Approximating Systems Fed by Poisson Processes with Rapidly Changing Arrival Rates

Zeyu Zheng
Department of Management Science and Engineering
Stanford University, zyzheng@stanford.edu

Harsha Honnappa
School of Industrial Engineering
Purdue University, honnappa@purdue.edu

Peter W. Glynn
Department of Management Science and Engineering
Stanford University, glynn@stanford.edu

July 17, 2018

Abstract

This paper introduces a new asymptotic regime for simplifying stochastic models having non-stationary 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 non-stationary 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.

Key words: point 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 non-stationarities. These non-stationarities 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 non-stationary 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 non-stationary 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 involve a stationary arrival rate calculation. This intuition has been validated rigorously by Khasminskii, Yin, and Zhang (1996), Massey and Whitt (1998), and, more recently, by Zheng, Honnappa, and Glynn (2018).

In this paper, we show that arrival rate modeling also simplifies 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 systems fed by Poisson processes in which the intensity at time t is given by $\lambda(t/\epsilon)$, where ϵ is a small parameter and $\lambda = (\lambda(s) : s \geq 0)$ is a fixed process. The process λ could be a deterministic periodic function, 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. Thus, this paper provides a second rigorously supported asymptotic regime within which the dynamics of a service system with a non-stationary arrival process can be approximated by a simpler system with stationary dynamics. We note that despite the practical importance of such non-stationary models, very few analytical approximations are available for such systems.

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.

This note is organized as follows. Section 2 provides our main weak convergence theorem, establishing that point 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 point process and the Poisson process in the Markov-modulated doubt 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

To construct our point process with a rapidly fluctuating arrival rate, 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 exactly by 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 posesesses a right continuous non-decreasing $\mathcal{F}$-compensator $A = (A(t) : t \geq 0)$, so that $M = (M(t) : t \geq 0)$ is a martingale adapted to $\mathcal{F}$, where

$$M(t) = N(t) - A(t).$$

Note that N need not be a doubly stochastic Poisson process (e.g. N could be a Hawkes process; see Hawkes (1971) for the definition).

For $1 \geq \epsilon > 0$, let $\beta^\epsilon = (\beta_i^\epsilon; i \geq 1)$ be an independent and identically distributed (iid) sequence of Bernoulli(ϵ) random variables (rv's) 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 $\sigma(\beta_i^\epsilon : 1 \leq i \leq N(t/\epsilon))$. Put $\mathcal{G}^\epsilon = (\mathcal{G}_t^\epsilon : t \geq 0)$ 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 point process for which

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

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 last equality (see p.87 of Kallenberg (1997)), and the fact that M is an $\mathcal{F}$-adapted martingale in the second 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 $\lambda_\epsilon(t) = \lambda(t/\epsilon)$. 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:

Assumption 1. *There exists a deterministic $\lambda^* \in (0, \infty)$ such that*

$$\frac{1}{t} A(t) \Rightarrow \lambda^*$$

as $t \to \infty$, where $\Rightarrow$ denotes weak convergence.

Here is our main result of this section. 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. *In the presence of Assumption 1,*

$$N_\epsilon \Rightarrow N_0$$

in $D[0, \infty)$ as $\epsilon \downarrow 0$, where $N_0 = (N_0(t) : t \geq 0)$ is a Poisson process with constant intensity λ^*.

Proof. We note that Assumption 1 implies that for each $t \geq 0$,

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

as $\epsilon \downarrow 0$. We now apply Theorem 13.4.IV of Daley and Vere-Jones (1988) to obtain the result. □

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) a.s., 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 point process N_ϵ is Poisson with rate λ^* (when ϵ is small).

In the remainder of this section, we make clear that while 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 complete separable 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

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

for $t \geq 0$ and $x \in S$, where $\mathbb{E}_x(\cdot) \triangleq \mathbb{E}(\,\cdot\,|X(0) = x)$; see Down, Meyn, and Tweedie (1995) for sufficient conditions assuring such geometric ergodicity.

We assume that:

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

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 iid sequence of S-valued rv's having common distribution π (independent of N_0).

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

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

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

Proof. The change-of-measure formula for doubly stochastic Poisson processes (see, for example, p.241 of Brémaud (1981)) asserts that

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

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, for example, Gibbs and Su (2002))

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

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

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

Since $I(f(\cdot) \le y)$ is upper bounded by v, Assumption 2 ensures that

$$p_\epsilon(s,x,y) \triangleq P(f(X(s/\epsilon)) \le y \,|\, X(0)=x) \to P(f(X(\infty)) \le y)$$

as $\epsilon \downarrow 0$, so that

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

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

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

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

$$(\lambda_\epsilon(T_1), \lambda_\epsilon(T_2), \ldots, \lambda_\epsilon(T_{N_0(t)})) \Rightarrow (f(X_1(\infty)), f(X_2(\infty)), \ldots, f(X_{N_0(t)}(\infty))) \tag{3}$$

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 (13). Chebyshev's inequality threfore implies that

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

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

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

as $\epsilon \downarrow 0$.

Finally,

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

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, due to 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.

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 on S, note that v-geometric ergodicity guarantees that if $f_c(x) = f(x) - \mathbb{E}f(X(\infty))$, then

$$|\mathbb{E}_x f_c(X(t))| \leq \|f\| d\, v(x) e^{-\alpha t} \tag{5}$$

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.

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

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

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

Proof. If we condition on X, we find that

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

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

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

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

Note that (5) implies that

$$\mathbb{E}_x\int_0^{t/\epsilon} f_c(X(s))ds = \int_0^{t/\epsilon}\mathbb{E}_x f_c(X(s))ds = g(x) + o(1)v(x) \tag{7}$$

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

$$\begin{aligned} &\epsilon\mathbb{E}_x\left(\int_0^{t/\epsilon} f_c(X(s))ds\right)^2 \\ &= 2\epsilon\int_0^{t/\epsilon}\int_s^{t/\epsilon}\mathbb{E}_x f_c(X(s))f_c(X(u))duds \\ &= 2\epsilon\int_0^{t/\epsilon}\mathbb{E}_x f_c(X(s))\int_0^\infty \mathbb{E}_x[f_c(X(s+u))|X(s)]duds \\ &\quad - 2\epsilon\int_0^{t/\epsilon}\mathbb{E}_x f_c(X(s))\int_0^\infty \mathbb{E}_x[\mathbb{E}_x[f_c(X(t/\epsilon+u))|X(t/\epsilon)]|X(s)]duds \\ &= 2\epsilon\int_0^{t/\epsilon}\mathbb{E}_x f_c(X(s))g(X(s))ds - 2\epsilon\int_0^{t/\epsilon}\mathbb{E}_x f_c(X(s))g(X(t/\epsilon))ds. \end{aligned} \tag{8}$$

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

$$\epsilon\int_0^{t/\epsilon}\mathbb{E}_x f_c(X(s))g(X(s))ds = t\mathbb{E}f_c(X(\infty))g(X(\infty)) + o(1) \tag{9}$$

as $\epsilon \downarrow 0$. Also,

$$\begin{aligned}&\epsilon \int_0^{t/\epsilon} \mathbb{E}_x f_c(X(s)) g(X(t/\epsilon)) ds \\ &= \epsilon \int_0^{t/\epsilon - \epsilon^{-1/2}} \mathbb{E}_x f_c(X(s)) \mathbb{E}_x[g(X(t/\epsilon))|X(s)] ds + \epsilon\, \mathbb{E}_x \int_{t/\epsilon - \epsilon^{-1/2}}^{t/\epsilon} f_c(X(s)) g(X(t/\epsilon)) ds. \qquad (10)\end{aligned}$$

Since $\mathbb{E} g(X(\infty)) = 0$, (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

$$\begin{aligned}&\left|\epsilon \int_0^{t/\epsilon - \epsilon^{-1/2}} \mathbb{E}_x f_c(X(s)) \mathbb{E}_x[g(X(t/\epsilon))|X(s)] ds\right| \\ &\le \|f\|^2 d\, \epsilon e^{-\alpha \epsilon^{-1/2}} \int_0^{t/\epsilon} \mathbb{E}_x v(X(s)) ds \\ &= \|f\|^2 d\, e^{-\alpha \epsilon^{-1/2}} \mathbb{E} v(X(\infty)) + o(1) v(x) \\ &= o(1) v(x) \qquad (11)\end{aligned}$$

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

$$\left|\epsilon \mathbb{E}_x \int_{t/\epsilon - \epsilon^{-1/2}}^{t/\epsilon} f_c(X(s)) g(X(t/\epsilon)) ds\right| \le \epsilon^{\frac{1}{2}} \|f\| \mathbb{E}_x g(X(t/\epsilon)) = o(1) v(x) \qquad (12)$$

as $\epsilon \downarrow 0$, and consequently, (8) through (12) yield

$$\epsilon\, \mathbb{E}_x \Big(\int_0^{t/\epsilon} f_c(X(s)) ds \Big)^2 = 2t\, \mathbb{E} f_c(X(\infty)) g(X(\infty)) + o(1) v(x) \qquad (13)$$

as $\epsilon \downarrow 0$.

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

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

where $y \vee 1 \triangleq \max(y, 1)$. Since f is bounded, it is evident that $h^{(3)}(\xi(\epsilon))$ is a bounded rv. Given

(6), our theorem follows if we prove that

$$\epsilon^2 \mathbb{E}_x \left(\int_0^{t/\epsilon} f_c(X(s))ds \right)^3 = o(1) \tag{14}$$

as $\epsilon \downarrow 0$. But (13) implies that

$$\begin{aligned}
&\epsilon^2 \mathbb{E}_x \left(\int_0^{t/\epsilon} f_c(X(s))ds \right)^3 \\
&= 6\epsilon^2 \int_0^{t/\epsilon - \epsilon^{-1/2}} \mathbb{E}_x f_c(X(s_1)) \int_{s_1}^{t/\epsilon} f_c(X(s_2)) \int_{s_2}^{t/\epsilon} f_c(X(s_3))\, ds_3 ds_2 ds_1 \\
&\quad + \epsilon^2 \mathbb{E}_x \left(\int_{t/\epsilon - \epsilon^{-1/2}}^{t/\epsilon} f_c(X(s))ds \right)^3 \\
&= 6\epsilon \int_0^{t/\epsilon - \epsilon^{-1/2}} \mathbb{E}_x f_c(X(s))[(t - \epsilon s)\mathbb{E} f_c(X(\infty))g(X(\infty)) + \epsilon o(1) v(X(s))]ds \\
&\quad + \epsilon^2 \mathbb{E}_x \left(\int_{t/\epsilon - \epsilon^{-1/2}}^{t/\epsilon} f_c(X(s))ds \right)^3,
\end{aligned} \tag{15}$$

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

$$\epsilon^2 \mathbb{E}_x \left(\int_{t/\epsilon - \epsilon^{-1/2}}^{t/\epsilon} f_c(X(s))ds \right)^3 \le \epsilon^{1/2} \|f\|^3 \to 0 \tag{16}$$

as $\epsilon \downarrow 0$. On the other hand, (5) implies that

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

so we conclude that

$$\epsilon \int_0^{t/\epsilon - \epsilon^{-1/2}} |\mathbb{E}_x f_c(X(s))|(t - \epsilon s)ds \to 0 \tag{17}$$

as $\epsilon \downarrow 0$. Also,

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

as $\epsilon \downarrow 0$, proving (14) in view of (15), (16), and (17), 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,

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

as $\epsilon\downarrow 0$, where $\lambda^*=\int_0^t\lambda(r)dr$ 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 $V_1, V_2, \ldots$ assigned to arriving consecutive customers are iid and independent of N_ϵ. Our goal in this section is to study the number-in-system process $Q_\epsilon=(Q_\epsilon(t):t\geq 0)$, when Q_ϵ has arrival process N_ϵ and service time sequence $V=(V_n:n\geq 1)$. Let $Q_0=(Q_0(t):t\geq 0)$ 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 $k=(k(x):x\geq 0)$, and set $\bar{K}(x)\triangleq P(V_1>x)$. If $\lambda_\epsilon(t)=f(X(t/\epsilon))$, then*

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

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

Proof. 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

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

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 $u \le t - \sqrt{\epsilon}$. 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

$$\begin{aligned} &\epsilon \mathbb{E}_x \left(\int_0^\infty k(u)[A_c(t/\epsilon) - A_c((t-u)/\epsilon)]du \right)^2 \\ &= 2\epsilon \int_0^\infty k(u_1) \int_{u_1}^\infty k(u_2)\mathbb{E}_x[(A_c(t/\epsilon) - A_c((t-u_1)/\epsilon)(A_c(t/\epsilon) - A_c((t-u_2)/\epsilon)]du_2 du_1. \end{aligned} \tag{18}$$

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

$$\begin{aligned} &\epsilon \mathbb{E}_x (A_c(t/\epsilon) - A_c((t-u_1)/\epsilon)(A_c((t-u_1)/\epsilon) - A_c((t-u_2)/\epsilon)) \\ &= \epsilon \int_{(t-u_2)/\epsilon}^{(t-u_1)/\epsilon} \int_{(t-u_1)/\epsilon}^{t/\epsilon} I(|s_1 - s_2| \le \epsilon^{-1/2})\mathbb{E}_x f_c(X(s_1))f_c(X(s_2))ds_1 ds_2 \\ &\quad + \epsilon \int_{(t-u_2)/\epsilon}^{(t-u_1)/\epsilon} \int_{(t-u_1)/\epsilon}^{t/\epsilon} I(|s_1 - s_2| > \epsilon^{-1/2})\mathbb{E}_x f_c(X(s_1))f_c(X(s_2))ds_1 ds_2. \end{aligned} \tag{19}$$

The first term on the right-hand side of (19) can be upper bounded by

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

as $\epsilon \downarrow 0$. For the second term, we use (5) to obtain the upper bound

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

as $\epsilon \downarrow 0$. Consequently, (18) equals

$$2\,\epsilon \int_0^\infty k(u_1) \int_{u_1}^\infty k(u_2)\mathbb{E}_x(A_c(t/\epsilon) - A_c((t-u_1)/\epsilon))^2 du_2 du_1 + o(1)$$

as $\epsilon \downarrow 0$. But (13) 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, (18) equals $\eta^2 + o(1)$ as $\epsilon \downarrow 0$.

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

References

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

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

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

Ethier, S. N. and T. G. Kurtz (1986). *Markov Processes, Characterization and Convergence.* John Wiley & Sons, New York.

Gibbs, A. L. and F. E. Su (2002). On choosing and bounding probability metrics. *International Statistical Review 70*(3), 419–435.

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

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

Khasminskii, R. Z., G. Yin, and Q. Zhang (1996). Asymptotic expansions of singularly perturbed systems involving rapidly fluctuating Markov chains. *SIAM Journal on Applied Mathematics 56*(1), 277–293.

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

Massey, W. A. and W. Whitt (1998). Uniform acceleration expansions for Markov chains with time-varying rates. *Annals of Applied Probability 8*(4), 1130–1155.

Zheng, Z., H. Honnappa, and P. W. Glynn (2018). Approximating performance measures for slowly changing non-stationary Markov chains. *arXiv preprint, 1805.01662.*