

QUEUES WITH PATH-DEPENDENT ARRIVAL PROCESSES

KERRY FENDICK,* *Johns Hopkins University Applied Physics Laboratory*

WARD WHITT[†],** *Columbia University*

Abstract

We study the transient and limiting behavior of a queue with a Pólya arrival process. The Pólya process is interesting because it exhibits path-dependent behavior, e.g. it satisfies a non-ergodic law of large numbers: the average number of arrivals over time $[0, t]$ converges almost surely to a nondegenerate limit as $t \rightarrow \infty$. We establish a heavy-traffic diffusion limit for the $\sum_{i=1}^n P_i/GI/1$ queue, with arrivals occurring exogenously according to the superposition of n independent and identically distributed Pólya point processes. That limit yields a tractable approximation for the transient queue-length distribution, because the limiting net input process is a Gaussian Markov process with stationary increments. We also provide insight into the long-run performance of queues with path-dependent arrival processes. We show how Little's law can be stated in this context, and we provide conditions under which there is stability for a queue with a Pólya arrival process.

Keywords: path-dependent stochastic processes; generalized Pólya process; Gaussian Markov process; diffusion approximations; queues; heavy-traffic limit

2010 Mathematics Subject Classification: Primary 60K25

Secondary 60F17; 90B22

1. Introduction

In almost all queueing models, the impact of initial conditions dissipates as time evolves. Thus, for stationary models interest usually centers on the steady-state distribution and convergence to it for various initial conditions. The asymptotic loss of memory (appropriately defined) is also anticipated in queueing models with time-varying arrival rates, as evidenced by results for the time-varying $G_t/M_t/s_t + GI_t$ many-server fluid model in [29] and the weak ergodicity results for nonhomogeneous Markov chains in [26, Chapter V]. In contrast, here we consider a queueing model in which the long-run behavior of the arrival process depends critically on the early history of the arrival process.

Such a stochastic process is said to exhibit path-dependent behavior. There has long been substantial interest in systems with path-dependent behavior, as can be seen from [1, 2] and the citations of them. The classic example is the familiar Pólya urn model discussed in [14], first studied in [34]. There is an urn containing r red balls and b blue balls. At each step,

Received 4 February 2020; revision received 27 October 2020.

* Postal address: Communications Systems Branch, Johns Hopkins University Applied Physics Laboratory, Laurel, MD 20723, USA. Email address: Kerry.Fendick@jhuapl.edu

** Postal address: Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027-6699, USA. Email address: ww2040@columbia.edu

© The Author(s), 2021. Published by Cambridge University Press on behalf of Applied Probability Trust.

we select one ball from the urn at random and then return that ball and one new ball of the selected color to the urn. The proportion of balls of any given color has a path-dependent limit, converging almost surely to a random limit, which has the beta distribution, depending on the parameters r and b . The different converging paths depending on the early history are shown in [2, Figure 1]. (Figure 1 in Section 2 here is an analog.)

It is natural for queues to arise in the allocation of scarce resources in systems with path-dependent behavior. Thus we are motivated to consider a queue with a path-dependent arrival process. Hence we consider the $\sum_{i=1}^n P_i/GI/1$ queue, which is a single-server queue with unlimited waiting space, the first-come first-served service discipline and independent and identically distributed (i.i.d.) service times with a general distribution, with arrivals according to the superposition of n i.i.d. Pólya point processes. A Pólya point process can be represented as a limit of Pólya urn models as indicated in [14, p. 480]. Theorem 1 shows that the Pólya point process is a stationary point process, while Theorem 3 shows that it satisfies a non-ergodic law of large numbers (LLN), which we use as our definition of path dependence. Proposition 3 (from [8]) shows that the superposition process is a special generalized Pólya process as considered by [7, 27], which inherits those properties.

We establish several results for queues with Pólya point process arrival processes. Our main contribution here is Theorem 5, which establishes a heavy-traffic limit that provides, via Corollary 6, a tractable description of the transient queue length distribution in the $\sum_{i=1}^n P_i/GI/1$ queueing model when the service-time distribution has finite second moment, exposing the performance impact of the path-dependent behavior of the arrival process. That follows from Theorem 4, which establishes a functional central limit theorem (FCLT) for the $\sum_{i=1}^n P_i$ superposition process, showing that the limit is a Gaussian Markov process with stationary increments (Ψ -GMP, where Ψ is a mnemonic for SI to denote stationary increments) studied in [16].

In Section 2 we place our superposition process in the context of generalized Pólya processes, as developed in [7, 8, 27]. In Theorems 1 and 2 we identify generalized Pólya processes that are stationary point processes and show how this class can be used to represent any generalized Pólya process. In Section 3 we establish the FCLT for the superposition process and state various consequences. In Section 4 we establish the FCLT with an extra drift and exhibit some striking properties of the Ψ -GMP with drift, further exposing the path-dependent behavior. In Section 5 we establish the associated heavy-traffic limit for the $\sum_{i=1}^n P_i/GI/1$ queue.

Afterwards, we provide additional results and discussion. In Section 6 we obtain stability results for the single-server queue and then establish steady-state results for queues with Pólya arrival processes when either (i) there are infinitely many servers or (ii) there is a single server with an adaptive rate-matching service-rate control in the spirit of [37]. In Section 7 we present some remaining technical details. In Section 8 we present conclusions and discussion.

We conclude this introduction by discussing related work. First, we note that queues with Pólya arrival processes have been previously considered as a way to capture exceptional variability [31, 32, 33]; we discuss that earlier work in Remark 2. Second, the heavy-traffic limit of the $\sum_{i=1}^n P_i/GI/1$ queue can be regarded as a Gaussian queue with a net input process that is a Ψ -GMP with drift. Thus, this paper is related to the large literature on Gaussian queues, which can be seen from [10, 11]. The FCLT here yields a Ψ -GMP with positive dependence, i.e. in which the increments over disjoint intervals are positively correlated; see Corollary 4. As we indicate in Section 8, similar limits hold for processes with negative dependence. That leads to convergence of empirical processes to the Brownian bridge and related queueing heavy-traffic limits as in [22, 25].

2. Generalized Pólya point process with stationary increments: Ψ -GPP

The Pólya point process has been extended to the generalized Pólya process (GPP) [7, 27]. A GPP $N \equiv \{N(t) : t \geq 0\}$ is a Markov point process with stochastic intensity (defined in terms of the internal histories \mathcal{H}_t ; see, e.g., [3, Section 1.8]) by

$$\lambda(t) \equiv \lambda(t | \mathcal{H}_t) \equiv (\gamma N(t-) + \beta)\kappa(t), \quad (1)$$

where $N(0) = 0$, γ and β are positive constants, $\kappa(t)$ is a positive integrable real-valued function and \equiv denotes equality by definition. The classical Pólya point process is the special case of (1) with $\beta = 1$ and

$$\kappa(t) = \frac{1}{\gamma t + 1}, \quad t \geq 0. \quad (2)$$

Many properties of the GPP were deduced in [7] by exploiting the restarting property.

Proposition 1. (The restarting property [7]) *If N is a GPP with parameter triple $(\kappa(t), \gamma, \beta)$, then the conditional future process $N_u(t) \equiv N(u+t) - N(u)$ given $N(u) = n$ and the history up to time u is itself a GPP with parameter triple $(\kappa(u+t), \gamma, \beta + n\gamma)$.*

Theorem 1 of [7] establishes the joint distribution of a GPP by exploiting the restarting property. As a consequence, the marginal distribution of a GPP starting at $N(0) = 0$ has a simple form.

Proposition 2. (Negative binomial marginal distribution [7]) *If N is a GPP with parameter triple $(\kappa(t), \gamma, \beta)$, then $N(t)$ has a negative binomial distribution with probability mass function $\mathbb{P}(N(t) = k) \equiv f(k; r, p(t)) = C(\beta, \gamma, k)(1 - p(t))^r p(t)^k$, $k = 0, 1, 2, \dots$, where $r = \beta/\gamma$, $p(t) = 1 - \exp[-\gamma K(t)]$, $K(t) \equiv \int_0^t \kappa(s) ds$, $t \geq 0$, and $C(\beta, \gamma, k) = \Gamma((\beta/\gamma) + k) / \Gamma((\beta/\gamma))k!$, with Γ being the gamma function, so $N(t)$ has mean and variance*

$$\mathbb{E}[N(t)] = \frac{rp(t)}{1 - p(t)}, \quad \text{Var}(N(t)) = \frac{rp(t)}{(1 - p(t))^2}, \quad t \geq 0.$$

For a general function $\kappa(t)$ the time-varying behavior can be complicated, but it simplifies for the classical Pólya point process and closely related processes (allowing $\beta \neq 1$). Indeed, for $\kappa(t)$ in (2) the GPP is a (strictly) stationary point process, i.e. the joint distribution of any k increments is independent of time shifts, as we show next. In the spirit of [16], we thus call the GPP with triple $(\kappa(t), \gamma, \beta)$ for $\kappa(t)$ in (2) a (β, γ) Ψ -GPP (again, with Ψ as a mnemonic for SI).

Theorem 1. (A stationary point process: Ψ -GPP) *Consider a GPP with parameter triple $(\kappa(t), \gamma, \beta)$. If $\kappa(t)$ is given by (2), then $1 - p(t) = \kappa(t)$,*

$$\mathbb{E}[N(t)] = \beta t, \quad \text{Var}(N(t)) = \beta t(1 + \gamma t), \quad t \geq 0. \quad (3)$$

Moreover, the joint distribution of k increments, $N(s_i + t_i + h) - N(s_i + h)$, $1 \leq i \leq k$, is independent of $h > 0$ for all h , so that N is a stationary stochastic point process with

$$\text{Cov}(N(s), N(t)) = \beta s(1 + \gamma t), \quad 0 \leq s \leq t < \infty. \quad (4)$$

Proof. For κ in (2),

$$K(t) \equiv \int_0^t (\gamma s + 1)^{-1} ds = \gamma^{-1} \log(\gamma t + 1), \quad t \geq 0,$$

so that $1 - p(t) = e^{-\gamma K(t)} = \kappa(t)$. Then the conclusion about the distribution of a single increment follows from the displayed distribution of an increment in [7, Theorem 1(ii)]. For the covariance in (4), write $\text{Var}(N(t - s)) = \text{Var}(N(t) - N(s)) = \text{Var}(N(t)) + \text{Var}(N(s)) - 2\text{Cov}(N(s), N(t))$, where the first equality follows from stationary increments, to obtain $\text{Cov}(N(s), N(t)) = [\text{Var}(N(t)) + \text{Var}(N(s)) - \text{Var}(N(t - s))]/2$, and then use the variance formula in (3).

For the joint distribution of k increments, we first observe that, without loss of generality, we can assume that the k increments are disjoint and contiguous, so that they represent a partition of a fixed interval $(s, s + t]$ into finitely many subintervals. (We are thus initially adding intervals and determining the joint distribution for more intervals.) We then apply [7, Theorem 3 and Remark 3] to conclude that the conditional distribution for the sequence of times when N increases on $(s, s + t]$ given that $N(s + t) - N(s) = k$ is, first, independent of s , and second, is itself the same as that of the order statistics of k i.i.d. random variables, each with probability density function

$$f(x) \equiv \frac{\gamma \kappa(x) \exp(\gamma K(x))}{\exp(\gamma K(t)) - 1}, \quad 0 \leq x \leq t,$$

assuming $s = 0$. If $\kappa(t)$ is given by (2), then $f(x) = 1/t$; i.e. the probability density function (PDF) of the uniform distribution. That in turn implies that the joint distribution of the k disjoint and contiguous increments $N(s_i + t_i + h) - N(s_i + h)$, $1 \leq i \leq k$, all within some larger interval $(s + h, s + t + h)$, is independent of $h > 0$ for all h . That follows because the conditioning event that $N(s + t + h) - N(s + h) = k$ has a distribution that is independent of h , and then the conditional distribution of the points within the interval given that $N(s + t + h) - N(s + h) = k$ is also independent of h . Finally, we go from the case of contiguous intervals to the original case by integrating out in order to obtain the desired marginal distribution; e.g. when we let the probability that one interval is $\leq \infty$, then that interval drops out. \square

We now show that any GPP with general $\kappa(t)$ in (1) can be expressed as a unique deterministic time-transformation of a Ψ -GPP based on (2). As a consequence, limits established for Ψ -GPPs will extend to GPPs by the continuous mapping theorem. Let $\stackrel{d}{=}$ denote equality in distribution, including for stochastic processes.

Theorem 2. (Characterization of GPP) *Let \hat{N} be the GPP with parameter triple $(\kappa(t), \gamma, \beta)$ satisfying (1); let N be the Ψ -GPP with parameter triple $(\kappa(t), \gamma, \beta)$ satisfying (1) and (2); and let*

$$M(t) \equiv \int_0^t \mu(s) ds, \quad t \geq 0, \tag{5}$$

where μ is a continuous positive function on $[0, \infty)$. Then $\{\hat{N}(t) : t \geq 0\} \stackrel{d}{=} \{N(M(t)) : t \geq 0\}$ if and only if

$$M(t) = \gamma^{-1}(e^{\gamma K(t)} - 1), \quad t \geq 0, \quad \text{where } K(t) \equiv \int_0^t \kappa(s) ds. \tag{6}$$

Proof. The stochastic intensity of N can be interpreted as

$$\lambda(t) \equiv \lambda(t \mid \mathcal{H}_{t-}) \equiv \lim_{h \rightarrow 0} \frac{\mathbb{P}(N(t+h) - N(t) = 1 \mid \mathcal{H}_{t-})}{h},$$

where $\{\mathcal{H}_t : t \geq 0\}$ is the family of histories associated with N . Let $C(t) = N(M(t))$, $t \geq 0$, be the composition process and let $\lambda_C(t)$ be its stochastic intensity function. Let $\{\mathcal{C}_t : t > 0\}$ be the histories associated with C . Because of (5), $\mathcal{C}_t = \mathcal{N}_{M(t)}$, $t \geq 0$. Then,

$$\begin{aligned} \lambda_C(t) &\equiv \lambda_C(t \mid \mathcal{C}_{t-}) \equiv \lim_{h \rightarrow 0} \frac{\mathbb{P}(C(t+h) - C(t) = 1 \mid \mathcal{C}_{t-})}{h} \\ &= \lim_{h \rightarrow 0} \frac{\mathbb{P}(N(M(t+h)) - N(M(t)) = 1 \mid \mathcal{N}_{M(t-)})}{h} \\ &= \lim_{h \rightarrow 0} \frac{\mathbb{P}(N(M(t) + \delta(h)) - N(M(t)) = 1 \mid \mathcal{N}_{M(t-)}) \delta(h)}{\delta(h) h}, \end{aligned}$$

where $\delta \equiv \delta(h) \equiv M(t+h) - M(t) = \mu(t)h + o(h)$ as $h \downarrow 0$. Hence,

$$\lambda_C(t) = \lambda(M(t))\mu(t) = \frac{(\gamma N((M(t)) + \beta)\mu(t)}{\gamma M(t) + 1},$$

where we use the specific form of $\kappa(t)$ in (2) because N is a Ψ -GPP.

Now let $\hat{\lambda}(t)$ be the stochastic intensity associated with \hat{N} . Then $\hat{\lambda}(t) = (\gamma \hat{N}(t) + \beta)\kappa(t)$, $t \geq 0$. Under the condition that $\hat{N}(t) = N(M(t))$, we have $\gamma \hat{N}(t) + \beta = \gamma N((M(t)) + \beta)$. Hence, $\lambda_C(t) = \hat{\lambda}(t)$ for all $t \geq 0$ if and only if

$$\frac{\mu(t)}{\gamma M(t) + 1} = \kappa(t), \quad t \geq 0. \tag{7}$$

Integrating both sides of (7), we get (6), as claimed. In closing, we mention two alternative proofs. The first can be based on the martingale characterization of the stochastic intensity as in [3, Section 1.8]. Then $\lambda(t) \equiv \lambda(t \mid \mathcal{H}_t)$ is the stochastic intensity of N in (1) if

$$\mathbb{E}[N(b) - N(a) \mid \mathcal{H}_a] = \mathbb{E}\left[\int_a^b \lambda(s) ds \mid \mathcal{H}_a\right]$$

for all intervals (a, b) . The second is to directly show that the finite-dimensional distributions are the same. □

Remark 1. (*Index of dispersion*) To better understand the impact of the variability as a function of time in an arrival process on the performance of a queueing model, we have shown in [17, 38, 39] that it is often helpful to look at the index of dispersion for counts, which for a (β, γ) Ψ -GPP is

$$I(t) \equiv \frac{\text{Var}(N(t))}{EN(t)} = 1 + \gamma t, \quad t \geq 0.$$

From this equation, we see that the variability increases without bound as t increases by this measure, consistent with Theorem 3. The index of dispersion for counts is also considered in [31, 32, 33] under the name ‘peakedness’, which is often used to describe traffic variability, but more commonly in a different way; see [28] and references therein.

We next apply Theorem 1 to characterize the nature of the path-dependent behavior for a (β, γ) Ψ -GPP. We do that by the following non-ergodic LLN; see, e.g., [20, Section 5.1] and references therein.

Theorem 3. (Non-ergodic LLN) *If $N(t)$ is a (β, γ) Ψ -GPP, then $t^{-1}N(t) \rightarrow L(\gamma, \beta)$ as $t \rightarrow \infty$ with probability 1, where L has a gamma distribution with shape β/γ and rate $1/\gamma$, and thus mean $\mathbb{E}[L] = \beta$ and variance $\text{Var}(L) = \beta\gamma$.*

Proof. Because the increments $N(n) - N(n - 1), n \geq 1$, form a stationary sequence, we can apply the Birkhoff ergodic theorem as in [6, Theorem 6.2.1] to establish the almost sure convergence of $N(n)/n$ as $n \rightarrow \infty$. The almost sure convergence of $N(t)/t$ as $t \rightarrow \infty$ is an easy consequence, e.g. using $N(\lfloor nt \rfloor) \leq N(t) \leq N(\lfloor nt \rfloor + 1)$ for $\lfloor nt \rfloor \leq t < \lfloor nt \rfloor + 1$. Next, from (3), we see that $\mathbb{E}[N(t)/t] = \beta$ for all $t > 0$ and $\text{Var}(N(t)/t) = \beta(1 + \gamma t)/t \rightarrow \beta\gamma$ as $t \rightarrow \infty$. The limiting gamma distribution for $N(t)/t$ is obtained by taking the limit as $t \rightarrow \infty$ of the characteristic function $\phi_{N(t)/t}(s) = \phi_{N(t)}(s/t)$. In particular, using Taylor-series asymptotics in the last step, we obtain

$$\begin{aligned} \phi_{t^{-1}N(t)}(s) &= \left(\frac{1 - p(t)}{1 - p(t)e^{is/t}} \right)^{\beta/\gamma} = \left(\frac{1}{1 + [p(t)/(1 - p(t))](1 - e^{is/t})} \right)^{\beta/\gamma} \\ &= \left(\frac{1}{1 + \gamma t(1 - e^{is/t})} \right)^{\beta/\gamma} = \left(\frac{1}{1 + \gamma t(1 - (1 + (is/t) + O(1/t^2)))} \right)^{\beta/\gamma} \\ &\rightarrow \left(\frac{1}{1 - is\gamma} \right)^{\beta/\gamma} \quad \text{as } t \rightarrow \infty. \end{aligned}$$

Finally, we recognize the limit as the characteristic function of the claimed gamma distribution. Convergence of characteristic functions then applies the convergence in distribution by [15, Theorem XV.2]. □

Theorem 3 implies that the pure birth process N has a limiting rate as $t \rightarrow \infty$, but that rate is random.

Corollary 1. (Asymptotically Poisson with a random rate.) *If $N(t)$ is a (β, γ) Ψ -GPP, with stochastic intensity $\lambda(t)$ in (1), then $\lambda(t) \rightarrow L(\gamma, \beta)$ as $t \rightarrow \infty$ with probability 1, where L is the gamma random variable from Theorem 3 with shape β/γ and rate $1/\gamma$. Hence, asymptotically as $t \rightarrow \infty$ the point process behaves as a Poisson process at random rate $L(\gamma, \beta)$.*

Proof. Multiply and divide by t in (1) and observe that the numerator converges to γL by Theorem 3, while the denominator converges to γ . □

Remark 2. (Instability of the $P/GI/1$ queue.) Theorem 3 and Corollary 1 imply that the queue length process is not stable in the $P/GI/1$ queue with a Pólya arrival process; i.e. there does not exist a random variable Q with $\mathbb{P}(Q < \infty) = 1$ such that $Q(t) \Rightarrow Q$ as $t \rightarrow \infty$, where \Rightarrow denotes convergence in distribution. That contradicts various conclusions about steady-state performance in [31, 32, 33]. We elaborate on stability and discuss ways to stabilize performance in queues with Pólya arrival processes in Section 6, but our main goal is to obtain a tractable approximation for transient performance in a class of $P/GI/1$ models. That is established by Theorem 5 and Corollary 6.

We illustrate Theorem 3 by showing the results of a simulation experiment. Figure 1 plots the empirical distribution of $N(100)/100$ based on 50,000 i.i.d. samples (left) and 25 individual

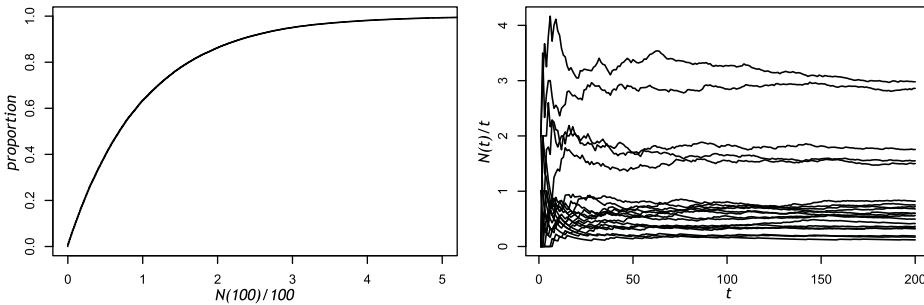


FIGURE 1: Display of the empirical distribution of $N(100)/100$ (left) and 25 individual sample paths of $N(t)/t$ over $[0, 200]$ (right) for a stationary Pólya point process N with $(\beta, \gamma) = (1, 1)$.

sample paths of $N(t)/t$ over $[0, 200]$ for a stationary Pólya point process with parameter pair $(\beta, \gamma) = (1, 1)$, assuming that (2) holds (right).

Figure 1 shows that much of the uncertainty about the long-run average disappears after an initial time t , e.g. $t = 50$, but the process $N_u(t) \equiv N(t + u) - N(u)$, $t \geq 0$, given that $N(u) = m$, is itself a new GPP by the restart property in Proposition 1. Additional results about this conditional process are given in [7]. We now apply [7] to further quantify the impact of starting at time u . The following result is elementary to prove, but the relations are revealing.

Corollary 2. (Impact of starting later) *If $N_u(t) \equiv N(u + t) - N(u)$ conditioned on $N(u) = m$, where N is the Ψ -GPP with parameter triple $(\kappa(t), \beta, \gamma)$ for $\kappa(t)$ in (2), then N_u is a new Ψ -GPP with parameter triple $(\kappa_u(t), \beta_u, \gamma_u)$ for $\kappa_u(t) = \kappa(u + t)$, $\beta_u \equiv \beta + m\gamma$ and $\gamma_u \equiv \gamma$. Consequently, $N_u(t)$ has a negative binomial distribution for each $t \geq 0$ with*

$$\mathbb{E}[N_u(t)] = \left(\frac{\beta_u}{\gamma_u}\right) \left(\frac{p_u(t)}{1 - p_u(t)}\right) = \frac{(\beta + m\gamma)t}{\gamma u + 1},$$

where

$$p_u(t) \equiv \frac{\eta_u t}{\eta_u t + 1}, \quad \eta_u \equiv \frac{\gamma}{\gamma u + 1},$$

so that $\mathbb{E}[N_u(t)] \leq \mathbb{E}[N(t)]$ if and only if $(m/u) \leq \gamma$. Moreover, for $0 \leq s \leq t$,

$$\text{Cov}(N_u(s), N_u(t)) = (\beta_u \eta_u / \gamma_u) s(1 + \eta_u t) = \mathbb{E}[N_u(s)](1 + \eta_u t),$$

$$I_u(t) \equiv \frac{\text{Var}(N_u(t))}{\mathbb{E}[N_u(t)]} = 1 + \eta_u t,$$

where η_u is strictly decreasing in u with $\eta_0 = \gamma$ and $\eta_u \rightarrow 1$ as $u \rightarrow \infty$.

Proof. This result mostly follows from the restart property from [7] stated here in Proposition 1. We also apply the explicit distribution of $N_u(t)$ given in [7, (20)]. The stationarity property follows from essentially the same proof as for Theorem 1. \square

We close this section by comparing the Ψ -GPP to another self-exciting process.

Remark 3. (Comparison to the Hawkes process.) The nondegenerate limit in Theorem 2 makes the Ψ -GPP quite different from the widely applied Hawkes [23, 24] process and most of its variants. For the basic Hawkes process in [23, (8)], instead of (1) we have

$$\lambda(t) \equiv \lambda(t | \mathcal{H}_{t-}) \equiv \nu + \int_{-\infty}^{t-} g(t - u) dN(u),$$

where N is defined over the entire real line and g is a nonnegative kernel satisfying $\eta \equiv \int_0^\infty g(u) du < 1$, so that the stationary rate is $\nu/(1 - \eta)$. Hawkes processes are alternative self-exciting processes, but they are stationary and ergodic point processes [4]. Thus, Hawkes processes are not path dependent. For applications of Hawkes processes to queues, see [9, 18].

3. Convergence to a Ψ -GMP

We now show that a properly scaled sequence of the superposition of i.i.d. Ψ -GPPs in Section 2 converges to a Ψ -GMP, the Gaussian Markov process with stationary increments studied in [16]. (In fact, [16] focuses on a multivariate Ψ -GMP.) We obtain all possible univariate Ψ -GMPs exhibiting positive dependence, as we explain in Section 8.

For $n \geq 1$, let

$$A^n = N^1 + \dots + N^n \tag{8}$$

be the sum of n i.i.d. GPPs each with parameter triple $(\kappa(t), \gamma, \beta)$. We first note that our superposition process is another GPP.

Proposition 3. (Superposition, Theorem 1 of [8].) *The superposition of two independent GPPs with parameter triples $(\kappa(t), \gamma, \beta_i), i = 1, 2$, is itself a GPP with parameter triple $(\kappa(t), \gamma, \beta_1 + \beta_2)$. If each GPP is a Ψ -GPP, then so is the superposition process. Then the superposition process satisfies the non-ergodic LLN in Theorem 3.*

We now apply the usual FCLT spatial scaling, but without scaling time by n (as in [36, (2.1) on p. 226 or (8.4) on p. 320]). In particular, for $n \geq 1$, let

$$A_n(t) \equiv n^{-1/2}(A^n(t) - \beta nt), \quad t \geq 0. \tag{9}$$

Let \Rightarrow denote convergence in distribution and let $D \equiv D[0, \infty)$ be the usual function space of right-continuous real-valued functions [5, 36].

Theorem 4. (FCLT for the superposition process.) *Consider the scaled superposition process $A_n(t)$ in (9). For $\kappa(t)$ in (2), so that in (8) N^1 is a (β, γ) Ψ -GPP while A^n is an $(n\beta, \gamma)$ Ψ -GPP,*

$$A_n \Rightarrow A \text{ in } D \quad \text{as } n \rightarrow \infty, \tag{10}$$

where A is a Ψ -GMP, i.e. a zero-mean Gaussian Markov process with stationary increments and covariance function

$$\text{Cov}(A(s), A(t)) = \mathbb{E}[A(s)A(t)] = \beta s(1 + \gamma t) = \text{Cov}(N^1(s), N^1(t)). \tag{11}$$

The limit A also satisfies the stochastic differential equation

$$dA(t) = \mu(t)A(t) dt + \sigma dB(t), \quad t \geq 0, \tag{12}$$

where $A(0) \equiv 0$, B is standard Brownian motion,

$$\mu(t) \equiv \frac{\beta\gamma}{\beta + \beta\gamma t} = (t + (1/\gamma))^{-1}, \quad \sigma = \sqrt{\beta}. \tag{13}$$

Proof. For the limit in (10), we apply Hahn’s [21] FCLT for sums of processes in [36, Theorem 7.2.1]. We verify the moment inequality conditions in that theorem in Section 7. The

stochastic differential equation characterization in (12) and (13) follows from [16, Theorem 3]. A Gaussian process with that covariance kernel is a Markov process by [12, Theorem 8.1, p. 233]. □

Remark 4. (*Parameters.*) Even though (11) shows identical structure in the covariance functions of the Ψ -GPP N^1 in (8) and the Ψ -GMP A in (10), the conventions here for the parameters are not the same as in [16]. When N^1 is a (β, γ) Ψ -GPP, A is an $(\alpha^*, \beta^*) = (\beta, -\beta\gamma)$ Ψ -GMP in [16].

Remark 5. (*Structural analogs.*) Many properties of the Ψ -GMP were established in [16]. Properties can also be deduced as a consequence of Theorem 4. Lemma 4 of [16] established an analog of the restarting property in Proposition 1. A variant of the proof of Theorem 3 shows that the Ψ -GMP also satisfies a non-ergodic LLN, with a Gaussian limit instead of a gamma distribution. Additional properties of a Ψ -GMP with drift are established in the next section.

4. Convergence to a Ψ -GMP with drift

For stable queueing models, there tends to be a negative drift in the potential net input process. Hence, in this section we consider a modification of the FCLT in Theorem 4 to produce a drift in the Ψ -GMP limit process. For that purpose, let $\{\mu_n : n \geq 1\}$ be a sequence of real numbers that satisfies

$$\mu_n \rightarrow 1 \quad \text{and} \quad \sqrt{n}(\mu_n - 1) \rightarrow \mu \quad \text{as } n \rightarrow \infty. \tag{14}$$

We are primarily interested in the case $\mu < 0$. Let $A^{d,n}(t) = A^n(\mu_n t)$ and

$$A_n^d(t) = n^{-1/2}(A^{d,n}(t) - \beta nt) = n^{-1/2}(A^n(\mu_n t) - \beta nt), \quad t \geq 0. \tag{15}$$

Let $e \equiv e(t) = t, t \geq 0$, be the identity function in D , and let D^k be the usual k -fold product space.

Corollary 3. (FCLT with a drift.) *If (14) holds in addition to the assumptions of Theorem 4, then $A_n^d \Rightarrow A + \omega e$ in D as $n \rightarrow \infty$ for A_n^d in (15) and $\omega \equiv \beta\mu$.*

Proof. We apply the continuous mapping argument for composition with centering as in [36, Section 13.3]. For that purpose, let $(\bar{M}_n(t), M_n(t)) \equiv (\mu_n t, \sqrt{n}(\mu_n - 1)t)$. Note that $A_n^d = A_n \circ \bar{M}_n + \beta M_n$, where \circ denotes the composition map. It is elementary that $(\bar{M}_n, M_n) \rightarrow (e, \mu e)$ in D^2 as $n \rightarrow \infty$. Then apply [36, Theorem 11.4.5] with Theorem 4 above to get the joint convergence $(A_n, \bar{M}_n, M_n) \Rightarrow (A, e, \mu e)$ in D^3 . Then the limit preservation in [36, Theorem 13.3.1] yields $A_n^d = (A_n \circ \bar{M}_n + \beta M_n) \Rightarrow A + \mu\beta e$ in D as $n \rightarrow \infty$. □

We now state three properties of a GMP with drift. The first two provide additional characterization of the path-dependent behavior.

Proposition 4. (Conditional mean, Lemma 4 from [16].) *If $A^d \equiv A + \omega e$ as in Corollary 3, where A is a Ψ -GMP satisfying (11) in Theorem 4, then $\mathbb{E}[A^d(s+t) - A^d(s) | A^d(u), 0 \leq u \leq s] = \omega(s)t$ for all $s, t \geq 0$, where $\omega(s) \equiv \omega + \gamma(1 + \gamma s)^{-1}(A^d(s) - s\omega)$.*

Proposition 4 shows that conditioning on the history induces the process to have a new constant drift.

Let $\text{Cor}(X, Y) \equiv \text{Cov}(X, Y) / \sqrt{\text{Var}(X)\text{Var}(Y)}$ be the correlation function.

Corollary 4. (Correlation between non-overlapping time intervals, Proposition 2 of [16].) For A^d as in Proposition 4,

$$\text{Cor}(A^d(t+s+u) - A^d(t+u), A^d(t+s) - A^d(t)) = \frac{\gamma s}{\gamma s + 1}$$

for all $t \geq 0$ and $u \geq s \geq 0$.

Corollary 4 concludes that the correlation between increments over non-overlapping intervals of equal lengths depends on the length of the intervals (s here) but not at all on the separation between the intervals (u here). The following corollary gives the limiting distribution. Let $\Phi(x) \equiv \mathbb{P}(N(0, 1) \leq x)$ be the standard normal cumulative distribution function (CDF) and let $\Phi^c(x) \equiv 1 - \Phi(x)$.

Corollary 5. (Asymptotics for the CDF.) For A^d as in Proposition 4,

$$\lim_{t \rightarrow \infty} \mathbb{P}(A^d(t) \leq x) = \Phi(-\omega/\sqrt{\beta\gamma})$$

for all x and all $t \geq 0$. Moreover,

$$\begin{aligned} \lim_{x \rightarrow -\infty} \lim_{t \rightarrow \infty} \mathbb{P}(A^d(t) \leq x) &= \Phi(-\omega/\sqrt{\beta\gamma}), \\ \lim_{x \rightarrow +\infty} \lim_{t \rightarrow \infty} \mathbb{P}(A^d(t) > x) &= \Phi^c(-\omega/\sqrt{\beta\gamma}). \end{aligned} \tag{16}$$

Proof. To get the first result we can directly take the limit in the Gaussian distribution of $A^d(t)$, which is

$$\mathbb{P}(A^d(t) \leq x) = \Phi\left(\frac{(x - \omega t)}{\sqrt{t(\beta + \beta\gamma t)}}\right),$$

from which (16) follows directly. □

Corollary 5 can be understood by recognizing that the standard deviation is the same order t as the mean, i.e. $\mathbb{E}[A^d(t)] = \omega t$ and $t^{-1}\sqrt{\text{Var}(A^d(t))} \rightarrow \sqrt{\beta\gamma}$ as $t \rightarrow \infty$. Thus, $\mathbb{P}(A^d(t) > 0)$ does not approach 0 or 1 as t increases.

5. Heavy-traffic limit for the $\sum_{i=1}^n P_i/GI/1$ queue

We now consider the single-server queue with arrival process $A^{d,n}(t) = A^n(\mu_n t)$ defined before (15). We assume that the service times are independent of the arrival process, mutually i.i.d. with a general distribution having mean $1/\beta$ and squared coefficient of variation (variance divided by the square of the mean) c_s^2 , where the service times are independent of the arrival times. We work with the associated renewal counting process $C(t)$. Since we center with $\beta n t$ in (15), we assume that the rate of this renewal process is also β . Let the scaled service process be

$$S_n(t) \equiv n^{-1/2}(C(nt) - \beta n t), \quad t \geq 0, \tag{17}$$

As in [36, Section 9.3], for the service process we only require a standard FCLT. Thus, that part of the following theorem can easily be generalized.

Let $X_n \equiv A_n^d - S_n$, $n \geq 1$, and let

$$Q_n(t) \equiv n^{-1/2}Q^n(t), \quad t \geq 0, \tag{18}$$

where $\{Q^n(t) : t \geq 0\}$ is the queue length (number in system) process in the system with initial queue length $Q^n(0)$, arrival process $\{A_n^d(t) : t \geq 0\}$ defined before (15) and scaled service renewal counting process $\{S_n(t) : t \geq 0\}$ defined in (17).

With those definitions, we apply the one-dimensional reflection map in [36, Section 13.5]. Let $\phi : D \times R \rightarrow D$ be the reflection map mapping a net input function x with $x(0) = 0$ and an initial queue length $q(0)$ with $q(0) \geq 0$ into $q(t)$ for $t > 0$ by

$$\phi(x)(t, q(0)) = q(0) + x(t) - \inf_{0 \leq s \leq t} \{\min\{q(0) + x(s), 0\}\}, \quad t \geq 0. \tag{19}$$

The reflection map in (19) is a continuous function on its domain.

Theorem 5. (Heavy-traffic FCLT for the $\sum_{i=1}^n P_i/GI/1$ queue.) *Consider a sequence of $\sum_{i=1}^n P_i/GI/1$ queues indexed by n , where the n th arrival process is the scaled superposition process $A_n^d(t)$ in (15) and (8), while the scaled service process is the scaled renewal counting process $S_n(t)$ in (17). We make the following assumptions about the initial conditions: let the arrival process after time 0 be independent of $Q_n(0)$; let the remaining service time in the process at time 0, if any, have finite mean; let all customers enter service in order of arrival from the service renewal counting process. If $n^{(-1/2)}Q^n(0) = Q_n(0)$ as $n \rightarrow \infty$ and $X_n \equiv A_n^d - S_n$, then $(A_n^d, S_n, X_n, Q_n) \Rightarrow (A + \omega e, S, X, Q)$ in D^4 as $n \rightarrow \infty$, where $\omega \equiv \mu\beta$ and A is a Ψ -GMP, while $S = \beta^{3/2}c_s B$ with B being standard Brownian motion, $X \equiv Y + \omega e$, $Y \equiv A - S$ and $Q \equiv \phi(X, Q(0))$ for ϕ in (19). In particular, Y is a Ψ -GMP with $\mathbb{E}[Y(s)Y(t)] = \beta s(1 + \gamma t) + \beta^3 c_s^2 s$ and $\text{Var}(Y(t)) = \beta t(1 + \gamma t) + \beta^3 c_s^2 t$ for $0 \leq s \leq t$, and so parameter pair $(\alpha^*, \beta^*) = (\beta + \beta^3 c_s^2, -\beta\gamma)$ in the terminology of [16], while X is a Ψ -GMP with $\text{Var}(X(t)) = \text{Var}(Y(t))$ and deterministic drift $\omega \equiv \mu\beta$.*

Proof. We apply the standard methodology for establishing a heavy-traffic FCLT for a single-server queue. We apply Donsker’s FCLT for the service times in [36, Section 4.3] and the inverse equivalence in [36, Theorem 7.3.2], in particular [36, Corollary 7.3.2 on p. 236], to obtain $S_n \Rightarrow \beta^{3/2}c_s B$ in D as $n \rightarrow \infty$, where B is a standard Brownian motion. Then we can apply Corollary 3 to obtain the limit $A_n^d \Rightarrow A + \omega e$ in D . Joint convergence for (A_n^d, S_n) then follows from independence and [36, Theorem 11.4.4]. We can then apply [36, Theorems 9.3.3, 9.3.4, 9.8.2]. □

Remark 6. (Other processes) As indicated in [36, Theorems 9.3.3 and 9.3.4], we can also obtain a heavy-traffic limit for other processes, including the workload and departure process, jointly with the limit above, by the same argument.

By the continuous mapping theorem with the projection map, we have the following corollary providing a limit for the marginal distributions. Theorem 5 of [16] provides the explicit form of the marginal distribution of the limit process, so that it can provide useful numerical results. Let the PDF of the joint limiting distribution be denoted by $f(x_s, q_s, q_{s+t}) \equiv f_{X(s), Q(s), Q(s+t)}(x_s, q_s, q_{s+t})$, and similarly for the associated marginal PDFs and conditional PDFs. We express the limiting distribution in terms of the exponential function and the standard normal CDF $\Phi(x) \equiv \mathbb{P}(N(0, 1) \leq x)$ and PDF $\phi(x)$. Let the associated CDF of $N(m, \sigma^2) \stackrel{d}{=} m + \sigma N(0, 1)$ be denoted by $\Phi(x; m, \sigma^2)$, and similarly for the others. To connect with [16], let

$$\omega_s \equiv \frac{\alpha^* \omega - \beta^* x_s}{\alpha^* - \beta^* s}, \quad \beta_s^* \equiv \frac{\alpha^* \beta^*}{\alpha^* - \beta^* s}, \tag{20}$$

where

$$\alpha^* \equiv \beta, \quad \beta^* \equiv -\beta\gamma, \tag{21}$$

as in Remark 4. Let $\delta(\cdot)$ be the Dirac delta function and let $\mathbf{1}_A$ be the indicator function, equal to 1 on A and 0 elsewhere.

Corollary 6. (Marginal limiting distributions.) *Under the conditions of Theorem 5, $(X_n(s), Q_n(s), Q_n(s+t)) \Rightarrow (X(s), Q(s), Q(s+t))$ in \mathbb{R}^3 as $n \rightarrow \infty$, where $X(s)$ is a (mean- ωs , variance $s(\alpha^* - \beta_s^*)$) Gaussian random variable for β_s^* in (20), while the joint limiting distribution has joint PDF*

$$f(x_s, q_s, q_{s+t}) = f(x_s)f(q_s | x_s)f(q_{s+t} | x_s, q_s), \tag{22}$$

where, assuming $\mathbb{P}(Q(0) = q_0) = 1$,

$$\begin{aligned} f(x_s) &\equiv \phi(x_s; \omega s, s(\alpha^* - \beta_s^*)) = \frac{1}{\sqrt{s(\alpha^* - \beta_s^*)}} \phi\left(\frac{x_s - \omega s}{\sqrt{s(\alpha^* - \beta_s^*)}}\right), \\ f(q_s | x_s) &\equiv \left(1 - e^{\{-2q_s(q_s - x_s)/(\alpha^* s)\}}\right) (\delta(q_s - q_0 - x_s)\mathbf{1}_{\{q_s \geq 0\}} + \mathbf{1}_{\{q_s - q_0 - x_s \geq 0\}}\delta(q_s)) \\ &\quad + \left(\frac{(4q_s - 2x_s)e^{\{-2q_s(q_s - x_s)/(\alpha^* s)\}}}{\alpha^* s}\right) \mathbf{1}_{\{q_s - q_0 - x_s \geq 0\}} \mathbf{1}_{\{q_s \geq 0\}}, \\ f(q_{s+t} | x_s, q_s) &\equiv \frac{1}{\sqrt{t(\alpha^* - \beta_s^* t)}} \phi\left(\frac{q_{s+t} - q_s - \omega s t}{\sqrt{t(\alpha^* - \beta_s^* t)}}\right) \\ &\quad + e^{\{-2q_{s+t}(\beta_s^* q_{s+t} - \alpha^* \omega s)/\alpha^{*2}\}} (A_1 + A_2) \end{aligned}$$

for

$$\begin{aligned} A_1 &\equiv \left(\frac{4\beta_s^* q_{s+t} - 2\alpha^* \omega s}{\alpha^{*2}}\right) \Phi\left(\frac{(2\beta_s^* q_{s+t} - \alpha^* \omega s)t - \alpha^*(q_{s+t} + q_s)}{\alpha^* \sqrt{t(\alpha^* - \beta_s^* t)}}\right), \\ A_2 &\equiv \left(\frac{\alpha^* - 2\beta_s^* t}{\alpha^* \sqrt{t(\alpha^* - \beta_s^* t)}}\right) \phi\left(\frac{(2\beta_s^* q_{s+t} - \alpha^* \omega s)t - \alpha^*(q_{s+t} + q_s)}{\alpha^* \sqrt{t(\alpha^* - \beta_s^* t)}}\right), \end{aligned}$$

$q_s, q_{s+t} \geq 0$. As a consequence, the associated conditional CDF, for $q_{s+t} \geq 0$, is

$$\begin{aligned} \mathbb{P}(Q(s+t) \leq q_{s+t} | X(s) = x_s, Q(s) = q_s) &= \Phi\left(\frac{q_{s+t} - q_s - \omega s t}{\sqrt{t(\alpha^* - \beta_s^* t)}}\right) \\ &\quad - e^{\left(\frac{-2q_{s+t}(\beta_s^* q_{s+t} - \alpha^* \omega s)}{\alpha^{*2}}\right)} \Phi\left(\frac{(2\beta_s^* q_{s+t} - \alpha^* \omega s)t - \alpha^*(q_{s+t} + q_s)}{\alpha^* \sqrt{t(\alpha^* - \beta_s^* t)}}\right) \end{aligned} \tag{23}$$

and, for $q_t \geq 0$,

$$\begin{aligned} \mathbb{P}(Q(t) \leq q_t) &= \Phi\left(\frac{q_t - q_0 - \omega t}{\sqrt{t(\alpha^* - \beta_s^* t)}}\right) \\ &\quad - e^{\left(\frac{-2q_t(\beta_s^* q_t - \alpha^* \omega)}{\alpha^{*2}}\right)} \Phi\left(\frac{(2\beta_s^* q_t - \alpha^* \omega)t - \alpha^*(q_t + q_0)}{\alpha^* \sqrt{t(\alpha^* - \beta_s^* t)}}\right). \end{aligned}$$

Proof. We give a brief overview of the proof in [16] (appearing on the last two pages), which draws heavily on [22]. We focus on $X(t) \equiv Y(t) + \omega t$, where Y is an (α^*, β^*) Ψ -GMP for $(\alpha^*, \beta^*) = (\beta + \beta^3 c_s^2, -\beta\gamma)$ and ωt is the drift, as in Theorem 5. We exploit known results for the Brownian bridge by looking at increments from the past conditioned on later process values. For that purpose, let $X^{(s)}(t) \equiv X(s+t) - X(s)$ for some (s, t) with $0 \leq s \leq t$. We observe that conditioning $X^{(s)}$ on both $X(s) = x_s$ and $Q(s) = q_s$ results in a new (α^*, β_s^*) Ψ -GMP for β_s^* in (20) that depends on x_s but not q_s . Further conditioning it on $X^{(s)}(t) = x_t^{(s)}$ for some $t \geq s$ results in yet another $(\alpha^*, t^{-1}\alpha^*)$ Ψ -GMP with drift $t^{-1}x_t^{(s)}$ on $[0, t]$ (which no longer depends on β_s^*). That process depends on $x_t^{(s)}$ but on neither q_s nor x_s . Therefore, the process obtained by conditioning $\{Q(u) : s \leq u \leq s+t\}$ on $X(s) = x_s, Q(s) = q_s$ and $X^{(s)}(t) = x_t^{(s)}$ begins at state q_s at time s and evolves according to a net input process that is an $(\alpha^*, t^{-1}\alpha^*)$ Ψ -GMP on the interval $[0, t]$. That process is a scaled Brownian bridge, for which the distribution of the queue length was previously obtained by [22]. The result in (23) is obtained from that conditional queue length distribution using the law of total probability. We remark that $f(q_s | x_s)$ in (22) is the density of the CDF

$$\mathbb{P}(Q(s) \leq q_s | X(s) = x_s) = \left(1 - e^{\{-2q_s(q_s - x_s)/(\alpha^* s)\}}\right) \mathbf{1}_{\{q_s - q_0 - x_s \geq 0\}} \mathbf{1}_{\{q_s \geq 0\}}$$

as derived by [22]. It suffices to apply the product rule for differentiation in this equation. \square

Some further insight can be gained from further Gaussian LLN limits for the limit process. For that purpose, let $X_s(t) \equiv (X(s+t) | X(s) = x_s)$ and $Q_s(t) \equiv (Q(s+t) | X(s) = x_s, Q(s) = q_s)$. Let $N(m, \sigma^2)$ denote a normal random variable with mean m and variance σ^2 , and let $(x)^+ \equiv \max\{x, 0\}$, so that $\mathbb{P}(N(m, \sigma^2)^+ = 0) = \mathbb{P}(N(m, \sigma^2) \leq 0)$.

Corollary 7. (LLN limit for the limit process.) *Given the Ψ -GMP limit in Theorem 4 and the associated heavy-traffic limit in Theorem 5 and Corollary 6, we have the following limit:*

$$t^{-1}(A^d(t), S(t), X(t), Q(t), X_s(t), Q_s(t)) \Rightarrow (N_1(\omega, \beta\gamma), 0, N_1(\omega, \beta\gamma), N_1(\omega, \beta\gamma)^+, N_2(\omega_s, -\beta_s^*), N_2(\omega_s, -\beta_s^*)^+) \text{ in } \mathbb{R}^6 \quad (24)$$

as $t \rightarrow \infty$ for some constant $s > 0$ and (ω_s, β_s^*) in (21), where (N_1, N_2) is a random vector in \mathbb{R}^2 with Gaussian one-dimensional marginal distributions.

Proof. For the first three processes, we exploit the distribution as a function of t as in the proof of Corollary 5. For the next two processes, note that $\mathbb{P}(X(t) \leq xt) \rightarrow \Phi((x - \omega)/\sqrt{\beta\gamma})$ as $t \rightarrow \infty$ as just described, and

$$\begin{aligned} \mathbb{P}(Q(t) \leq tq | X(t) = tx) &= (1 - e^{-\{2t^2q(q-x)/\alpha^*s\}}) \mathbf{1}_{\{tq-tx \geq q_0\}} \mathbf{1}_{\{q \geq 0\}} \\ &= (1 - e^{-\{2t^2q(q-x)/\alpha^*s\}}) \mathbf{1}_{\{t \geq q_0/(q-x)\}} \mathbf{1}_{\{q-x > 0\}} \mathbf{1}_{\{q \geq 0\}} \\ &\rightarrow \mathbf{1}_{\{q-x > 0\}} \mathbf{1}_{\{q \geq 0\}} \quad \text{as } t \rightarrow \infty. \end{aligned}$$

The joint PDF for the two limits is therefore

$$\tilde{f}(x, q) \equiv \tilde{f}(x)\tilde{f}(q | x) = \frac{1}{\sqrt{\beta\gamma}} \phi((x - \omega)/\sqrt{\beta\gamma})(\delta(q - x)\mathbf{1}_{\{q \geq 0\}} + \delta(q)\mathbf{1}_{\{q > x\}}).$$

The joint PDF for the limits of the last two processes is similarly obtained. \square

Figure 2 illustrates the heavy-traffic limits established in this section by plotting results of simulations of the scaled queue-length process in the $\sum_{i=1}^n P_i/D/1$ queue.

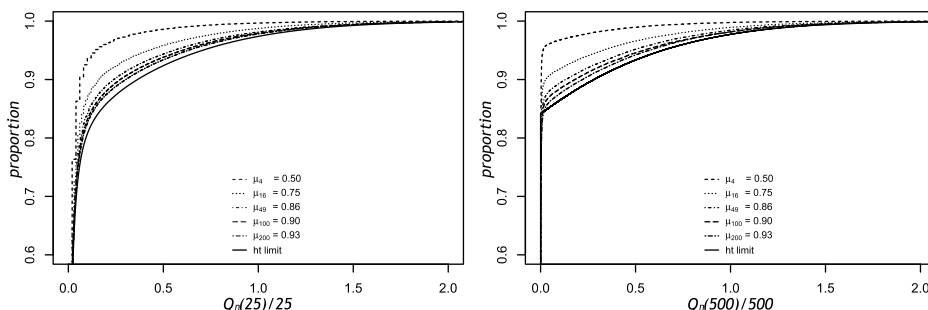


FIGURE 2: Display of the empirical CDF of $Q_n(t)/t$ for $0 \leq t \leq T$ in the $\sum P_i/D/1$ queue with $Q_n(t)$ in (18), $\beta = \gamma = \mu = 1$ for $T = 25$ (left) and 500 (right), compared to $Q(t)/t$ in (24).

We consider the (β, γ) Ψ -GPP with parameters $\beta = \gamma = 1$ as a function of the number n of processes superposed. We introduce a negative drift in the limit process by using (14) with $\mu = 1$. We let the service times initially have mean (and fixed value) 1, but that becomes $1/n$ with the scaling in (17). We look at n ranging from 4 to 200. We consider two values of T : 25 and 500. The similarity of the plots supports Corollary 7. The convergence as n increases supports Corollary 6. Numerical study indicates errors roughly of order $1/\sqrt{n}$.

Remark 7. (Insights from Corollary 7) Contrary to the usual queueing behavior, the limits are continuous functions of the drift constant ω , showing that the usual critical level for stability does not apply to the limit process. We elaborate on this property in Section 6.1. The limit for the last two components in Corollary 7 shows that the limit as $t \rightarrow \infty$ is not independent of fixed s . Corollary 7 also quantifies the growing variability in all processes except for the service-time process as t increases. The conditional random variable $Q_s(t)$ is highly variable as t increases for any given conditioning event $(X(s), Q(s)) = (x_s, q_s)$.

6. Steady-state results for queues with Ψ -GPP arrivals

This section is something of a departure from Sections 3–5, because we no longer consider FCLTs or heavy-traffic limits. In this section we present further results for queues with Ψ -GPP arrivals. In Section 6.1 we obtain some elementary stability results and show that we can obtain both positive and negative results from a sample-path version of Little’s law. In Section 6.2 we show that stability can be achieved when there are infinitely many servers. In Section 6.3 we show that it is also possible to devise adaptive service policies that stabilize performance in a single-server queue with Ψ -GPP arrivals. We use new (but similar) notation in this section.

6.1. Stability and Little’s law

We first discuss the implications of the LLN in Theorem 3 for the $P/GI/1$ queue with a Pólya arrival process. For this purpose, let the service times have mean 1. Given Theorem 3, the following is a standard heavy-traffic law of large numbers for overloaded queues, as in [36, Theorem 5.3.2].

Corollary 8. (Explosion.) *Let $Q(t)$ be the queue length process, starting with $Q(0) = 0$, in the $P/GI/1$ queue with a (β, γ) Ψ -GPP arrival process, where $0 < \beta < \infty$ and $0 < \gamma < \infty$, and i.i.d. service times with mean 1. Then $t^{-1}Q(t) \rightarrow \max\{L(\beta, \gamma) - 1, 0\}$ as $t \rightarrow \infty$ with probability 1, so that $\mathbb{P}(Q(t) \rightarrow \infty \text{ as } t \rightarrow \infty) = \mathbb{P}(L(\gamma, \beta) > 1)$, where $0 < \mathbb{P}(L(\gamma, \beta) > 1) < 1$.*

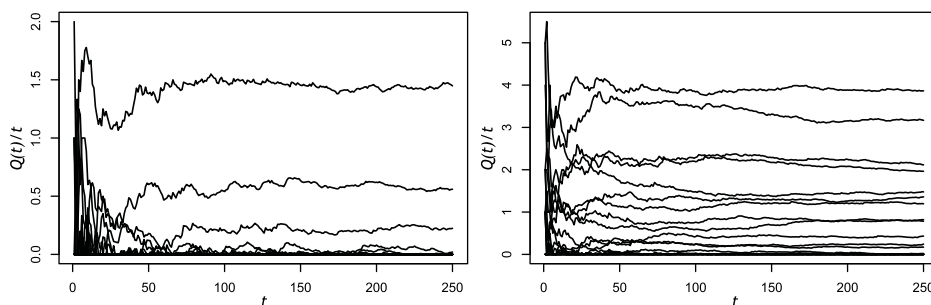


FIGURE 3: Display of 25 individual sample paths of $Q(t)/t$ for $0 \leq t \leq 250$, starting empty, for a $P/D/1$ queue with parameter pairs $(\beta, \gamma) = (0.5, 1)$ (left) and $(\beta, \gamma) = (1.5, 1)$ (right).

We illustrate the less critical role of the usual traffic intensity by showing the results of a simulation experiment. Figure 3 plots the 25 individual sample paths of $Q(t)/t$ over $[0, 250]$, starting empty, for a $P/D/1$ queue with parameter pairs $(\beta, \gamma) = (0.5, 1)$ (left) and $(\beta, \gamma) = (1.5, 1)$ (right). These plots only clearly display the paths that tend to be unstable by time $t = 250$. There are 3 of these in the 25 plots for $\beta = 0.5$ on the left, but about 12 in the 25 plots for $\beta = 1.5$ on the right. However, Corollary 2 implies that we cannot draw a firm conclusion about ultimate stability by what we see over the interval $[0, 250]$.

Nevertheless, there is a version of Little’s law in this setting that can accommodate quite general initial conditions, but for simplicity assume that the system starts empty. Let W_k be the waiting time of customer k , and let

$$L^* \equiv \lim_{t \rightarrow \infty} t^{-1} \int_0^t Q(s) \, ds, \quad W^* \equiv \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n W_k, \quad \lambda^* \equiv L(\beta, \gamma). \quad (25)$$

Corollary 9. (Little’s law.) Consider the $P/GI/1$ queue with a (β, γ) Ψ -GPP arrival process starting with $Q(0) = 0$ as above, where $0 < \beta < \infty$ and $0 < \gamma < \infty$.

(a) If W^* in (25) is well defined and finite and if $L(\beta, \gamma) < 1$, which occurs with positive probability, then L^* in (25) is well defined and $L^* = \lambda^* W^* < \infty$.

(b) If $L(\beta, \gamma) > 1$, as occurs with positive probability, then L^* in (25) is well defined but $L^* = \infty$. Moreover, W^* is then also well defined and $W^* = L^* = \infty$.

Proof. For part (a), we apply the sample-path version of Little’s law from [35]. For part (b), we apply Corollary 8 together with [40, (6)]. Sufficient conditions for more results are also given in [40]. □

6.2. The Ψ -GPP/GI/ ∞ queue

We next consider the Ψ -GPP/GI/ ∞ infinite-server queue. We show that bounded service times ensure reaching steady state in finite time. Let $\stackrel{d}{=}$ denote equality in distribution.

Theorem 6. (Infinite-server queue with bounded service times.) For an infinite-server queue with a (β, γ) Ψ -GPP arrival process, if the service times are (i) independent of the arrival process, (ii) mutually i.i.d., each distributed as a random variable V , and (iii) $\mathbb{P}(V \leq \zeta) = 1$ for some $\zeta < \infty$, then the number of busy servers $Q(t)$ reaches steady state by time ζ , i.e.

$$Q(u) \stackrel{d}{=} Q(t), \quad \mathbb{E}[Q(u)] = \beta \mathbb{E}[V] \quad \text{for all } u \geq t \geq \zeta. \quad (26)$$

Proof. We apply a useful device from [19, Section 2]. We note that if the service times are deterministic with $\mathbb{P}(V = d) = 1$, then

$$Q(t) = N(t) - N(t - d) \quad \text{for all } t \geq d. \tag{27}$$

Because the Ψ -GPP arrival process has stationary increments, the distribution of $Q(t)$ in (27) is independent of t for $t \geq d$, and so has reached steady state by time d .

We next observe that $Q(t)$ also reaches steady state in finite time when the service time distribution has finite support within the interval $[0, \zeta]$. Hence, suppose that $P(V = d_i) = q_i$, $1 \leq i \leq n$. Now we can classify the arrivals by ‘type’ according to their service time, where this type assignment is done independently of the arrival process. Thus, the distribution of $Q(t)$ is the sum of the number of each type in the system at time t , so that we can write

$$Q(t) = \sum_{i=1}^n Q_i(t) = \sum_{i=1}^n [N(t) - N(t - d_i)]T_i(t, d_i),$$

where $T_i(t, d_i)$ is the proportion of the $N(t) - N(t - d_i)$ arrivals in $[t - d_i, t]$ that are of type i , i.e. have service times d_i , which has a multinomial distribution.

We then observe that the distribution of $Q(t)$ is independent of t , because the random vector $(N(t) - N(t - d_i) : 1 \leq i \leq n)$ is independent of t in \mathbb{R}^n for all $t \geq \max \{d_i : 1 \leq i \leq n\}$.

The joint distribution of $(T_i(t, d_i) : 1 \leq i \leq n)$ is somewhat complicated, but we can directly deduce that it has property (26) and we can write down the mean:

$$\begin{aligned} \mathbb{E}[Q(t)] &= \sum_{i=1}^n \mathbb{E}[Q_i(t)] = \sum_{i=1}^n \mathbb{E}[N(t) - N(t - d_i)]\mathbb{E}[T_i(t, d_i)] \\ &= \sum_{i=1}^n \beta d_i \mathbb{E}[T_i(t, d_i)] = \beta \mathbb{E}[V]. \end{aligned}$$

Since the probability distributions with finite support in $[0, \zeta]$ are dense in the space of all probability distributions with support in $[0, \zeta]$, we obtain the general results by taking a limit. To obtain the limit, note that we can express $Q(t)$ in terms of the arrival times T_i associated with the arrival process $N(t)$ and service times V_i by $Q(t) = \sum_{i=1}^{N(t)} \mathbf{1}_{\{T_i + V_i > t\}}$, which representation implies that $Q(t)$ is almost surely a continuous function of the service times with respect to the limit process, so that we can apply the generalized continuous mapping theorem from [36, Theorem 3.4.4]. In particular, we see that $Q(t)$ is almost surely a continuous function of the service times except when $T_i + V_i = t$ for some i . For any CDF of V , that almost surely does not occur because

$$\sum_{i=1}^{\infty} \mathbb{P}(t \leq T_i \leq t + \varepsilon) = \mathbb{P}(N(t + \varepsilon) - N(t) \geq 1) \leq \mathbb{E}[N(t + \varepsilon) - N(t)] = \beta \varepsilon$$

by (3). We see that the probability is 0 by letting $\varepsilon \downarrow 0$. Hence we can apply the generalized continuous mapping theorem to deduce that the distribution of $Q(t)$ is almost surely a continuous function of the distribution of V in this setting. \square

6.3. Adaptive service processes to enforce stability

Suppose that we let the service rate at time t depend on the history of the arrival process up to that time. In particular, let the arrival-history-dependent service rate be

$$\mu(t) \equiv \mu(t | \mathcal{H}_t) = \frac{\lambda(t)}{\rho} = \frac{\gamma N(t) + \beta}{\rho(\gamma(t) + 1)} \quad \text{for all } t \geq 0. \tag{28}$$

The proposed control is a variant of the rate-matching service-rate control in [30, 37]. The following is an analog of [37, Theorem 3.1].

Theorem 7. *If the arrival-history-dependent service process in (28) with $\rho < 1$ is used in the single-server queue with $Q(0) = 0$ and the GPP arrival process having parameter triple $(\kappa(t), \gamma, \beta)$ with κ in (2), then the pair of rates satisfies the joint limit*

$$(\lambda(t), \mu(t)) \rightarrow (L(\gamma, \beta), L(\gamma, \beta)/\rho) \quad \text{as } t \rightarrow \infty \text{ with probability 1,} \tag{29}$$

where $L(\gamma, \beta)$ is the gamma random variable in Theorem 3. Moreover, the queue length process $\{Q(t) : t \geq 0\}$ is distributed the same as in an $M/M/1$ queue in a random environment, so that

$$Q(t) \Rightarrow Q \text{ as } t \rightarrow \infty, \quad \text{where } \mathbb{P}(Q = k) = (1 - \rho)\rho^k, \quad k \geq 0, \tag{30}$$

as in the $M/M/1$ queue with traffic intensity (arrival rate divided by the maximum potential service rate) ρ .

Proof. First, the joint limit in (29) follows easily from Corollary 1 and (28). For the connection to $M/M/1$, we start by representing the counting processes $N(t)$ and $S(t)$ in terms of Poisson processes, using the martingale representation for point processes as in [3, Section 1.8]. Let $\Pi_1(t)$ and $\Pi_2(t)$ be two independent rate-1 Poisson counting processes. Then the counting process N can be represented as $\Pi_1(\Lambda(t))$, $t \geq 0$, where $\Lambda(t)$ is the cumulative rate function associated with $\lambda(t)$ and the compensator of the point process, i.e. $\Lambda(t) \equiv \int_0^t \lambda(s) ds$, $t \geq 0$. Because of (28), we thus have $\{(N(t), S(t)) : t \geq 0\} \stackrel{d}{=} \{(\Pi_1(\Lambda(t)), \Pi_2(\rho^{-1}\Lambda(t))) : t \geq 0\}$; see, e.g., [13, Theorem 6.4.1]. Then the queue length process $Q(t)$ can be expressed as the reflection map applied to the net-input process $Y(t) \equiv N(t) - S(t)$. Let $\tilde{Q}(t)$ and $\tilde{Y}(t)$ be the corresponding processes in an $M/M/1$ queue with arrival rate 1 and service rate $1/\rho$, i.e. $\tilde{Y}(t) \equiv \Pi_1(t) - \Pi_2(\rho^{-1}t)$. With this representation we have $\{Y(t), Q(t) : t \geq 0\} \stackrel{d}{=} \{\tilde{Y}(\Lambda(t)), \tilde{Q}(\Lambda(t)) : t \geq 0\}$, where $\{\Lambda(t) : t \geq 0\}$ is independent of $\{\tilde{Y}(t), \tilde{Q}(t) : t \geq 0\}$. Because $\Lambda(t) \rightarrow \infty$ as $t \rightarrow \infty$ with probability 1, we have the claimed representation and convergence in (30). □

7. Completing the proof of Theorem 4

We complete the proof of Theorem 4 by verifying the two inequality conditions in Hahn’s theorem [21] as in [36, Theorem 7.2.1]. To do so, we prove convergence in $D[0, 1]$, as in [21], but we note that essentially the same arguments show convergence in $D[0, T]$ for any $T > 0$ and therefore implies convergence in $D[0, \infty)$; see [36, Section 12.9].

By Theorem 1, $\mathbb{E}[A_1(s)A_1(t)] = \beta s(1 + \gamma t)$ when A_1 is the centered process defined in (9). First, for $0 \leq t_1 \leq t_2 \leq 1$,

$$\begin{aligned} \mathbb{E} \left[(A_1(t_2) - A_1(t_1))^2 \right] &= \beta (t_2 - t_1) (1 + \gamma (t_2 - t_1)) \\ &= \beta (t_2 + \gamma t_2^2) - \beta (t_1 + \gamma t_1^2) - 2\beta\gamma t_1 (t_2 - t_1) \\ &\leq \beta (t_2 + \gamma t_2^2) - \beta (t_1 + \gamma t_1^2), \end{aligned}$$

so that condition (2.3) of [36, Theorem 7.2.1] is met. We will also show that

$$\mathbb{E}[(A_1(t) - A_1(t_1))^2 (A_1(t_2) - A_1(t))^2] \leq c (t_2 - t_1)^2 \tag{31}$$

for $0 \leq t_1 \leq t \leq t_2 \leq 1$ and a constant c , so that condition (2.4) of [36, Theorem 7.2.1] will be met as well.

To do the explicit calculations, we again apply [7, Theorem 3 and Remark 3], as in the proof of Theorem 1. That applies for any GPP N^1 , the conditional distribution for the sequence of times when N^1 increases on $(0, 1)$ given that $N^1(1) = k$ is the same as that of the order statistics of k i.i.d. random variables, each with probability density function

$$f(x) \equiv \frac{\gamma \kappa(x) \exp(\gamma K(x))}{\exp(\gamma K(1)) - 1}, \quad 0 \leq x \leq 1.$$

For a Ψ -GPP, $f(x) = 1$. Hence, the conditional distribution for the ordered sequence of times when the Ψ -GPP N^1 increases on $(0, 1)$ given that $N^1(1) = k$ is the same as that of the order statistics of i.i.d. uniform random variables U_j on $[0, 1]$.

Following the proof of [5, Theorem 14.3], let $p_1 = t - t_1$, $p_2 = t - t_2$; let V_j be $(1 - p_1)$ or $-p_1$ as U_j lies in $[t_1, t)$ or not; let W_j be $(1 - p_2)$ or $-p_2$ as U_j lies in $[t, t_2]$ or not. Also, let $V = \sum_{j=1}^k V_j$, $W = \sum_{j=1}^k W_j$, $\tilde{V} = (N^1(1) - \beta) p_1$ and $\tilde{W} = (N^1(1) - \beta) p_2$. Then

$$\begin{aligned} & \mathbb{E} \left[(A_1(t) - A_1(t_1))^2 (A_1(t_2) - A_1(t))^2 \mid N^1(1) = k \right] \\ &= \mathbb{E} \left[\left(\sum_{j=1}^k \left(V_j + \frac{(k - \beta) p_1}{k} \right) \right)^2 \left(\sum_{j=1}^k \left(W_j + \frac{(k - \beta) p_2}{k} \right) \right)^2 \mid N^1(1) = k \right] \\ &= \mathbb{E} \left[(V + \tilde{V})^2 (W + \tilde{W})^2 \mid N^1(1) = k \right] \\ &= \mathbb{E} \left[((VW + V\tilde{W}) + W\tilde{V} + \tilde{V}\tilde{W})^2 \mid N^1(1) = k \right] \\ &\leq 4 \mathbb{E} \left[(VW)^2 + (V\tilde{W})^2 + (W\tilde{V})^2 + (\tilde{V}\tilde{W})^2 \mid N^1(1) = k \right], \tag{32} \end{aligned}$$

where the last inequality follows from the Cauchy–Schwartz inequality.

It follows from [7, Theorem 1(i)] that $N \equiv N^1(1)$ has a negative binomial distribution with moment-generating function

$$M(s) = \left(\frac{\theta}{1 - (1 - \theta) e^s} \right)^r,$$

where $r = \beta/\gamma$ and $\theta = (1 + \gamma)^{-1}$. Then,

$$\mathbb{E}(N^p) = \left(\frac{d^p}{ds^p} M(s) \right)_{s=0} \quad \text{for } p \geq 1,$$

which implies that

$$\begin{aligned} \mathbb{E}[N] &= \beta, \\ \mathbb{E}[N^2] &= \beta\gamma + \beta^2 + \beta, \\ \mathbb{E}[N^3] &= \beta(2\gamma^2 + 3\beta\gamma + 3\gamma + \beta^2 + 3\beta + 1), \\ \mathbb{E}[N^4] &= \beta(6\gamma^3 + 11\beta\gamma^2 + 12\gamma^2 + 6\beta^2\gamma + 18\beta\gamma + 7\gamma + \beta^3 + 6\beta^2 + 7\beta + 1). \end{aligned} \tag{33}$$

By [5, (14.10)], $\mathbb{E}[(VW)^2 | N^1(1) = k] \leq 6k^2p_1p_2$, so that

$$\mathbb{E}[(VW)^2] \leq 6E[N^2]p_1p_2. \tag{34}$$

Similarly,

$$\begin{aligned} \mathbb{E}[(V\tilde{W})^2] &= \mathbb{E}[Np_1(1-p_1)((N-\beta)p_2)^2], \\ \mathbb{E}[(\tilde{V}W)^2] &= \mathbb{E}[Np_2(1-p_2)((N-\beta)p_1)^2], \end{aligned}$$

and

$$\mathbb{E}[(\tilde{V}\tilde{W})^2] = \mathbb{E}[(N-\beta)p_1)^2((N-\beta)p_2)^2]. \tag{35}$$

Using Macsyma for algebraic simplification, we find first, by (32),

$$\begin{aligned} &\mathbb{E}[(A_1(t) - A_1(t_1))^2(A_1(t_2) - A_1(t))^2] \\ &\leq \Omega \equiv 4 \left(\mathbb{E}[(VW)^2] + \mathbb{E}[(V\tilde{W})^2] + \mathbb{E}[(\tilde{V}W)^2] + \mathbb{E}[(\tilde{V}\tilde{W})^2] \right). \end{aligned}$$

Then, by (33)–(35), we find that

$$\begin{aligned} \Omega &= 4\beta \left(6\gamma^3p_1p_2 + 3\beta\gamma^2p_1p_2 + 8\gamma^2p_1p_2 + 4\beta\gamma p_1p_2 + \gamma p_1p_2 + \beta p_1p_2 - p_1p_2 + 2\gamma^2p_2 \right. \\ &\quad \left. + \beta\gamma p_2 + 3\gamma p_2 + \beta p_2 + p_2 + 2\gamma^2p_1 + \beta\gamma p_1 + 3\gamma p_1 + \beta p_1 + p_1 + 6\gamma + 6\beta + 6 \right) p_1p_2 \\ &\leq 4\beta(6\gamma^3 + 3\beta\gamma^2 + 12\gamma^2 + 6\beta\gamma + 13\gamma + 9\beta + 7)p_1p_2 \\ &= 4\beta(6\gamma^3 + 3\beta\gamma^2 + 12\gamma^2 + 6\beta\gamma + 13\gamma + 9\beta + 7)(t - t_1)(t_2 - t) \\ &\leq 4\beta(6\gamma^3 + 3\beta\gamma^2 + 12\gamma^2 + 6\beta\gamma + 13\gamma + 9\beta + 7)(t_2 - t_1)^2. \end{aligned}$$

Thus, we conclude that condition (2.4) of [36, Theorem 7.2.1] holds for $c = 4\beta(6\gamma^3 + 3\beta\gamma^2 + 12\gamma^2 + 6\beta\gamma + 13\gamma + 9\beta + 7)$. Under those conditions, [36, Theorem 7.2.1] shows that $A_n \Rightarrow A$, where A is a zero-mean Gaussian process with the same covariance kernel as A_1 .

Remark 8. (*Easier proof of Theorem 3 fails.*) A candidate easier proof of Theorem 4 for sums of Markov processes is provided by [36, Theorem 7.2.2], but conditions (2.9) and (2.10) there are not satisfied in our case. The proof of Theorem 4 shows that the expectation in (32) conditional on $N = N^1(1)$ does not have a uniform bound of the form of the right-hand side of (31). The uniform bound in (31) is obtained from the conditional expectation only after accounting for the distribution of N . Therefore, the candidate easier proof for sums of Markov processes fails in our case.

8. Conclusions

In this paper we have helped expose the performance consequence on a single-server queue of having a path-dependent arrival process. Our first main result is Theorem 4, which shows that a superposition of Ψ -GPPs, which is itself a Ψ -GPP, converges to a Ψ -GMP as studied in [16]; i.e. the limit process is a Gaussian Markov process with stationary increments. Our second main result is the consequence for queueing models in Theorem 5. Corollaries 6 and 7 provide explicit performance approximations for the queueing processes. Such explicit representations are unusual in the theory of Gaussian queues.

In the process of establishing the results above, we also clarified the role of Ψ -GPPs within the larger class of GPPs in Theorems 1 and 2. The Ψ -GPP and the limiting Ψ -GMP exhibit positive dependence, but the class of Ψ -GMPs considered in [16] also include processes exhibiting negative dependence. We close by observing that it is possible to obtain all possible Ψ -GMPs considered in [16] by limits like that in Theorem 4. Indeed, all possible limits are obtained by considering linear combinations of uniform empirical processes and superpositions of Ψ -GPPs. That is natural because both processes can be regarded as superposition processes. The case of negative dependence connects with previous heavy-traffic limits for queues in [22, 25]. In fact, the proof of the explicit form of the distribution of the reflected Ψ -GMP in Corollary 6 already drew on the structure of the Brownian bridge, as can be seen from our sketch of that proof.

Additional insight into the steady-state performance of queues with path-dependent arrival processes was provided by the results in Section 6. We showed how Little's law can be stated in this context, and we provided conditions under which there is stability for a queue with a Pólya arrival process.

References

- [1] ARTHUR, W. B. (1988). Self-reinforcing mechanisms in economics. In *The Economy as an Evolving Complex System*, eds P. W. Anderson, K. Arrow and D. Pines. CRC Press, Boca Raton, FL.
- [2] ARTHUR, W. B., ERMOLIEV, YU. M. AND KANJOVSKI, YU. M. (1987). Path-dependent processes and the emergence of macro-structure. *Eur. J. Operat. Res.* **30**, 294–303.
- [3] BACCELLI, F. AND BREMAUD, P. (1994). *Elements of Queueing Theory*. Springer, New York.
- [4] BACRAY, E., DELATTRE, S., HOFFMAN, M. AND MUZY, J. F. (2013). Some limit theorems for Hawkes processes and applications to financial statistics. *Stoch. Process. Appl.* **123**, 2475–2499.
- [5] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd edn. John Wiley, New York.
- [6] BREIMAN, L. (1968). *Probability*. Addison Wesley, Reading, MA.
- [7] CHA, J. H. (2014). Characterization of the generalized Pólya process and its applications. *Adv. Appl. Prob.* **46**, 1148–1171.
- [8] CHA, J. H. AND BADIA, F. G. (2019). On a multivariate generalized Pólya process without regularity property. *Prob. Eng. Inf. Sci.* **34**, 484–506.
- [9] DAW, A. AND PENDER, J. (2018). Queues driven by Hawkes processes. *Stoch. Systems* **8**, 192–229.
- [10] DEBICKI, K. AND ROLSKI, T. (2002). A note on transient Gaussian fluid models. *Queueing Systems* **41**, 321–342.
- [11] DEBICKI, K., KOSINSKI, K. AND MANDJES, M. (2012). Gaussian queues in light and heavy traffic. *Queueing Systems* **71**, 137–149.
- [12] DOOB, J. L. (1953). *Stochastic Processes*, Wiley, New York.
- [13] ETHIER, S. N. AND KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence*, Wiley, New York.
- [14] FELLER, W. (1968). *An Introduction to Probability Theory and its Applications*, Vol. I, 3rd edn. John Wiley, New York.
- [15] FELLER, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. II, 2nd edn. John Wiley, New York.
- [16] FENDICK, K. W. (2020). Brownian motion minus the independent increments: Representation and queueing application. *Prob. Eng. Inf. Sci.*, doi: [10.1017/S0269964820000388](https://doi.org/10.1017/S0269964820000388).

- [17] FENDICK, K. W. AND WHITT, W. (1989). Measurements and approximations to describe the offered traffic and predict the average workload in a single-server queue. *Proc. IEEE* **71**, 171–194.
- [18] GAO, X. AND ZHU, L. (2018). Functional central limit theorems for stationary Hawkes processes and application to infinite-server queues. *Queueing Systems* **90**, 161–206.
- [19] GLYNN, P. W. AND WHITT, W. (1991). A new view of the heavy-traffic limit for infinite-server queues. *Adv. Appl. Prob.* **23**, 188–209.
- [20] GREGOIRE, G. (1983). Negative binomial distributions for point processes. *Stochastic Proc. Appl.* **16**, 179–188.
- [21] HAHN, M. G. (1978). Central limit theorems in $D[0,1]$. *Z. Wahrscheinlichkeitsth.* **44**, 89–101.
- [22] HAJEK, B. (1994). A queue with periodic arrivals and constant service. In *Probability, Statistics and Optimization: A Tribute to Peter Whittle*, ed. F. P. Kelley. Wiley, Chichester, pp. 147–157.
- [23] HAWKES, A. G. (1971a). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58**, 83–90.
- [24] HAWKES, A. G. (1971b). Point spectra of some mutually exciting point processes. *J. R. Statist. Soc.* **33**, 438–443.
- [25] HONNAPPA, H., JAIN, R. AND WARD, A. (2015). A queueing model with independent arrivals and its fluid and diffusion limits. *Queueing Systems* **80**, 71–103.
- [26] ISAACSON, D. AND MADSEN, R. (1976). *Markov Chains: Theory and Applications*. Wiley, New York.
- [27] KONNO, T. H. (2010). On the exact solution of a generalized Polya process. *Adv. Math. Phys.* **2010**, 504267.
- [28] LI, A. AND WHITT, W. (2014). Approximate blocking probabilities for loss models with independence and distribution assumptions relaxed. *Performance Evaluation* **80**, 82–101.
- [29] LIU, Y. AND WHITT, W. (2011). Large-time asymptotics for the $G_t/M_t/s_t + GI_t$ many-server fluid model with customer abandonment. *Queueing Systems* **67**, 145–182.
- [30] MA, N. AND WHITT, W. (2019). Minimizing the maximum expected waiting time in a periodic single-server queue with a service-rate control. *Stochastic Systems* **9**, 261–290.
- [31] MIRTCEV, S. T. (2019). Study of preemptive priority single-server queue with peaked arrival flow. In *Proc. X Nat. Conf. with Int. Participation 'Electronica 2019'*, May 16–17, 2019, Sofia, Bulgaria.
- [32] MIRTCEV, S. T. AND GOLEVA, R. (2013). New constant service time Pólya/ D/n traffic model with peaked input stream. *Simul. Model. Pract. Theory* **34**, 200–207.
- [33] MIRTCEV, S. T. AND GANCHEV, I. (2016). Generalised Pollaczek–Khinchin formula for the Pólya/ $G/1$ queue. *Electron. Lett.* **53**, 27–29.
- [34] PÓLYA, G. AND EGGENBERGER, F. (1923). Über die Statistik verketteter Vorgänge. *Z. angewandte Mathematische Mechanik*, **3**, 279–289.
- [35] STIDHAM, S. (1974). A last word on $L = \lambda W$. *Operat. Res.* **22**, 417–421.
- [36] WHITT, W. (2002). *Stochastic Process Limits*. Springer, New York.
- [37] WHITT, W. (2015). Stabilizing performance in a single-server queue with time-varying arrival rate. *Queueing Systems* **81**, 341–378.
- [38] WHITT, W. AND YOU, W. (2018). Using robust queueing to expose the impact of dependence in single-server queues. *Operat. Res.* **66**, 184–199.
- [39] WHITT, W. AND YOU, W. (2019). The advantage of indices of dispersion in queueing approximations. *Operat. Res. Lett.* **47**, 99–104.
- [40] WOLFF, R. W. AND YAO, Y. (2014). Little's law when the average waiting time is infinite. *Queueing Systems* **76**, 267–281.