org/10.1287/opre.2020.2031

Full terms and conditions of use: https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2021, INFORMS

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

informs
http://pubsonline.informs.org/journal/opre

OPERATIONS RESEARCH
Vol. 69, No. 5, September–October 2021, pp. 1566–1574
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Methods

Technical Note—Approximating Systems Fed by Poisson Processes with Rapidly Changing Arrival Rates

Zeyu Zheng,[a] Harsha Honnappa,[b] Peter W. Glynn[c]

[a] Department of Industrial Engineering and Operations Research, University of California, Berkeley, Berkeley, California 94720; [b] School of Industrial Engineering, Purdue University, West Lafayette, Indiana 47906; [c] Department of Management Science and Engineering, Stanford University, Stanford, California 94305

Contact: zyzheng@berkeley.edu, https://orcid.org/0000-0001-5653-152X (ZZ); honnappa@purdue.edu, https://orcid.org/0000-0002-0834-054X (HH); glynn@stanford.edu, https://orcid.org/0000-0003-1370-6638 (PWG)

Received: July 18, 2018
Revised: October 24, 2018; June 6, 2019
Accepted: April 1, 2020
Published Online in Articles in Advance: June 30, 2021

Subject Classifications: probability: stochastic model applications
Area of Review: Stochastic Modeling

https://doi.org/10.1287/opre.2020.2031

Copyright: © 2021 INFORMS

Abstract. This paper introduces a new asymptotic regime for simplifying stochastic models having nonstationary effects, such as those that arise in the presence of time-of-day effects. This regime describes an operating environment within which the arrival process to a service system has an arrival intensity that is fluctuating rapidly. We show that such a service system is well approximated by the corresponding model in which the arrival process is Poisson with a constant arrival rate. In addition to the basic weak convergence theorem, we also establish a first order correction for the distribution of the cumulative number of arrivals over $[0, t]$, as well as the number-in-system process for an infinite-server queue fed by an arrival process having a rapidly changing arrival rate. This new asymptotic regime provides a second regime within which nonstationary stochastic models can be reasonably approximated by a process with stationary dynamics, thereby complementing the previously studied setting within which rates vary slowly in time.

Keywords: counting processes • Poisson process • weak convergence • total variation convergence • compensator • intensity • infinite-server queue

1. Introduction

In many operations management settings, the arrival process to the system exhibits clear nonstationarities. These nonstationarities may arise as a consequence of time-of-day effects, day-of-week effects, seasonalities, or stochastic fluctuations in the arrival rate. One mathematical vehicle for studying such nonstationary arrival processes is to consider the setting in which the arrival rate changes slowly in time. In this setting, it is intuitively clear that the nonstationary system can be viewed as a small perturbation of a constant arrival rate system. Consequently, it seems conceptually reasonable that one should be able to study such slowly changing arrival rate models via an asymptotic expansion in which each of the terms in the expansion involves a stationary arrival rate calculation. This intuition has been validated rigorously by Khasminskii et al. (1996), Massey and Whitt (1998), and, more recently, Zheng et al. (2018).

In this paper, we show that arrival rate modeling also can be simplified significantly at the opposite end of the asymptotic spectrum in which the arrival rates fluctuate rapidly. Thus, we can view the results of this paper as complementing the existing literature on slowly varying arrival rate modeling. In particular, we study counting processes in which the intensity at time t is given by $\lambda_\epsilon(t) := \lambda(t/\epsilon)$, where ϵ is small and $\lambda = (\lambda(s) : s \geq 0)$ is a fixed process. We also study stochastic systems that are fed by such counting processes. The process λ could be a deterministic periodic function intended to model time-of-day effects, or it could be a functional of a positive recurrent Markov process. In either case, we show that when $\epsilon \downarrow 0$, we may view the system as one fed by a constant rate Poisson process with rate λ^* given by the long-run time average of λ; see Theorem 1 for details. Our theory applies not only to processes with deterministic and doubly stochastic intensities, but also to self-exciting counting processes like Hawkes processes. Indeed, there is increasing interest in studying service systems fed by self-exciting Hawkes processes; see Gao and Zhu (2016), Daw and Pender (2018), and Koops et al. (2018). Thus, this paper provides a second rigorously supported asymptotic regime within which the dynamics of a service system with a nonstationary arrival process can be approximated by a simpler system with stationary dynamics. We observe that the importance of such arrival processes has been acknowledged in the literature, though much of the recent literature focuses on deterministic time-of-day dependencies (see the recent survey by Whitt 2018).

Note that in our setting, the number of arrivals per time period remains bounded as $\epsilon \to 0$, so that the intensity fluctuates rapidly within a typical interarrival time.

Thus, our asymptotic regime describes systems wherein the time scale over which the intensity is fluctuating is much smaller than the time scale at which arrivals occur. These high-frequency fluctuations in the arrival rate may be a consequence of a short period, stochastic effects, or some combination of high-frequency periodicity and rapid stochastic fluctuations. As an example of a real-world system in which such an asymptotic regime may be appropriate, consider a construction equipment leasing company. If the leases tend to be of long duration (e.g., on the order of months), our theory suggests that in the analysis of a queueing model intended to predict lost sales (due to all the available equipment having been rented), one can safely ignore the daily periodicity in the arrival rate describing exogenous demand for the company's equipment. As a further example, consider a motor vehicle company that must handle warranty claims for repairs. Although warranty-covered failures likely exhibit daily periodicities because of customer vehicle usage patterns, the failure time for vehicles will be large relative to the period of the intensity, so could safely be ignored by the insurance company (according to the results of this paper). On the other hand, a towing company would take the intraday variations into consideration while making operational decisions on where to deploy its resources. Power-related failures of satellites in orbit (Landis et al. 2006) provide another vivid example in the context of reliability. Satellite failures are rare, but the fact that low Earth orbit satellites rapidly cycle between bright light and complete darkness places stress on the electronics and solar arrays that form the power systems. Although the intensity of light changes rapidly (in relation to the lifetime of the satellite), there are very few points of failure, suggesting that the number of failures in a time period corresponding to the typical lifetime of a satellite should be approximately Poisson.

It is important to observe that there are critical qualitative differences between the setting in this paper and the substantial literature on "rapid switching" diffusion and counting processes; see Khasminskii et al. (2007), Khasminskii (2012), Blom et al. (2014), Heemskerk et al. (2017), Koops et al. (2017, 2018), and Spreij and Storm (2018). In particular, in the rapid switching literature on counting processes, the process N itself is accelerated by a factor of $1/\epsilon$ (rather than its intensity) so that the arrivals and intensity are both fluctuating on the same time scale. The analysis there exploits an averaging effect (due, essentially, to the law of large numbers as applied to N) to establish weak limits for the rescaled counting processes. A consequence of the acceleration in the rapid switching regime is that the scaled counting processes $\epsilon N(t/\epsilon)$ appearing there cannot converge to a counting process, because it has jumps of size ϵ, not 1 as for a counting process. On the other hand, the "rapid fluctuation" in our setting is describing a rapid change in the intensity process, whereas the counting process itself is not rescaled, and our analysis identifies a constant rate Poisson process as the limit. Our results also depend on an "averaging effect," but our averaging arises as a consequence of the fact that the compensator of a rapidly fluctuating counting process converges to the "constant rate" compensator associated with the Poisson process. In particular, the averaging in our setting relates to the compensator rather than the counting process itself. Our results establish an alternative regime to the Palm–Khintchine superposition theorem (Cinlar and Agnew 1968) in which a Poisson process emerges as a weak limit of a sequence of counting processes. Crucially, unlike in the Palm–Khintchine setting, there is no superposition of independent and identically distributed (i.i.d.) counting processes in this setting.

This paper is organized as follows. Section 2 provides our main weak convergence theorem, establishing that counting processes with rapidly fluctuating intensities can be weakly approximated by a constant rate Poisson process (Theorem 1). In the remainder of the section, we compute the total variation (TV) distance between the counting process and the Poisson process in the Markov-modulated doubly stochastic setting, and prove that the TV distance does not tend to zero, thereby showing that one can expect to use the constant rate Poisson approximation only for suitably continuous path functionals. In Section 3, we study the distribution of the total number of arrivals in an interval $[0, t]$ and obtain a first order refinement to the weak convergence theorem that reflects the first order impact of the high-frequency fluctuations in the arrival rate; see Theorem 3. Finally, Section 4 provides a similar first order refinement in the setting of the number-in-system process for the infinite-server queue; see Theorem 4.

2. Weak Convergence to a Constant Rate Poisson Process

Our goal is to study the counting process N_ϵ having intensity λ_ϵ, where $\lambda_\epsilon(t) = \lambda(t/\epsilon)$. There are multiple different ways of constructing the process N_ϵ. For example, when the process N has a deterministic intensity or is doubly stochastic, we may generate N_ϵ from a Poisson process via inversion of the integrated intensity function; see Asmussen and Glynn (2007, p. 60) for details. Another alternative is to define N_ϵ via a change-of-measure argument; see Brémaud (1981, p. 165). However, the approach to constructing N_ϵ that works most generally (e.g., for Hawkes processes; see Hawkes 1971) is to construct N_ϵ by "thinning" the counting process N. We start with a fixed arrival counting process $N = (N(t) : t \geq 0)$. We assume that N is *simple*, in the sense that N increases by exactly one at each arrival epoch (and hence no batch arrivals are possible). We further require that N be adapted to a filtration $\mathcal{F} = (\mathcal{F}_t : t \geq 0)$, and that N possess a right-continuous nondecreasing $\mathcal{F}$-compensator $A = (A(t) : t \geq 0)$, so that $M = (M(t) : t \geq 0)$ is a martingale adapted to $\mathcal{F}$, where $M(t) = N(t) - A(t)$.

For $1 \geq \epsilon > 0$, let $\beta^\epsilon = (\beta_i^\epsilon; i \geq 1)$ be an i.i.d. sequence of Bernoulli(ϵ) random variables (RVs) independent of N. For $t \geq 0$, let $A_\epsilon(t) = \epsilon A(t/\epsilon)$, and let $\mathcal{G}_t^\epsilon$ be the smallest σ-algebra containing $\mathcal{F}_{t/\epsilon}$ and the σ-algebra $\sigma(\beta_i^\epsilon : 1 \leq i \leq N(t/\epsilon))$. Let $\mathcal{G}^\epsilon = (\mathcal{G}_t^\epsilon : t \geq 0)$ and

$$N_\epsilon(t) = \sum_{i=1}^{N(t/\epsilon)} \beta_i^\epsilon.$$

Then, $N_\epsilon = (N_\epsilon(t) : t \geq 0)$ is a simple counting process for which

$$\begin{aligned}
&\mathbb{E}\big[N_\epsilon(t+s) - A_\epsilon(t+s) | \mathcal{G}_t^\epsilon\big] \\
&= N_\epsilon(t) + \mathbb{E}\left[\sum_{i=N(t/\epsilon)+1}^{N((t+s)/\epsilon)} \beta_i^\epsilon \middle| \mathcal{G}_t^\epsilon\right] - \mathbb{E}\big[A_\epsilon(t+s)|\mathcal{G}_t^\epsilon\big] \\
&= N_\epsilon(t) + \mathbb{E}\beta_1^\epsilon \mathbb{E}\big[\big(N((t+s)/\epsilon) - N(t/\epsilon)|\mathcal{G}_t^\epsilon\big] \\
&\quad - \epsilon\mathbb{E}\big[A((t+s)/\epsilon)|\mathcal{G}_t^\epsilon\big] \\
&= N_\epsilon(t) + \epsilon\mathbb{E}\big[N((t+s)/\epsilon) - A((t+s)/\epsilon)|\mathcal{G}_t^\epsilon\big] \\
&\quad - \epsilon N(t/\epsilon) \\
&= N_\epsilon(t) + \epsilon\mathbb{E}\big[N((t+s)/\epsilon) - A((t+s)/\epsilon)|\mathcal{F}_{t/\epsilon}\big] \\
&\quad - \epsilon N(t/\epsilon) \\
&= N_\epsilon(t) + \epsilon(N(t/\epsilon) - A(t/\epsilon)) - \epsilon N(t/\epsilon) \\
&= N_\epsilon(t) - A_\epsilon(t),
\end{aligned}$$

for $s, t \geq 0$, so that A_ϵ is the $\mathcal{G}^\epsilon$-compensator of N_ϵ. (Here, we used the independence of β^ϵ from N in the third-to-last equality (see Kallenberg 1997, p. 87) and the fact that M is an $\mathcal{F}$-adapted martingale in the second-to-last equality.)

An important special case is when the compensator A can be written in the form $A(t) = \int_0^t \lambda(s)ds$, in which case $\lambda = (\lambda(t) : t \geq 0)$ is the $\mathcal{F}$-intensity of N. Then, N_ϵ has $\mathcal{G}^\epsilon$-intensity $\lambda_\epsilon = (\lambda_\epsilon(t) : t \geq 0)$, where $\lambda_\epsilon(t) = \lambda(t/\epsilon)$, and compensator $A_\epsilon(t) = \int_0^t \lambda(s/\epsilon)ds$ for $t \geq 0$. We can see clearly, in this setting, that N_ϵ has a rapidly fluctuating intensity as $\epsilon \downarrow 0$, so that this framework is indeed modeling such an asymptotic regime.

We now assume the following.

Assumption 1. *There exists a deterministic $\lambda^* \in (0, \infty)$ such that $\frac{1}{t}A(t) \Rightarrow \lambda^*$ as $t \to \infty$.*

Recall that $D[0, \infty)$ is the space of right-continuous functions on $[0, \infty)$ having left limits, endowed with the Skorohod J_1 topology; see Ethier and Kurtz (1986) for details.

Theorem 1. *Suppose that the compensator A satisfies Assumption 1. Then, $N_\epsilon \Rightarrow N_0$ in $D[0, \infty)$ as $\epsilon \to 0$, where $N_0 = (N_0(t) : t \geq 0)$ is a homogeneous Poisson process with rate λ^*.*

Proof of Theorem 1. Fix $t \geq 0$. Then, by definition, $\lambda^* t \in \mathcal{G}_0^\epsilon$ for every $\epsilon > 0$. Furthermore, we note that Assumption 1 implies that

$$A_\epsilon(t) = \epsilon A(t/\epsilon) = \left(\frac{\epsilon}{t}\right) A(t/\epsilon) \cdot t \Rightarrow \lambda^* t,$$

as $\epsilon \downarrow 0$. It follows that $A(t)$ and $A_\epsilon(t)$ satisfy the conditions (i) and (ii) of theorem 13.4.IV of Daley and Vere-Jones (1988), yielding the result. □

The following examples illustrate the settings in which Assumption 1 is satisfied.

Example 1. A nonhomogeneous counting process with deterministic and periodic intensity function $\lambda(t)$ that satisfies $T^{-1}\int_0^T \lambda(t)dt = a$, where T is the period, is an easy (if obvious) example of a stochastic process satisfying this condition.

Example 2. Let $N = (N(t) : t \geq 0)$ be an exponential Hawkes process (Laub et al. 2015) with conditional intensity function $\lambda(t) = \nu + \int_0^t \alpha e^{-\beta(t-s)} dN(s)$. When $\alpha/\beta < 1$, the strong law of large numbers for continuous time local martingales (Liptser 1980) and Hawkes processes (Bacry et al. 2013) together imply that $A(t)/t \Rightarrow \lambda^* := \nu/(1 - \alpha/\beta)$ as $t \to \infty$.

Of course, arrival processes typically serve as models describing exogenous inputs to queueing systems or service systems. Other sources of randomness described, say, by a random sequence (such as service time requirements, abandonment times, etc.) will typically also be present. If Z is independent of N_ϵ, it follows from Theorem 1 that

$$(Z, N_\epsilon) \Rightarrow (Z, N_0),$$

in $\mathbb{R}^\infty \times D[0, \infty)$ as $\epsilon \downarrow 0$. It follows that if $h : \mathbb{R}^\infty \times D[0, \infty) \to \mathbb{R}$ is continuous in the product topology at (Z, N_0) almost surely, then $h(Z, N_\epsilon) \Rightarrow h(Z, N_0)$ as $\epsilon \downarrow 0$ (via the continuous mapping principle; see Billingsley 1968, p. 21).

Consequently, if h is a map that sends (Z, N_ϵ) into some associated performance measure (e.g., the number-in-system at time t), we may infer that the performance measure can be computed as if the counting process N_ϵ is Poisson with rate λ^* (when ϵ is small).

In the remainder of this section, we make clear that although N_ϵ converges weakly to N_0 in $D[0, \infty)$ as $\epsilon \downarrow 0$, no convergence typically takes place in the total variation norm. More specifically, suppose that N_ϵ is a doubly stochastic Poisson process with stochastic intensity $\lambda_\epsilon = (\lambda_\epsilon(t) : t \geq 0)$, where $\lambda_\epsilon(t) = \lambda(t/\epsilon)$ for some fixed intensity λ. Suppose that S is a metric space. Recall that an S-valued Markov process $X = (X(t) : t \geq 0)$ is said to be *v-geometrically ergodic* if there exists a (measurable) function $v \geq 1$, a probability π on S, $d < \infty$, and $\alpha > 0$ such that

$$\sup_{|g|\leq v}\left|\mathbb{E}_x g(X(t)) - \int_S g(y)\pi(dy)\right| \leq d\, v(x)e^{-\alpha t}, \qquad (1)$$

for $t \geq 0$ and $x \in S$, where $\mathbb{E}_x(\cdot) \triangleq \mathbb{E}(\cdot|X(0) = x)$; see Down et al. (1995) for sufficient conditions assuring such geometric ergodicity. Stable queueing networks in which the service time RVs have exponential moments typically are v-geometrically ergodic for some v; see Kumar and Meyn (1995) and Dai and Meyn (1995).

We assume the following.

Assumption 2. *$\lambda(t) = f(X(t))$ for some bounded continuous $f: S \to \mathbb{R}_+$, where X is φ-irreducible and v-geometrically ergodic.*

Observe that the v-geometric ergodicity implies that $t^{-1}\int_0^t \lambda(s)\,ds \to \lambda^* := \int f(x)\pi(dx)$ (see Down et al. 1995; Meyn and Tweedie 2009, chapter 13), where π is the stationary distribution of X, and Assumption 1 is satisfied. Therefore, the cumulative intensity process A_ϵ converges to $\lambda^* t$ in distribution and $N_\epsilon \Rightarrow N_0$ as $\epsilon \downarrow 0$.

To state our next result on the total variation distance between N_ϵ and N_0, we let $X_1(\infty), X_2(\infty), \ldots$ be an i.i.d. sequence of S-valued RVs having common distribution π (independent of N_0).

Theorem 2. *Suppose Assumption 2 holds and $\mathbb{E}f(X_1(\infty)) > 0$. Then,*

$$\begin{aligned}\sup_A &|P((N_\epsilon(s) : 0 \leq s \leq t) \in A) \\ &- P((N_0(s) : 0 \leq s \leq t) \in A)| \\ &\to \frac{1}{2}\mathbb{E}\left|\prod_{j=1}^{N_0(t)} \frac{f(X_j(\infty))}{\mathbb{E}f(X_1(\infty))} - 1\right|,\end{aligned}$$

as $\epsilon \downarrow 0$, where the supremum is taken over the Borel subsets of $D[0, t]$.

Proof of Theorem 2. The change-of-measure formula for doubly stochastic Poisson processes (see, e.g., equation (3.2) of Brémaud 1981, p. 241) asserts that

$$\begin{aligned}P&((N_\epsilon(s) : 0 \leq s \leq t) \in A) \\ &= \mathbb{E}I((N_0(s) : 0 \leq s \leq t) \in A)\exp\left(-\int_0^t \tilde{\lambda}_\epsilon(s)ds\right) \\ &\cdot \prod_{j=1}^{N_0(t)}\left(\frac{\lambda_\epsilon(T_j)}{\lambda^*}\right),\end{aligned}$$

where $T_1, T_2, \ldots$ are the consecutive jump times of N_0, $\lambda^* = \mathbb{E}f(X_1(\infty))$, N_0 is a Poisson process with constant rate λ^* under P, and $\tilde{\lambda}_\epsilon(s) = \lambda_\epsilon(s) - \lambda^*$. It follows that (see, e.g., Gibbs and Su 2002, p. 7)

$$\begin{aligned}\sup_A &|P((N_\epsilon(s) : 0 \leq s \leq t) \in A) - P((N_0(s) : 0 \leq s \leq t) \in A)| \\ &= \frac{1}{2}\mathbb{E}|\exp\left(-\int_0^t \tilde{\lambda}_\epsilon(s)ds\right)\prod_{j=1}^{N_0(t)}\left(\frac{\lambda_\epsilon(T_j)}{\lambda^*}\right) - 1|.\end{aligned} \qquad (2)$$

Let $\mathcal{H}$ be the σ-algebra generated by $T_1, T_2, \ldots, T_{N_0(t)}, N_0(t)$. Conditional on $\mathcal{H}$, Assumption 2 implies that

$$\begin{aligned}P(&\lambda_\epsilon(T_i) \leq x_i,\ 1 \leq i \leq N_0(t)\,|\,\mathcal{H}) \\ &= \mathbb{E}\big(I\big(f(X(T_i/\epsilon)) \leq x_i,\ 1 \leq i \leq N_0(t) - 1\big) \\ &\quad \times P\big(f\big(X\big(T_{N_0(t)}/\epsilon\big)\big) \leq x_{N_0(t)}\,|\,X\big(T_{N_0(t)-1}/\epsilon\big)\big)\,|\,\mathcal{H}\big).\end{aligned}$$

Because $I(f(\cdot) \leq y)$ is bounded from above by v, Assumption 2 ensures that

$$\begin{aligned}p_\epsilon(s, x, y) &\triangleq P\big(f(X(s/\epsilon)) \leq y\,|\,X(0) = x\big) \\ &\to P\big(f(X(\infty)) \leq y\big),\end{aligned}$$

as $\epsilon \downarrow 0$, so that

$$\begin{aligned}|\,&P(\lambda_\epsilon(T_i) \leq x_i,\ 1 \leq i \leq N_0(t)\,|\,\mathcal{H}) - P \\ &\times(\lambda_\epsilon(T_i) \leq x_i,\ 1 \leq i \leq N_0(t) - 1)\,|\,\mathcal{H})P \\ &\times\big(f(X(\infty)) \leq x_{N_0(t)}\big)\,| \\ &= |\,\mathbb{E}(I(\lambda_\epsilon(T_i)) \leq x_i,\ 1 \leq i \leq N_0(t) - 1)\big(p_\epsilon \\ &\times\big(T_{N_0(t)} - T_{N_0(t)-1}, X\big(T_{N_0(t)-1}/\epsilon\big), x_{N_0(t)}\big) \\ &- P\big(f(X(\infty)) \leq x_{N_0(t)}\big)\big)|\mathcal{H})| \\ &\leq \mathbb{E}|\,p_\epsilon\big(T_{N_0(t)} - T_{N_0(t)-1}, X\big(T_{N_0(t)-1}/\epsilon\big), x_{N_0(t)}\big) \\ &- P\big(f(X(\infty)) \leq x_{N_0(t)}\big)\,| \to 0,\end{aligned}$$

as $\epsilon \downarrow 0$. We now repeat this argument $N_0(t) - 1$ additional times, thereby yielding

$$\begin{aligned}P\big(&\lambda_\epsilon(T_i) \leq x_i,\ 1 \leq i \leq N_0(t)\,|\,T_1, T_2, \ldots, T_{N_0(t)}, N_0(t)\big) \\ &\to \prod_{i=1}^{N_0(t)} P\big(f(X_i(\infty)) \leq x_i\big),\end{aligned}$$

as $\epsilon \downarrow 0$. Hence, conditional on $T_1, \ldots, T_{N_0(t)}, N_0(t)$,

$$\begin{aligned}&\big(\lambda_\epsilon(T_1), \lambda_\epsilon(T_2), \ldots, \lambda_\epsilon\big(T_{N_0(t)}\big)\big) \\ &\quad\Rightarrow \big(f(X_1(\infty)), f(X_2(\infty)), \ldots, f\big(X_{N_0(t)}(\infty)\big)\big),\end{aligned} \qquad (3)$$

as $\epsilon \downarrow 0$.

The proof of Theorem 3 establishes that $\mathbb{E}(\int_0^t \tilde{\lambda}_\epsilon(s)ds)^2 \to 0$ as $\epsilon \downarrow 0$; see (14). Chebyshev's inequality therefore implies that

$$\int_0^t \tilde{\lambda}_\epsilon(s)ds \Rightarrow 0, \qquad (4)$$

as $\epsilon \downarrow 0$. Relations (3) and (4) yield the conclusion that

$$\exp\left(-\int_0^t \tilde{\lambda}_\epsilon(s)ds\right)\prod_{j=1}^{N_0(t)}\left(\frac{\lambda_\epsilon(T_j)}{\lambda^*}\right) \Rightarrow \prod_{j=1}^{N_0(t)}\left(\frac{f(X_j(\infty))}{\mathbb{E}f(X_1(\infty))}\right),$$

as $\epsilon \downarrow 0$.

Finally,

$$\left|\exp\left(-\int_0^t \tilde{\lambda}_\epsilon(s)ds\right)\prod_{j=1}^{N_0(t)}\left(\frac{\lambda_\epsilon(T_j)}{\lambda^*}\right) - 1\right| \le 1 + \exp(\|f\|t)\left(\frac{\|f\|}{\lambda^*}\right)^{N_0(t)},$$

where $\|f\| \triangleq \max\{|f(x) : x \in S|\}$, so that the integrand of the right-hand side of (2) is bounded uniformly in ϵ by an integrable RV. Consequently, the dominated convergence theorem applies to the right-hand side of (2), yielding the theorem. □

It is evident that N_ϵ does not converge to N_0 in total variation, because of the rapid fluctuations in the intensity λ_ϵ at any $\epsilon > 0$. However, these rapid fluctuations are "smoothed out" by path functionals that are suitably continuous, yielding the weak convergence associated with Theorem 1.

Example 3. Let $(N_\epsilon(t) : t \ge 0)$, with $1 \ge \epsilon > 0$, be a counting process with a periodic intensity function

$$\lambda^*_\epsilon(t + k\epsilon) = \begin{cases} 1 & t \in [0, \epsilon/2), \\ 0 & t \in [\epsilon/2, \epsilon), \end{cases} \tag{5}$$

for $k \ge 1$. Notice that the corresponding compensator satisfies Assumption 1. Observe that the intensity undergoes a "rapid" shift from a positive level to zero halfway through the period. Consequently, this counting process will not converge to a Poisson process in total variation, because there are deterministic intervals of time where there can be no points.

3. An Asymptotic Refinement for the Distribution of $N_\epsilon(t)$

In this section, we show how the approximation of Theorem 1 can be improved via a "first order" refinement that reflects the impact of the high-frequency fluctuations. Recall that $o(a(\epsilon))$ represents a function of ϵ such that $o(a(\epsilon))/(a(\epsilon)) \to 0$ as $\epsilon \downarrow 0$. Also, for a bounded (measurable) function f on S, note that v-geometric ergodicity guarantees that if $f_c(x) = f(x) - \mathbb{E}f(X(\infty))$, then

$$|\mathbb{E}_x f_c(X(t))| \le \|f\| d\, v(x) e^{-\alpha t}, \tag{6}$$

and hence the integral defining

$$g(x) \triangleq \int_0^\infty \mathbb{E}_x f_c(X(t))dt,$$

converges absolutely and is bounded by a multiple of v. In fact, the function g defined above is the solution to Poisson's equation for the function f_c, that is, $-(\mathcal{A}g)(x) = f_c(x)$, in which $\mathcal{A}$ is the generator of the Markov process X (for details, see Glynn and Meyn 1996; Meyn and Tweedie 2009, chapter 17.4). Note that the solution of Poisson's equation is fundamental to the analysis of the additive functional of $\int_0^t f(X(s))ds$, which in our case is the compensator of the counting process. In particular, the process constructed by $M(t) \triangleq g(X(t)) + \int_0^t f_c(X(s))ds$ is a martingale.

Theorem 3. *Suppose Assumption 2 holds with* $\mathbb{E}f(X(\infty)) > 0$. *If* $\lambda_\epsilon(t) = f(X(t/\epsilon))$, *then*

$$P(N_\epsilon(t) = k) = P(N_0(t) = k)\Bigg(1 + \epsilon\Bigg[\left(\frac{k}{\lambda^* t} - 1\right)g(x) + \frac{1}{2}\left(1 - \frac{2k}{\lambda^* t} + \frac{k(k-1)}{(\lambda^* t)^2}\right)\sigma^2 t\Bigg] + o(\epsilon)\Bigg),$$

as $\epsilon \downarrow 0$, *where* $\sigma^2 = 2\mathbb{E}f_c(X(\infty))g(X(\infty))$.

Note that the variance term σ^2 emerges as the limiting variance in the martingale central limit theorem as applied to $M(t)$. In fact, there are a number of models where Poisson's equation can be explicitly solved and σ^2 can be explicitly computed, for example, birth-and-death stochastic processes; see Whitt (1992) for details.

Proof of Theorem 3. If we condition on X, we find that

$$P_x(N_\epsilon(t) = k) = \mathbb{E}_x \exp\left(-\int_0^t \lambda_\epsilon(s)ds\right)\frac{\left(\int_0^t \lambda_\epsilon(s)ds\right)^k}{k!}.$$

Set $h_k(y) = e^{-y}y^k/k!$, and note that for $y > 0$,

$$h_k^{(1)}(y) = h_k(y)\left(\frac{k}{y} - 1\right),$$
$$h_k^{(2)}(y) = h_k(y)\left(1 - \frac{2k}{y} + \frac{k(k-1)}{y^2}\right),$$
$$h_k^{(3)}(y) = h_k(y)\left(\frac{k(k-1)(k-2)}{y^3} - \frac{3k(k-1)}{y^2} + \frac{3k}{y} - 1\right).$$

Hence, a Taylor expansion of h_k about $t\mathbb{E}f(X(\infty))$ implies that

$$\begin{aligned} h_k&\left(\int_0^t \lambda_\epsilon(s)ds\right) \\ &= h_k\left(\epsilon\int_0^{t/\epsilon} f(X(s))ds\right) \\ &= h_k(t\mathbb{E}f(X(\infty))) + h_k^{(1)}(t\mathbb{E}f(X(\infty))) \\ &\quad\times\left(\epsilon\int_0^{t/\epsilon} f_c(X(s))ds\right) \\ &\quad+ \frac{h_k^{(2)}(t\mathbb{E}f(X(\infty)))}{2}\left(\epsilon\int_0^{t/\epsilon} f_c(X(s))ds\right)^2 \\ &\quad+ \frac{h_k^{(3)}(\xi(\epsilon))}{6}\left(\epsilon\int_0^{t/\epsilon} f_c(X(s))ds\right)^3, \end{aligned} \tag{7}$$

where $\xi(\epsilon)$ lies between $\int_0^t \lambda_\epsilon(s)ds$ and $t\mathbb{E}f(X(\infty))$.

Note that (6) implies that

$$\mathbb{E}_x \int_0^{t/\epsilon} f_c(X(s))ds = \int_0^{t/\epsilon} \mathbb{E}_x f_c(X(s))ds = g(x) + o(1)v(x), \tag{8}$$

as $\epsilon \downarrow 0$. Also, the Markov property implies that

$$\begin{aligned}\epsilon\mathbb{E}_x\left(\int_0^{t/\epsilon} f_c(X(s))ds\right)^2 &= 2\epsilon\int_0^{t/\epsilon}\int_s^{t/\epsilon}\mathbb{E}_x f_c(X(s))f_c(X(u))duds\\ &= 2\epsilon\int_0^{t/\epsilon}\mathbb{E}_x f_c(X(s))\int_0^{\infty}\mathbb{E}_x[f_c(X(s+u))|X(s)]duds\\ &\quad - 2\epsilon\int_0^{t/\epsilon}\mathbb{E}_x f_c(X(s))\int_0^{\infty}\mathbb{E}_x[\mathbb{E}_x[f_c(X(t/\epsilon+u))|\\ &\quad \times X(t/\epsilon)]|X(s)]duds\\ &= 2\epsilon\int_0^{t/\epsilon}\mathbb{E}_x f_c(X(s))g(X(s))ds - 2\epsilon\\ &\quad \times \int_0^{t/\epsilon}\mathbb{E}_x f_c(X(s))g(X(t/\epsilon))ds.\end{aligned} \tag{9}$$

Because f and g are each bounded by a multiple of v, it follows that fg is bounded by a multiple of v, so that (1) implies that

$$\begin{aligned}\epsilon\int_0^{t/\epsilon}&\mathbb{E}_x f_c(X(s))g(X(s))ds\\ &= t\mathbb{E}f_c(X(\infty))g(X(\infty)) + o(1),\end{aligned} \tag{10}$$

as $\epsilon \downarrow 0$. Also,

$$\begin{aligned}\epsilon\int_0^{t/\epsilon}&\mathbb{E}_x f_c(X(s))g(X(t/\epsilon))ds\\ &= \epsilon\int_0^{t/\epsilon-\epsilon^{-1/2}}\mathbb{E}_x f_c(X(s))\mathbb{E}_x[g(X(t/\epsilon))|X(s)]ds\\ &\quad + \epsilon\,\mathbb{E}_x\int_{t/\epsilon-\epsilon^{-1/2}}^{t/\epsilon} f_c(X(s))g(X(t/\epsilon))ds.\end{aligned} \tag{11}$$

Because $\mathbb{E}g(X(\infty)) = 0$, (1) implies that

$$|\mathbb{E}_x[g(X(t/\epsilon))|X(s)]| \le \|f\|d\,e^{-\alpha(t/\epsilon-s)}v(X(s)),$$

so that

$$\begin{aligned}&\left|\epsilon\int_0^{t/\epsilon-\epsilon^{-1/2}}\mathbb{E}_x f_c(X(s))\mathbb{E}_x[g(X(t/\epsilon))|X(s)]ds\right|\\ &\le \|f\|^2 d\,\epsilon\, e^{-\alpha\epsilon^{-1/2}}\int_0^{t/\epsilon}\mathbb{E}_x v(X(s))ds\\ &= \|f\|^2 d\,t\,e^{-\alpha\epsilon^{-1/2}}\mathbb{E}v(X(\infty)) + o(1)v(x)\\ &= o(1)v(x),\end{aligned} \tag{12}$$

as $\epsilon \downarrow 0$. Furthermore, (1) and the boundedness of f ensure that

$$\begin{aligned}&\left|\epsilon\mathbb{E}_x\int_{t/\epsilon-\epsilon^{-1/2}}^{t/\epsilon} f_c(X(s))g(X(t/\epsilon))ds\right| \le \epsilon^{\frac{1}{2}}\|f\|\mathbb{E}_x g(X(t/\epsilon))\\ &\quad = o(1)v(x),\end{aligned} \tag{13}$$

as $\epsilon \downarrow 0$, and consequently, (9) through (13) yield

$$\epsilon\mathbb{E}_x\left(\int_0^{t/\epsilon} f_c(X(s))ds\right)^2 = 2t\,\mathbb{E}f_c(X(\infty))g(X(\infty)) + o(1)v(x), \tag{14}$$

as $\epsilon \downarrow 0$, because of Assumption 2.

Finally, note that for $y \ge 0$,

$$\begin{aligned}\left|h_k^{(3)}(y)\right| &= \left|\frac{1}{k!}e^{-y}y^{k-3}\right|[-y^3 + 3ky^2 - 3k(k-1)y + k\\ &\qquad \times (k-1)(k-2)]\\ &\le \frac{(y\vee 1)^k}{k!}(1 + 3k + 3k(k-1) + k(k-1)(k-2))\\ &\le \frac{8(y\vee 1)^k}{(k-3)!}I(k\ge 3) + 8(y\vee 1)^k I(k\le 2),\end{aligned}$$

where $y \vee 1 \triangleq \max(y,1)$. Because f is bounded, it is evident that $h^{(3)}(\xi(\epsilon))$ is a bounded RV. Given (7), our theorem follows if we prove that

$$\epsilon^2\mathbb{E}_x\left(\int_0^{t/\epsilon} f_c(X(s))ds\right)^3 = o(1), \tag{15}$$

as $\epsilon \downarrow 0$. But (14) implies that

$$\begin{aligned}\epsilon^2&\mathbb{E}_x\left(\int_0^{t/\epsilon} f_c(X(s))ds\right)^3\\ &= 6\epsilon^2\int_0^{t/\epsilon-\epsilon^{-1/2}}\mathbb{E}_x f_c(X(s_1))\int_{s_1}^{t/\epsilon} f_c(X(s_2))\\ &\quad \times \int_{s_2}^{t/\epsilon} f_c(X(s_3))\,ds_3ds_2ds_1\\ &\quad + \epsilon^2\mathbb{E}_x\left(\int_{t/\epsilon-\epsilon^{-1/2}}^{t/\epsilon} f_c(X(s))ds\right)^3\\ &= 6\epsilon\int_0^{t/\epsilon-\epsilon^{-1/2}}\mathbb{E}_x f_c(X(s))[(t-\epsilon s)\mathbb{E}f_c(X(\infty))g\\ &\quad \times (X(\infty)) + \epsilon o(1)v(X(s))]ds\\ &\quad + \epsilon^2\mathbb{E}_x\left(\int_{t/\epsilon-\epsilon^{-1/2}}^{t/\epsilon} f_c(X(s))ds\right)^3,\end{aligned} \tag{16}$$

where the term $o(1)$ holds uniformly over $0 \le s \le t/\epsilon - \epsilon^{-1/2}$. The boundedness of f implies that

$$\epsilon^2\mathbb{E}_x\left(\int_{t/\epsilon-\epsilon^{-1/2}}^{t/\epsilon} f_c(X(s))ds\right)^3 \le \epsilon^{1/2}\|f\|^3 \to 0, \tag{17}$$

as $\epsilon \downarrow 0$. On the other hand, because (6) implies that

$$\int_0^\infty |\mathbb{E}_x f_c(X(s))|(1+s)ds < \infty,$$

we conclude that

$$\epsilon \int_0^{t/\epsilon - \epsilon^{-1/2}} |\mathbb{E}_x f_c(X(s))|(t - \epsilon s)ds \to 0, \tag{18}$$

as $\epsilon \downarrow 0$. Also,

$$\left|\epsilon \int_0^{t/\epsilon-\epsilon^{-1/2}} o(1)\mathbb{E}_x v(X(s)) f_c(X(s))ds\right|$$
$$\le o(1)\|f\|\epsilon \int_0^{t/\epsilon} \mathbb{E}_x v(X(s))ds$$
$$= o(1)\|f\|t\mathbb{E}v(X(\infty))(1+o(1)) \to 0,$$

as $\epsilon \downarrow 0$, proving (15) in view of (16), (17), and (18), and thereby establishing the theorem. □

A similar (but easier) calculation follows in the deterministic periodic setting in which $\lambda(\cdot)$ is deterministic with period 1, say. In this case,

$$P(N_\epsilon(t) = k) = P(N_0(t) = k)\left(1 + \epsilon\left(\frac{k}{\lambda^* t} - 1\right)\right.$$
$$\left. \times \int_{\lfloor t/\epsilon \rfloor}^{t/\epsilon} (\lambda(s) - \lambda^*)ds + o(\epsilon)\right),$$

as $\epsilon \downarrow 0$, where $\lambda^* = \int_0^1 \lambda(r)dr$, and $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x.

4. An Asymptotic Refinement for Infinite-Server Queues

In this section, we study our Poisson approximation (and its associated first order "error correction") in the setting of the infinite-server queue. Assume that the system starts empty at $t = 0$ and that the service times $V_1, V_2, \dots$ assigned to arriving consecutive customers are i.i.d. and independent of N_ϵ. Our goal in this section is to study the number-in-system process $Q_\epsilon = (Q_\epsilon(t) : t \ge 0)$ when Q_ϵ has arrival process N_ϵ and service time sequence $V = (V_n : n \ge 1)$ so that

$$Q_\epsilon(t) = \sum_{i=0}^{N_\epsilon(t)} 1_{\{T_i^{(\epsilon)} + V_i > t\}}, \tag{19}$$

where $t > 0$ and $T_1^{(\epsilon)}, T_2^{(\epsilon)}, \dots$ are the arrival epochs for the arrival process N_ϵ. Let $Q_0 = (Q_0(t) : t \ge 0)$ be the number-in-system process associated with the constant rate Poisson process N_0 and the same service time sequence V. Our main result in this section is our next theorem.

Theorem 4. *Assume Assumption 2, and suppose that f is bounded (and measurable) with $\mathbb{E}f(X(\infty)) > 0$. Suppose V_1 has a density $h = (h(x) : x \ge 0)$, and set $\bar{K}(x) \triangleq P(V_1 > x)$. If $\lambda_\epsilon(t) = f(X(t/\epsilon))$, then*

$$P(Q_\epsilon(t) = k)$$
$$= P(Q_0(t) = k)\left(1 + \epsilon\left[\left(\frac{k}{\mathbb{E}Q_0(t)} - 1\right)g(x)\bar{K}(t)\right.\right.$$
$$\left.\left. + \frac{1}{2}\left(1 - \frac{2k}{\mathbb{E}Q_0(t)} + \frac{k(k-1)}{(\mathbb{E}Q_0(t))^2}\right)\eta^2\right] + o(\epsilon)\right),$$

where $\eta^2 = 2\sigma^2 \int_0^t \bar{K}(s)h(s)s\,ds + \sigma^2 t\bar{K}(t)^2$, and σ^2 and g are as in Section 3.

Proof of Theorem 4. The argument closely follows that of Theorem 3. Because $Q_\epsilon(t)$ is, conditional on X, Poisson distributed (see Massey and Whitt 1993), it follows that

$$P(Q_\epsilon(t) = k) = \mathbb{E}_x \exp\left(-\int_0^t \lambda_\epsilon(s)\bar{K}(t-s)ds\right)$$
$$\cdot \left(\int_0^t \lambda_\epsilon(s)\bar{K}(t-s)ds\right)^k \frac{1}{k!}.$$

As in Theorem 3, we now Taylor expand $h_k(\cdot)$. In this setting, we expand about $\mathbb{E}Q_0(t)$. It follows that the first order term here is $h_k^{(1)}(\mathbb{E}Q_0(t))$ multiplied by

$$\int_0^t [\lambda_\epsilon(s) - \lambda^*]\bar{K}(t-s)ds$$
$$= \int_0^t f_c(X(s/\epsilon)) \int_{t-s}^\infty h(u)du\,ds$$
$$= \int_0^\infty \int_0^t I(s > t-u) f_c(X(s/\epsilon))ds\,h(u)du$$
$$= \int_0^\infty h(u)[\epsilon A_c(t/\epsilon) - \epsilon A_c((t-u)/\epsilon)]du,$$

where $A_c(r) = 0$ for $r \le 0$ and $A_c(r) = \int_0^r f_c(X(s))ds$ for $r \ge 0$. We note that

$$\mathbb{E}_x A(t/\epsilon) - \mathbb{E}_x A((t-u)/\epsilon) \to \begin{cases} 0, & 0 \le u \le t, \\ g(x), & u > t, \end{cases}$$

as $\epsilon \downarrow 0$, uniformly in $u \le t$. Accordingly,

$$\epsilon\mathbb{E}_x \int_0^t [\lambda_\epsilon(s) - \lambda^*]\bar{K}(t-s)ds = \epsilon g(x)\bar{K}(t)(1+o(1)),$$

as $\epsilon \downarrow 0$.

As for the second derivative term, we are led to the consideration of

$$\epsilon\mathbb{E}_x\left(\int_0^\infty h(u)[A_c(t/\epsilon) - A_c((t-u)/\epsilon)]du\right)^2$$
$$= 2\epsilon \int_0^\infty h(u_1) \int_{u_1}^\infty h(u_2)\mathbb{E}_x[(A_c(t/\epsilon)$$
$$- A_c((t-u_1)/\epsilon)(A_c(t/\epsilon) - A_c((t-u_2)/\epsilon)]du_2du_1. \tag{20}$$

Note that for $0 \le u_1 \le u_2 \le t$,

$$\begin{aligned}
&\epsilon \mathbb{E}_x(A_c(t/\epsilon) - A_c((t-u_1)/\epsilon)(A_c((t-u_1)/\epsilon) \\
&\quad - A_c((t-u_2)/\epsilon)) \\
&\quad = \epsilon \int_{(t-u_2)/\epsilon}^{(t-u_1)/\epsilon} \int_{(t-u_1)/\epsilon}^{t/\epsilon} I(|s_1 - s_2| \le \epsilon^{-1/2}) \mathbb{E}_x f_c \\
&\qquad \times (X(s_1)) f_c(X(s_2)) ds_1 ds_2 \\
&\qquad + \epsilon \int_{(t-u_2)/\epsilon}^{(t-u_1)/\epsilon} \int_{(t-u_1)/\epsilon}^{t/\epsilon} I(|s_1 - s_2| > \epsilon^{-1/2}) \mathbb{E}_x f_c \\
&\qquad \times (X(s_1)) f_c(X(s_2)) ds_1 ds_2. \qquad (21)
\end{aligned}$$

The first term on the right-hand side of (21) can be upper bounded by

$$\begin{aligned}
&\epsilon \|f\|^2 \int_{(t-u_2)/\epsilon}^{(t-u_1)/\epsilon} \int_{(t-u_1)/\epsilon}^{t/\epsilon} I(|s_1 - s_2| \le \epsilon^{-1/2}) ds_1 ds_2 \\
&\quad = O(\epsilon^{1/2}) \to 0,
\end{aligned}$$

as $\epsilon \downarrow 0$. For the second term, we use (6) to obtain the upper bound

$$\begin{aligned}
&\epsilon \int_{(t-u_2)/\epsilon}^{(t-u_1)/\epsilon} \int_{(t-u_1)/\epsilon}^{t/\epsilon} I(|s_1 - s_2| > \epsilon^{-1/2}) \mathbb{E}_x f_c(X(s_2)) \\
&\quad \times O(\mathbb{E}_x v(X(s_2))) e^{-\alpha(s_1 - s_2)} ds_1 ds_2 \\
&\quad \le \epsilon e^{-\alpha\epsilon^{-1/2}} \|f\| \int_{(t-u_2)/\epsilon}^{(t-u_1)/\epsilon} \int_{(t-u_1)/\epsilon}^{t/\epsilon} O(\mathbb{E}_x v(X(s_2))) \\
&\qquad \times ds_1 ds_2 \to 0,
\end{aligned}$$

as $\epsilon \downarrow 0$. Consequently, (20) equals

$$\begin{aligned}
&2\epsilon \int_0^\infty h(u_1) \int_{u_1}^\infty h(u_2) \mathbb{E}_x (A_c(t/\epsilon) - A_c \\
&\quad \times ((t-u_1)/\epsilon))^2 du_2 du_1 + o(1),
\end{aligned}$$

as $\epsilon \downarrow 0$. But (14) proves that

$$\epsilon \mathbb{E}_x (A_c(t/\epsilon) - A_c((t-u)/\epsilon))^2 \to \begin{cases} \sigma^2 u, & 0 \le u \le t, \\ \sigma^2 t, & u > t, \end{cases}$$

uniformly in $0 \le u \le t$. As a consequence, (20) equals $\eta^2 + o(1)$ as $\epsilon \downarrow 0$.

The third derivative term can be handled as in Theorem 3, thereby yielding the proof of the result. □

Acknowledgments

The authors thank the editors and referees for their careful reading of this paper and their help with improving the exposition, particularly as to how the limit regime in this paper compares with that associated with the "rapid switching" literature.

References

Asmussen S, Glynn PW (2007) *Stochastic Simulation: Algorithms and Analysis*, vol. 57 (Springer, New York).

Bacry E, Delattre S, Hoffmann M, Muzy JF (2013) Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes Their Appl.* 123(7):2475–2499.

Billingsley P (1968) *Convergence of Probability Measures* (John Wiley & Sons, New York).

Blom J, Kella O, Mandjes M, Thorsdottir H (2014) Markov-modulated infinite-server queues with general service times. *Queueing Systems* 76(4):403–424.

Brémaud P (1981) *Point Processes and Queues: Martingale Dynamics* (Springer, New York).

Cinlar E, Agnew R (1968) On the superposition of point processes. *J. Royal Statist. Soc. B* 30(3):576–581.

Dai JG, Meyn SP (1995) Stability and convergence of moments for multiclass queueing networks via fluid limit models. *IEEE Trans. Automatic Control* 40(11):1889–1904.

Daley DJ, Vere-Jones D (1988) *An Introduction to the Theory of Point Processes* (Springer, New York).

Daw A, Pender J (2018) Queues driven by Hawkes processes. *Stochastic Systems* 8(3):192–229.

Down D, Meyn SP, Tweedie RL (1995) Exponential and uniform ergodicity of Markov processes. *Ann. Probab.* 23(4):1671–1691.

Ethier SN, Kurtz TG (1986) *Markov Processes, Characterization and Convergence* (John Wiley & Sons, New York).

Gao X, Zhu L (2016) Functional central limit theorems for stationary Hawkes processes and application to infinite-server queues. Preprint, submitted July 22, https://arxiv.org/abs/1607.06624.

Gibbs AL, Su FE (2002) On choosing and bounding probability metrics. *Internat. Statist. Rev.* 70(3):419–435.

Glynn PW, Meyn SP (1996) A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.* 24(2):916–931.

Hawkes AG (1971) Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58(1):83–90.

Heemskerk M, van Leeuwaarden J, Mandjes M (2017) Scaling limits for infinite-server systems in a random environment. *Stochastic Systems* 7(1):1–31.

Kallenberg O (1997) *Foundations of Modern Probability* (Springer, New York).

Khasminskii RZ (2012) Stability of regime-switching stochastic differential equations. *Problems Inform. Transmission* 48(3):259–270.

Khasminskii RZ, Yin G, Zhang Q (1996) Asymptotic expansions of singularly perturbed systems involving rapidly fluctuating Markov chains. *SIAM J. Appl. Math.* 56(1):277–293.

Khasminskii RZ, Zhu C, Yin G (2007) Stability of regime-switching diffusions. *Stochastic Processes Their Appl.* 117(8):1037–1051.

Koops DT, Boxma OJ, Mandjes M (2017) Networks of $\cdot/G/\infty$ queues with shot-noise-driven arrival intensities. *Queueing Systems* 86(3–4):301–325.

Koops DT, Saxena M, Boxma O, Mandjes M (2018) Infinite-server queues with Hawkes input. *J. Appl. Probab.* 55(3):920–943.

Kumar P, Meyn SP (1995) Stability of queueing networks and scheduling policies. *IEEE Trans. Automat. Control* 40(2):251–260.

Landis GA, Bailey SG, Tischler R (2006) Causes of power-related satellite failures. *IEEE 4th World Conf. Photovoltaic Energy Conversion*, vol. 2 (Institute of Electrical and Electronics Engineers, Piscataway, NJ), 1943–1945.

Laub PJ, Taimre T, Pollett PK (2015) Hawkes processes. Preprint, submitted July 10, https://arxiv.org/abs/1507.02822.

Liptser RS (1980) A strong law of large numbers for local martingales. *Stochastics* 3(1–4):217–228.

Massey WA, Whitt W (1993) Networks of infinite-server queues with nonstationary Poisson input. *Queueing Systems* 13(1–3):183–250.

Massey WA, Whitt W (1998) Uniform acceleration expansions for Markov chains with time-varying rates. *Ann. Appl. Probab.* 8(4):1130–1155.

Meyn S, Tweedie RL (2009) *Markov Chains and Stochastic Stability* (Cambridge University Press, Cambridge, UK).

Spreij P, Storm J (2018) Diffusion limits for a Markov modulated counting process. Preprint, submitted January 11, https://arxiv.org/abs/1801.03682.

Whitt W (1992) Asymptotic formulas for Markov processes with applications to simulation. *Oper. Res.* 40(2):279–291.

Whitt W (2018) Time-varying queues. *Queueing Models Service Management* 1(2):79–164.

Zheng Z, Honnappa H, Glynn PW (2018) Approximating performance measures for slowly changing non-stationary Markov chains. Preprint, submitted May 4, https://arxiv.org/abs/1805.01662.

Zeyu Zheng is an assistant professor in the Department of Industrial Engineering and Operations Research at the University of California, Berkeley. He received his PhD in management science and engineering, PhD minor in statistics, and MA in economics from Stanford University, and a BS in mathematics from Peking University. He has done research in simulation, nonstationary stochastic modeling and decision making, data analytics, statistical learning, and over-the-counter financial markets.

Harsha Honnappa is an assistant professor in the School of Industrial Engineering, Purdue University. His research interests are in queueing theory, applied probability, game theory, and statistical modeling. He is also the recipient of the 2016 Lajos Takács Award for outstanding PhD thesis on queueing theory and its applications.

Peter W. Glynn is the Thomas Ford Professor in the Department of Management Science and Engineering at Stanford University. He was the co-winner of the John von Neumann Theory Prize from INFORMS and was elected to the National Academy of Engineering. He was Founding Editor-in-Chief of *Stochastic Systems* and served as Editor-in-Chief of *Journal of Applied Probability* and *Advances in Applied Probability*. His research interests lie in simulation, computational probability, queueing theory, statistical inference for stochastic processes, and stochastic modeling.