*Applied Probability Trust* (4 February 2020)

# QUEUES WITH PATH-DEPENDENT ARRIVAL PROCESSES

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

WARD WHITT,** *Columbia University*

### Abstract

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 $n$ i.i.d. Polya 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. The Polya process, the associated superposition process and the limit process are interesting because they exhibit path-dependent behavior; e.g., they each satisfy a non-ergodic law of large numbers. The average number of arrivals over time $[0, t]$ converges almost surely to a nondegenerate limit as $t \to \infty$.

*Keywords:* path-dependent stochastic processes, generalized Polya 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 condiitons. 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

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

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

Markov chains in Chapter V of [25]. 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 to them. The classic example is the familiar Polya urn model discussed in Feller [12], first studied by Polya and Eggenberger [34]. There is an urn containing $r$ red balls and $b$ blue balls. At each step, we select one ball in 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 Figure 1 of [2].

For the allocation of scarce resources in systems with path-dependent behavior, it is natural for queues to arise. 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. Polya point processes. A Polya point process can be represented as a limit of Polya urn models as indicated on p. 480 of [12]. Theorem 1 shows that the Polya point process is a stationary point process, while Theorem 2 shows that it satisfies a non-ergodic law of large numbers (LLN), which we use as our definition of path-dependence. Proposition 3 (from [7]) shows that the superposition process is a special generalized Polya process as considered by [27, 8], which inherits those properties.

We establish several results for queues with Polya point process arrival processes. Our main contribution here is Theorem 4, which establishes a heavy-traffic limit that provides, via Corollary 5, 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 3, 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, Ψ mnemonic for SI to denote Stationary Increments) studied in [14].

Here is how this paper is organized: In §2 we place our superposition process in the context of generalized Polya processes, as developed in [27, 8, 7]. (We regard §2 largely as a self-contained review.) In §3 we establish the FCLT for the superposition process and state various consequences. In §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 §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 §6 we obtain stability results for the single server queue and then establish steady-state results for queues with Polya 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 [38]. In §7 we present some remaining technical details. In §8 we present conclusions and discussion.

We conclude this introduction by discussing related work. First, we note that queues with Polya arrival processes have been considered previously as a way to capture exceptional variability by [31] and [32, 33]; we discuss that earlier work in Remark 3. Second, the heavy-traffic limit of the queue with a Ψ-GPP arrival process 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, 9]. The FCLT for the Ψ-GPP here yields a Ψ-GMP with positive dependence, i.e., in which the increments over disjoint intervals are positively correlated; see Corollary 3. As we indicate in §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 [19, 23].

## 2. Generalized Polya Point Process with Stationary Increments: Ψ-GPP

The Polya point process has been extended to the generalized Polya process (GPP) by Konno [27] and Cha [8]. A GPP $N \equiv \{N(t) : t \geq 0\}$ is a Markov point process with stochastic intensity (defined in terms of the internal histories $\mathcal{H}_t$; e.g., see §1.8 of [3]) by

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

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

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

Many properties of the GPP were deduced in [8] by exploiting the restarting property.

**Proposition 1.** (the restarting property, [8].) *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 [8] 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, [8].) *If $N$ is a GPP with parameter triple $(\kappa(t), \gamma, \beta)$, then $N(t)$ has a negative binomial distribution with probability mass function (pmf)*

$$P(N(t) = k) \equiv f(k; r, p(t)) = C(\beta, \gamma, k)(1 - p(t))^r p(t)^k, \quad k = 0, 1, 2, \ldots \tag{3}$$

*where*

$$r = \beta/\gamma, \quad p(t) = 1 - exp\{-\gamma K(t)\}, \quad K(t) \equiv \int_0^t \kappa(s)\, ds, \quad t \geq 0, \tag{4}$$

*and*

$$C(\beta, \gamma, k) = \Gamma((\beta/\gamma) + k)/\Gamma((\beta/\gamma))k! \tag{5}$$

*with $\Gamma$ being the gamma function, so $N(t)$ has mean and variance*

$$E[N(t)] = \frac{rp(t)}{1 - p(t)} \quad and \quad Var(N(t)) = \frac{rp(t)}{(1 - p(t))^2}, \quad t \geq 0. \tag{6}$$

For general function $\kappa(t)$, the time-varying behavior can be complicated, but it simplifies for the classical Polya 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 [14], we thus call the GPP with triple $(\kappa(t), \gamma, \beta)$ for $\kappa(t)$ in (2) a $(\beta, \gamma)$ $\Psi$-GPP.

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

$$E[N(t)] = \beta t \quad \textit{and} \quad Var(N(t)) = \beta t(1 + \gamma t), \quad t \geq 0. \tag{7}$$

*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*

$$Cov(N(s), N(t)) = \beta s(1 + \gamma t), \quad 0 \leq s \leq t < \infty. \tag{8}$$

*Proof.* For $\kappa$ in (2),

$$K(t) \equiv \int_0^t (\gamma s + 1)\, ds = \gamma^{-1} \log(\gamma t + 1), \quad t \geq 0, \tag{9}$$

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 Theorem 1 (ii) of [8]. For the covariance in (8), write

$$Cov(N(s), N(t)) = [Var(N(t)) + Var(N(s)) - Var(N(t - s))]/2$$

and then use the variance in (7). 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 the represent a partion of a fixed interval $(s, s + t]$ into finitely many subintervals. We then apply Theorem 3 and Remark 3 of [8] 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

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

If $\kappa(t)$ is given by (2), then $f(x) = 1/t$; i.e., the conditional distribution for the ordered sequence of times when the GPP $N$ increases on $(s, t]$ given that $N(t) - N(s) =$

$k$. That in turn implies that the joint distribution of the $k$ disjoint and contiguous increments $N(s_i+t_i+h)-N(s_i+h)$, $1 \le i \le k$, 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 $h$. □

**Remark 1.** (*application of strict stationarity.*) We are primarily interested in a single increment, so a weaker definition of stationarity than strict stationarity often suffices. However, we apply the stronger definition here in our proof of Theorem 5 about an infinite-server queue with a Polya arrival process in §6.2.

We next apply Theorem 1 to characterize the nature of the path-dependent behavior for a $(\beta,\gamma)$ $\Psi$-GPP. We do that from the following non-ergodic law of large numbers (LLN); e.g., see §5.1 of [17] and references there.

**Theorem 2.** (non-ergodic LLN.) *If $N(t)$ is $(\beta,\gamma)$ $\Psi$-GPP, then*

$$t^{-1}N(t) \to L(\gamma,\beta) \quad \textit{as} \quad t\to\infty \quad \textit{w.p.1,} \tag{10}$$

*where $L$ has a gamma distribution with shape $\beta/\gamma$ and rate $1/\gamma$, and thus mean $E[L]=\beta$ and variance $Var(L)=\beta\gamma$.*

*Proof.* Because the increments $N(n+1)-N(n)$ form a stationary sequence, we can apply the Birkhoff ergodic theorem as in Theorem 6.2.1 of [6] to establish the almost sure convergence. Next, from (7), we see that $E[N(t)/t]=\beta$ for all $t>0$ and $Var(N(t)/t)=\beta(1+\gamma t)/t \to \beta\gamma$ as $t\to\infty$. The limiting gamma distribution for $N(t)/t$ is obtained by taking a limit as $t\to\infty$ 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

$$\begin{aligned}
\phi_{t^{-1}N(t)}(s) &= \left(\frac{1-p(t)}{1-p(t)e^{is/t}}\right)^{\beta/\gamma} = \left(\frac{1}{1+[p(t)/(1-p(t))](1-e^{is/t})}\right)^{\beta/\gamma} \\
&= \left(\frac{1}{1+\gamma t(1-e^{is/t})}\right)^{\beta/\gamma} = \left(\frac{1}{1+\gamma t(1-(1+(is/t)+O(1/t^2))}\right)^{\beta/\gamma} \\
&\to \left(\frac{1}{1-is\gamma}\right)^{\beta/\gamma} \quad \text{as} \quad t\to\infty.
\end{aligned} \tag{11}$$

Finally, we recognize the limit as the cf of the claimed gamma distribution. Convergence of characteristic functions then applies the convergence in distribution by Theorem XV.2 of [13]. □

Theorem 2 implies that the pure birth process $N$ has a limiting rate as $t \to \infty$, but that rate is random.

**Corollary 1.** (asymptotically Poisson with a random rate.) *If $N(t)$ is a $(\beta, \gamma)$ $\Psi$-GPP, with stochastic intensity $\lambda(t)$ in* (1)*, then*

$$\lambda(t) \to L(\gamma, \beta) \quad as \quad t \to \infty \quad w.p.1, \tag{12}$$

*where $L$ is the gamma random variable in* (10) *above with shape $\beta/\gamma$ and rate $1/\gamma$. Hence, asymptotically as $t \to \infty$, 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 $\gamma L$ by Theorem *2*, while the denominator converges to $\gamma$. □

We conclude this section by making three remarks.

**Remark 2.** (*index of dispersion.*) To better understand the impact of the variability as a function of time in an arrival process upon the performance of a queueing model, we have shown in [15] and [39] that it is often helpful to look at the index of dispersion for counts, which for a $(\beta, \gamma)$ $\Psi$-GPP is

$$I(t) \equiv \frac{Var(N(t))}{EN(t)} = 1 + \gamma t, \quad t \geq 0. \tag{13}$$

From (13), we see that the variability increases without bound as $t$ increases by this measure, consistent with Theorem 2. The IDC is also considered in [31] and [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 there.

**Remark 3.** (*instability of the $P/GI/1$ queue.*) Theorem 2 and Corollary 1 imply that the queue length process is not stable in the $P/GI/1$ queue with a Polya arrival process; i.e., there does not exist a random variable $Q$ with $P(Q < \infty) = 1$ such that $Q(t) \Rightarrow Q$ as $t \to \infty$, where $\Rightarrow$ denotes convergence in distribution. That contradicts various conclusions about steady-state performance in [31] and [32, 33]. We elaborate on stability and discuss ways to stabilize performance in queues with Polya arrival processes in §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 4 and Corollary 5.

**Remark 4.** (*comparison to the Hawkes process.*) The non-degenerate limit in Theorem 2 makes the $\Psi$-GPP quite different from the widely applied Hawkes [21, 22] process and most of its variants. These are alternative self-exciting processes, but they are are stationary and ergodic point processes; e.g., see [4]. For the basic Hawkes process in (8) of [21], instead of (1) we have

$$\lambda(t) \equiv \lambda(t|\mathcal{H}_t) \equiv \nu + \int_{-\infty}^{t} g(t-u)\, dN(u) \tag{14}$$

where $g$ is a nonnegative kernel satisfying $\eta \equiv \int_0^\infty g(u)\, du < 1$, so that the stationary rate is $\nu/(1-\eta)$. The special case in which $g(u) = ae^{-bu}$, where $a, b > 0$ and $\eta = a/b < 1$ makes $(\lambda(t), N(t))$ jointly Markov. If we let $b$ increase, then the process behaves locally much like the Polya process. If we let $b \rightarrow \infty$ and $\nu \rightarrow 0$ such that $b\nu = 1$ and $b - a = 1$, then the process has rate 1 in all cases. Moreover, its local behavior approaches that of the Polya process.

## 3. Convergence to a $\Psi$-GMP

We now show that a properly scaled sequence of the superpositon of i.i.d. $\Psi$-GPP's in §2 converges to a $\Psi$-GMP, the Gaussian Markov process with stationary increments studied in [14]. (In fact, [14] focuses on a multivariate $\Psi$-GMP.) We obtain all possible univariate $\Psi$-GMP's exhibiting positive dependence, as we explain in §8.

For $n \geq 1$, let

$$A^n = N^1 + \cdots + N^n \tag{15}$$

be the sum of $n$ i.i.d. GPP's each with parameter triple $(\kappa(t), \gamma, \beta)$. We first note that our superposition process is another GPP.

**Proposition 3.** (superposition, Theorem 1 of [7].) *The superposition of two independent GPP's 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 $\Psi$-GPP, then so is the superposition process. Then the superposition process satisfies the non-ergodic LLN in Theorem 2.*

We now apply the usual FCLT spacial scaling, but without scaling time by $n$ (as in (2.1) on p. 226 or (8.4) on p. 320 of [37]). In particular, for $n \geq 1$, let

$$A_n(t) \equiv n^{-1/2}(A^n(t) - \beta n t), \quad t \geq 0, \tag{16}$$

Let $\Rightarrow$ denote convergence in distribution and let $D \equiv D[0, \infty)$ be the usual function space of right continuous real-valued functions; e.g., as in [5] or [37].

**Theorem 3.** (FCLT for the superposition process.) *Consider the scaled superposition process $A_n(t)$ in* (16). *For $\kappa(t)$ in* (2), *so that in* (15) *$N^1$ is a $(\beta, \gamma)$ $\Psi$-GPP while $A^n$ is an $(n\beta, \gamma)$ $\Psi$-GPP,*

$$A_n \Rightarrow A \quad in \quad D \quad as \quad n \to \infty, \tag{17}$$

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

$$Cov(A(s), A(t)) = E[A(s)A(t)] = \beta s(1 + \gamma t) = Cov(N^1(s), N^1(t)). \tag{18}$$

*The limit $A$ also satisfies the stochastic differential equation*

$$dA(t) = \mu(t)A(t) + \sigma B(t), \quad t \geq 0, \tag{19}$$

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

$$\mu(t) \equiv \frac{\beta\gamma}{\beta + \beta\gamma t} = (t + (1/\gamma)^{-1}) \quad and \quad \sigma = \sqrt{\beta}. \tag{20}$$

*Proof.* For the limit in (17), we apply Hahn's [18] FCLT for sums of processes in Theorem 7.2.1 of [37]. We verify the moment inequality conditions in that theorem in §7. The SDE characterization in (19) and (20) follows from Theorem 3 of [14]. A Gaussian process with that covariance kernel is a Markov process by Theorem 8.1 on p. 233 of [11]. □

**Remark 5.** (*parameters.*) Even though (18) shows identical structure in the covariance functions of the $\Psi$-GPP $N^1$ in (15) and the $\Psi$-GMP $A$ in (17), the conventions here for the parameters are not the same as in [14]. When $N^1$ is a $(\beta, \gamma)$ $\Psi$-GPP, $A$ is an $(\alpha^*, \beta^*) = (\beta, -\beta\gamma)$ $\Psi$-GMP in [14].

**Remark 6.** (*structural analogs.*) Many properties of the $\Psi$-GMP were established in [14]. Properties also can be deduced as a consequence of Theorem 3. Lemma 4 of

[14] established an analog of the restarting property in Proposition 1. A variant of the proof of Theorem 2 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 3 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

$$\mu_n \to 1 \quad \text{and} \quad \sqrt{n}(\mu_n - 1) \to \mu \quad \text{as} \quad n \to \infty. \tag{21}$$

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

$$A_n^d(t) = n^{-1/2}(A^{d,n}(t) - \beta n t) = n^{-1/2}(A^n(\mu_n t) - \beta n t), \quad t \geq 0. \tag{22}$$

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 2.** (FCLT with a drift.) *If* (21) *holds in addition to the assumptions of Theorem* 3*, then*

$$A_n^d \Rightarrow A + \mu\beta e \quad \textit{in} \quad D \quad \textit{as} \quad n \to \infty \tag{23}$$

*for* $A_n^d$ *in* (22)*.*

*Proof.* We apply the continuous mapping argument for composition with centering as in §13.3 of [37]. For that purpose, let

$$(\bar{M}_n(t), M_n(t)) \equiv \left(\mu_n t, \sqrt{n}(\mu_n - 1)t\right). \tag{24}$$

Note that $A_n^d = A_n \circ \bar{M}_n + M_n$, where $\circ$ denotes the composition map. It is elementary that

$$(\bar{M}_n, M_n) \to (e, \mu e) \quad \text{in} \quad D^2 \quad \text{as} \quad n \to \infty$$

Then apply Theorem 11.4.5 of [37] with Theorem 3 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 Theorem 13.3.1 of

[37] yields

$$A_n^d = (A_n \circ \bar{M}_n + \beta M_n) \Rightarrow A + \mu\beta e \quad \text{in} \quad D \quad \text{as} \quad n \to \infty.$$

□

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 [14].) *If $A^d \equiv A + \omega e$ as in Corollary 2, where $A$ is a $\Psi$-GMP satisfying* (18) *in Theorem* 3*, then*

$$E[A^d(s+t) - A^d(s) | A^d(u), 0 \le u \le s] = \omega(s)t \quad \textit{for all} \quad s, t \ge 0, \tag{25}$$

*where*

$$\omega(s) \equiv \omega + \gamma(1+\gamma s)^{-1}(A^d(s) - s\omega). \tag{26}$$

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

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

**Corollary 3.** (correlation between non-overlapping time intervals, Proposition 2 of [14].) *For $A^d$ as in Proposition* 4*,*

$$Cor(A^d(t+s+u) - A^d(t+u), A^d(t+s) - A^d(t)) = \frac{\gamma s}{\gamma s + 1} \tag{27}$$

*for all $t \ge 0$ and $u \ge s \ge 0$.*

Corollary 3 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 $\Phi(x) \equiv P(N(0,1) \le x)$ be the standard normal cdf and let $\Phi^c(x) \equiv 1 - \Phi(x)$.

**Corollary 4.** (limiting cdf.) *For $A^d$ as in Proposition* 4*,*

$$\lim_{t\to\infty} P(A^d(t) \le x) = \Phi(-\omega/\sqrt{\beta\gamma}) \tag{28}$$

*for all $x$ and all $t \ge 0$. Moreover,*

$$\begin{aligned} \lim_{x\to-\infty} \lim_{t\to\infty} P(A^d(t) \le x) &= \Phi(-\omega/\sqrt{\beta\gamma}) \\ \lim_{x\to+\infty} \lim_{t\to\infty} P(A^d(t) > x) &= \Phi^c(-\omega/\sqrt{\beta\gamma}). \end{aligned} \tag{29}$$

*Proof.* We can directly take the limit in the Guassian distribution of $A^d(t)$, which is

$$P(A^d(t) \leq x) = \Phi\left(\frac{(x - \omega t)}{\sqrt{t(1 + \beta\gamma t)}}\right),$$

getting (29). □

Corollary 4 can be understood by recognizing the the standard deviation is the same order $t$ as the mean; i.e., $E[A^d(t)] = \omega t$ and $t^{-1}\sqrt{Var(A^d(t)} \to \sqrt{\beta\gamma}$ as $t \to \infty$. Thus, $P(A^d(t) > 0)$ 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 (22). 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 (scv, 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 $\beta nt$ in (22), we assume that the rate of this renewal process is also $\beta$. Let the scaled service process be

$$S_n(t) \equiv n^{-1/2}(C(nt) - \beta nt), \quad t \geq 0, \tag{30}$$

As in §9.3 of [37], 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

$$Q_n(t) \equiv n^{-1/2}Q^n(nt), \quad t \geq 0, \tag{31}$$

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 $\{A^{d,n}(t) : t \geq 0\}$ defined before (22) and service renewal counting process $\{C(t) : t \geq 0\}$ defined above.

With those definitions, we apply a modification of the one-dimensional reflection map in §13.5 of [37]. Let $\phi : D \times R \to D$ be the reflection map mapping a net input function $x$ with $x(0) = 0$ and an initial queue length $q(0)$ into $q(t)$ for $t > 0$ by

$$\phi(x)(t, q(0)) = q(0) + x(t) - \inf_{0 \leq s \leq t}\{\min\{q(0) + x(s), 0\}\}, \quad t \geq 0. \tag{32}$$

Like the other reflection maps, the reflection map in (32) is a continuous function on its domain.

**Theorem 4.** (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 arrival process $n$ is the scaled superposition process $A_n^d(t)$ in* (22) *and* (15)*, while the scaled service process is the scaled renewal counting process $S_n(t)$ in* (30)*. For the initial conditions, let the arrival process after time* 0 *be independent of $Q_n(0)$; let the remaining service time in 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 $n^{-1}Q_n(0) \Rightarrow Q(0)$ as $n \to \infty$ and $X_n \equiv A_n^d - S_n$, then*

$$(A_n^d, S_n, X_n, Q_n) \Rightarrow (A + \omega e, S, X, Q) \quad in \quad D^4 \quad as \quad n \to \infty, \tag{33}$$

*where $\omega \equiv \mu\beta$, $A$ is a $\psi$-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 $\phi$ in* (32)*. In particular, $Y$ is a $\Psi$-GMP with*

$$E[Y(s)Y(t)] = \beta s(1 + \gamma t) + \beta^3 c_s^2 s \quad and \quad Var(Y(t)) = \beta t(1 + \gamma t) + \beta^3 c_s^2 t, \tag{34}$$

*for $0 \le s \le t$ and so parameter pair $(\alpha^*, \beta^*) = (\beta + \beta^3 c_s^2, -\beta\gamma)$ in the terminology of* [14]*, while $X$ is a $\Psi$-GMP with $Var(X(t)) = Var(Y(t))$ and deterministic drift $\omega \equiv \mu\beta$.*

*Proof.* We apply standard methodology for establishing a heavy-traffic FCLT for a single-server queue. We apply Donsker's FCLT for the service times in §4.3 of [37] and the inverse equivalence in Theorem 7.3.2 of [37], in particular Corollary 7.3.2 on p. 236 of [37], to obtain

$$S_n \Rightarrow \beta^{3/2} c_s B \quad \text{in} \quad D \quad \text{as} \quad n \to \infty, \tag{35}$$

where $B$ is a standard Brownian motion (BM). Then we can apply Corollary 2 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 Theorem 11.4.4 of [37]. We can then apply Theorems 9.3.3, 9.3.4 and 9.8.2 in [37]. □

By the continuous mapping theorem with the projection map, we have the following corollary providing a limit for the marginal distributions. Theorem 5 of [14] 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

$$f(x_s, q_s, q_{s+t}) \equiv f_{X(s),Q(s),Q(s+t)}(x_s, q_s, q_{s+t}) \tag{36}$$

and similarly for the associated marginal pdf's and conditional pdf's. We express the limiting distribution in terms of the exponential function and the standard normal cdf $\Phi(x) \equiv P(N(0,1) \leq x)$ and pdf $\phi(x)$. Let the associated cdf of $(N(m, \sigma^2) \overset{\mathrm{d}}{=} m + \sigma N(0,1)$ be denoted by $\Phi(x; m, \sigma^2)$ and similarly for the others. To connect with [14], let

$$\omega_s \equiv \frac{\alpha^* \omega - \beta^* x_s}{\alpha^* - \beta^* s} \quad \text{and} \quad \beta_s^* \equiv \frac{\alpha^* \beta^*}{\alpha^* - \beta^* s}, \tag{37}$$

where

$$\alpha^* \equiv \beta \quad \text{and} \quad \beta^* \equiv -\beta\gamma \tag{38}$$

as in Remark 5. Let $\delta(\cdot)$ be the Dirac delta function and let $1_A$ be the indicator function, equal to 1 on $A$ and 0 elsewhere.

**Corollary 5.** (marginal limiting distributions.) *Under the conditions of Theorem* 4,

$$(X_n(s), Q_n(s), Q_n(s+t)) \Rightarrow (X(s), Q(s), Q(s+t)) \quad in \quad \mathbb{R}^3 \quad as \quad n \to \infty, \tag{39}$$

*where $X(s)$ is a (mean-$\omega s$, variance $s(\alpha^* - \beta_s^* s)$) Gaussian random variable for $\beta_s^*$ in* (37)*, while the joint limiting distribution has joint pdf*

$$f(x_s, q_s, q_{s+t}) = f(x_s) f(q_s | x_s) f(q_{s+t} | x_s, q_s), \tag{40}$$

*where, assuming $P(Q(0) = q_0) = 1$,*

$$\begin{aligned}
f(x_s) &\equiv \phi(x_s; \omega s, s(\alpha^* - \beta^* s)) = \frac{1}{\sqrt{s(\alpha^* - \beta^* s)}} \phi\left(\frac{x_s - \omega s}{\sqrt{s(\alpha^* - \beta^* s)}}\right), \\
f(q_s | x_s) &\equiv \left(1 - e^{\{-2q_s)(q_s - x_s)/(\alpha^* s)\}}\right) \delta(q_s - q_0 - x_s) \\
&\quad + \left(\frac{(4q_s - 2x_s) e^{\{-2q_s(q_s - x_s)/(\alpha^* s)\}}}{\alpha^* s}\right) 1_{\{q_s \geq q_0 + x_s\}} \quad \mathit{for} \quad q_s \geq 0, \quad \mathit{and} \\
f(q_{s+t} | x_s, q_s) &\equiv \frac{1}{\sqrt{t(\alpha^* - \beta_s^* t)}} \phi\left(\frac{q_{s+t} - q_s - \omega_s t}{\sqrt{t(\alpha^* - \beta_s^* t)}}\right) + e^{\{-2q_{s+t}(\beta_s^* q_{s+t} - \alpha^* \omega_s)/\alpha^{*2}\}} (A_1 + A_2) \\
\mathit{for} \quad A_1 &\equiv \left(\frac{4\beta_s^* q_{s+t} - 2\alpha^* \omega_s}{\alpha^{*2}}\right) \Phi\left(\frac{(2\beta_s^* q_{s+t} - \alpha^* \omega_s) t - \alpha^* (q_{s+t} + q_s)}{\alpha^* \sqrt{t(\alpha^* - \beta_s^* t)}}\right) \\
\mathit{and} \quad A_2 &\equiv \left(\frac{\alpha^* - 2\beta_s^* t}{\alpha^* \sqrt{(t(\alpha^* - \beta_s^* t)}}\right) \phi\left(\frac{(2\beta_s^* q_{s+t} - \alpha^* \omega_s) t - \alpha^* (q_{s+t} + q_s)}{\alpha^* \sqrt{t(\alpha^* - \beta_s^* t)}}\right)
\end{aligned} \tag{41}$$

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

$$P(Q(s+t) \leq q_{s+t}|X(s) = x_s, Q(s) = q_s) = \Phi\left(\frac{q_{s+t} - q_s - \omega_s t}{\sqrt{t(\alpha^* - \beta_s^* t)}}\right)$$
$$-e^{\left(\frac{-2q_{s+t}(\beta_s^* q_{s+t} - \alpha^* \omega_s)}{\alpha^{*2}}\right)}\Phi\left(\frac{(2\beta_s^* q_{s+t} - \alpha^* \omega_s)t - \alpha^*(q_{s+t} + q_s)}{\alpha^* \sqrt{t(\alpha^* - \beta_s^* t)}}\right). \qquad (42)$$

*and, for* $q_t \geq 0$*,*

$$P(Q(t) \leq q_t) = \Phi\left(\frac{q_t - q_0 - \omega t}{\sqrt{t(\alpha^* - \beta^* t)}}\right)$$
$$-e^{\left(\frac{-2q_t(\beta^* q_t - \alpha^* \omega)}{\alpha^{*2}}\right)}\Phi\left(\frac{(2\beta^* q_t - \alpha^* \omega)t - \alpha^*(q_t + q_0)}{\alpha^* \sqrt{t(\alpha^* - \beta^* t)}}\right). \qquad (43)$$

*Proof.* We give a brief overview of the proof in [14]. We focus on $X(t) \equiv Y(t) + \omega t$, where $Y$ is an $(\alpha^*, \beta^*)$ $\Psi$-GMP for $(\alpha^*, \beta^*) = (\beta + \beta^3 c_s^2, -\beta\gamma)$ and $\omega t$ is the drift, as in Theorem 4. 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 $(\alpha^*, \beta_s^*)$ $\Psi$-GMP for $\beta_s^*$ in (37) 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 $(\alpha^*, t^{-1}\alpha^*)$ $\Psi$-GMP with drift $t^{-1}x_t^{(s)}$ on $[0, t]$ (which no longer depends on $\beta_s^*$). 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 $(\alpha^*, t^{-1}\alpha^*)$ $\Psi$-GMP on the interval $[0, t]$. That process is scaled Brownian bridge, for which the distribution of the queue length was previously obtained by [19]. The result in (42) is obtained from that conditional queue length distribution using the law of total probability. We remark that $f(q_s|x_s)$ in (40) is the density of the cdf

$$P(Q(s) \leq q_s|X(s) = x_s) = \left(1 - e^{\{-2q_s(q_s - x_s)/(\alpha^* s)\}}\right)1_{\{q_s \geq q_0 + x_s\}} \qquad (44)$$

as derived by [19].

□

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 $\sigma^2$ and let $(x)^+ \equiv \max\{x, 0\}$, so that $P(N(m,\sigma^2)^+ = 0) = P(N(m,\sigma^2) \leq 0)$.

**Corollary 6.** (LLN limit for the limit process.) *Given the $\Psi$-GMP limit in Theorem 3 and the associated heavy-traffic limit in Theorem 4 and Corollary 5, we have the following limit*

$$\begin{aligned} & t^{-1}(A^d(t), S(t), X(t), Q(t), X_s(t), Q_s(t)) && (45) \\ & \Rightarrow (N_1(\omega,\beta\gamma), 0, N_1(\omega,\beta\gamma), N_1(\omega,\beta\gamma)^+, N_2(\omega_s, -\beta_s^*), N_2(\omega_s, -\beta_s^*)^+) \quad in \quad \mathbb{R}^6 \end{aligned}$$

*as $t \to \infty$ for some constant $s > 0$ and $(\omega_s, \beta_s^*)$ in (38), 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 4. For the next two processes, note that $P(X(t) \leq xt) \to \Phi((x-\omega)/\sqrt{\beta\gamma})$ as $t \to \infty$ as just described and

$$\begin{aligned} P(Q(t) \leq tq | X(t) = tx) &= (1 - e^{-(2t^2 q(q-x))/\alpha s}) 1_{\{tq - tx \geq q_0\}} \\ &= (1 - e^{-(2t^2 q(q-x))/\alpha s}) 1_{\{t \geq q_0/(q-x)\}} 1_{\{q-x>0\}} \\ &\to 1_{\{q-x>0\}} \quad \text{as} \quad t \to \infty. \qquad (46) \end{aligned}$$

. The joint pdf for the two limits is therefore

$$\bar{f}(x,q) \equiv \bar{f}(x)\bar{f}(q|x) = \frac{1}{\sqrt{\beta\gamma}} \phi((x-\omega)/\sqrt{\beta\gamma})\delta(q-x). \qquad (47)$$

The joint pdf for the limits of the last two processes is similarly obtained. □

The limit for the last two components in Corollary 6 shows that the limit as $t \to \infty$ is not independent of fixed $s$. Corollary 6 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

In this section we present further results for queues with Ψ-GPP arrivals. In §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 §6.2 we show that stability can be achieved when there are infinitely many servers. In §6.3 we show that it is also possible to devise adaptive service policies that stabilize performance in a single-server queue.

### 6.1. Stability and Little's Law

We first discuss the implications of the LLN in Theorem 2 for the $P/GI/1$ queue with a Polya arrival process. For this purpose, let the service times have mean 1. Given Theorem 2, the following is a standard heavy-traffic law of large numbers for overloaded queues, as in Theorem 5.3.2 of [37].

**Corollary 7.** (explosion.) *Let $Q(t)$ be the queue length process, starting with $Q(0) = 0$, in the $P/GI/1$ queue with a $(\beta, \gamma)$ $\Psi$-GPP arrival process and i.i.d. service times with mean* 1*. Then*

$$t^{-1}Q(t) \to \min\{L(\beta,\gamma) - 1, 0\} \quad \textit{as} \quad t \to \infty \quad \textit{w.p.1}, \tag{48}$$

*so that*

$$P(Q(t) \to \infty \quad \textit{as} \quad t \to \infty) = P(L(\beta,\gamma) > 1), \tag{49}$$

*where*

$$0 < P(L(\beta,\gamma) > 1) < 1. \tag{50}$$

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

$$L^* \equiv \lim_{t\to\infty} t^{-1} \int_0^t Q(s)\,ds, \quad W^* \equiv \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} W_k \quad \text{and} \quad \lambda^* \equiv L(\beta,\gamma). \tag{51}$$

**Corollary 8.** (Little's law.) *Consider the $P/GI/1$ queue with a $(\beta, \gamma)$ $\Psi$-GPP arrival process starting with $Q(0) = 0$ as above.*

(*a*) *If the definition of $W^*$ in* (51) *is well defined and finite and if $L(\beta,\gamma) < 1$, which occurs with positive probability, then the definition of $L^*$ in* (51) *is well defined and*

$$L^* = \lambda^* W^* < \infty. \tag{52}$$

(*b*) *If* $L(\beta, \gamma) > 1$*, as occurs with positive probability, then the definition of* $L^*$ *in* (51) *is well defined but* $L^* = \infty$*. Moreover, the definition of* $W^*$ *is then also well defined and*

$$W^* = L^* = \infty. \tag{53}$$

*Proof.* For part (a), we apply the sample-path version of Little's law from [35]. For part (b), we apply Corollary 7 together with (6) in [40]. 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 ensures reaching steady state in finite time. Let $\stackrel{\mathrm{d}}{=}$ denote equality in distribution.

**Theorem 5.** (infinite-server queue with bounded service times.) *For an infinite-server queue with a* $(\beta, \gamma)$ 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)* $P(V \le \zeta) = 1$ *for some* $\zeta < \infty$*, then the number of busy servers* $Q(t)$ *reaches steady state by time* $\zeta$*, i.e.,*

$$Q(u) \stackrel{\mathrm{d}}{=} Q(t) \quad \textit{and} \quad E[Q(u)] = \beta E[V] \quad \textit{for all} \quad u \ge t \ge \zeta. \tag{54}$$

*Proof.* We apply a useful device from §2 of [16]. We note that if the service times are deterministic with $P(V = d) = 1$, then

$$Q(t) = N(t) - N(t-d) \quad \text{for all} \quad t \ge d. \tag{55}$$

Because the Ψ-GPP arrival process has stationary increments, the distribution of $Q(t)$ in (55) is independent of $t$ for $t \ge d$, 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 $P(V = d_i) = q_i$, $1 \le i \le n$. 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

$$Q(t)=\sum_{i=1}^{n}Q_i(t)=\sum_{i=1}^{n}[N(t)-N(t-d_i)]T_i(t,d_i) \tag{56}$$

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 $(N(t)-N(t-d_i):1\le i\le n)$ is independent of $t$ in $\mathbb{R}^n$ for all $t\ge \max\{d_i:1\le i\le n\}$.

The joint distribution of $(T_i(t,d_i):1\le i\le n)$ is somewhat complicated, but we can directly deduce that it has property (54) and we can write down the mean

$$\begin{aligned} E[Q(t)] &= \sum_{i=1}^{n}E[Q_i(t)]=\sum_{i=1}^{n}E[N(t)-N(t-d_i)]E[T_i(t,d_i)] \\ &= \sum_{i=1}^{n}\beta d_i E[T_i(t,d_i)]=\beta E[V]. \end{aligned} \tag{57}$$

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

$$Q(t)=\sum_{i=1}^{N(t)}1_{\{T_i+V_i>t\}} \tag{58}$$

Representation (58) 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 in Theorem 3.4.4 of [37]. 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

$$\sum_{i=1}^{\infty}P(t\le T_i\le t+\epsilon)=P(N(t+\epsilon)-N(t)\ge 1)\le E[N(t+\epsilon)-N(t)]=\beta\epsilon \tag{59}$$

by (7). We see that the probability is 0 by letting $\epsilon\downarrow 0$. 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. □

### 6.3. Adaptive Service Processes to Enforce Stability

Given the non-ergodic LLN for the GPP in Theorem 2, we see that, for any constant service rate, there is a strictly positive probability that the $GPP/GI/1$ queue is stable in the long run, but also a strictly positive probability that it is unstable. That motivates us to consider alternative adaptive service processes that always achieve stability. We now show one way that can be done.

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

$$\mu(t) \equiv \mu(t|\mathcal{H}_t) = \frac{\lambda(t)}{\rho} = \frac{\gamma N(t) + \beta}{\rho(\gamma(t) + 1)} \quad \text{for all} \quad t \geq 0. \tag{60}$$

The proposed control is a variant of the rate-matching service-rate control in [30, 38]. The following is an analog of Theorem 3.1 of [38].

**Theorem 6.** *If the arrival-history-dependent service process in* (60) *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* $\kappa$ *in* (2)*, then the pair of arrival and service processes* $(N(t), S(t))$ *satisfies the joint non-ergodic LLN*

$$t^{-1}(N(t), S(t)) \to (L(\gamma, \beta), L(\gamma, \beta)/\rho) \quad \textit{as} \quad t \to \infty \quad \textit{w.p.1}, \tag{61}$$

*where* $L(\gamma, \beta)$ *is the gamma random variable in Theorem* 2*. 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*

$$Q(t) \Rightarrow Q \quad \textit{as} \quad t \to \infty, \quad \textit{where} \quad P(Q = k) = (1 - \rho)\rho^k, \quad k \geq 0, \tag{62}$$

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

*Proof.* Observe that definitions (1) and (60) make the the queue length process a birth and death process in a random environment, starting out at $Q(0) = 0$. Potential transitions occur at the random rate $R(t) \equiv \lambda(t)(1 + \rho^{-1})$ at time $t$, which converges almost surely to $L(\gamma, \beta)(1 + \rho^{-1})$ as $t \to \infty$. Whenever a transition occurs, it is a jump up with the constant probability $\rho/(1 + \rho)$; it is a jump down otherwise. When the queue is empty the transition down does not actually occur. Thus the evolution is the same as for the $M/M/1$ queue in the random environment with potential transitions occurring at $R(t)$, which converges to the random limit. □

## 7. Completing the Proof of Theorem 3

We complete the proof of Theorem 3 by verifying the two inequality conditions in Hahn's theorem [18] as in Theorem 7.2.1 of [37]. To do so, we prove convergence in $D[0,1]$, as in [18], but we note that essentially the same arguments shows convergence in $D[0,T]$ for any $T>0$ and therefore implies convergence in $D[0,\infty)$; see Section 12.9 of [37].

By Theorem 1, $E[A_1(s)A_1(t)] = \beta s(1+\gamma t)$, when $A_1$ is the centered process defined in (16). First, for $0 \leq t_1 \leq t_2 \leq 1$,

$$\begin{aligned} E\left[(A_1(t_2)-A_1(t_1))^2\right] &= \beta\left(t_2-t_1\right)\left(1+\gamma\left(t_2-t_1\right)\right) \\ &= \beta\left(t_2+\gamma t_2^2\right)-\beta\left(t_1+\gamma t_1^2\right)-2\beta\gamma t_1\left(t_2-t_1\right) \\ &\leq \beta\left(t_2+\gamma t_2^2\right)-\beta\left(t_1+\gamma t_1^2\right). \end{aligned} \tag{63}$$

so that condition (2.3) of Theorem 7.2.1 in [37] is met by (63) here.

We will also show that

$$E[(A_1(t)-A_1(t_1))^2(A_1(t_2)-A_1(t))^2] \leq c\left(t_2-t_1\right)^2 \tag{64}$$

for $0 \leq t_1 \leq t \leq t_2 \leq 1$ and a constant $c$, so that condition (2.4) of Theorem 7.2.1 in [37] will be met as well.

To do explicit calculations, we again apply Theorem 3 and Remark 3 of [8], 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

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

For a $\Psi$-GPP, $p(x)=1$. Hence, the conditional distribution for the ordered sequence of times when the $\Psi$-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 Theorem 14.3 of [5], 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 $V = \sum\limits_{j=1}^{k} V_j$, $W = \sum\limits_{j=1}^{k} W_j$, $\tilde{V} = \big(N^1(1) - \beta\big)\, p_1$ and $\tilde{W} = \big(N^1(1) - \beta\big)\, p_2$. Then

$$\begin{aligned}
&E\left[\left(A_1(t) - A_1(t_1)\right)^2 \left(A_1(t_2) - A_1(t)\right)^2 |N^1(1) = k\right] \\
&\quad = E\left[\left(\sum_{j=1}^{k}\left(V_j + \frac{(k-\beta)\,p_1}{k}\right)\right)^2 \left(\sum_{j=1}^{k}\left(W_j + \frac{(k-\beta)\,p_2}{k}\right)\right)^2 |N^1(1) = k\right] \\
&\quad = E\left[(V+\tilde{V})^2 (W+\tilde{W})^2 |N^1(1) = k\right] = E\left[\left((VW + V\tilde{W}) + W\tilde{V} + \tilde{V}\tilde{W}\right)^2 |N^1(1) = k\right] \\
&\quad \le 4E\left[(VW)^2 + (V\tilde{W})^2 + (W\tilde{V})^2 + (\tilde{V}\tilde{W})^2 |N^1(1) = k\right], \qquad (65)
\end{aligned}$$

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

It follows from Theorem 1 (i) of [8] 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,

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

which implies that

$$\begin{aligned}
E[N] &= \beta, \quad E\left[N^2\right] = \beta\gamma + \beta^2 + \beta \\
E\left[N^3\right] &= \beta(2\gamma^2 + 3\beta\gamma + 3\gamma + \beta^2 + 3\beta + 1) \quad \text{and} \\
E\left[N^4\right] &= \beta(6\gamma^3 + 11\beta\gamma^2 + 12\gamma^2 + 6\beta^2\gamma + 18\beta\gamma + 7\gamma + \beta^3 + 6\beta^2 + 7\beta + 1) (66)
\end{aligned}$$

By (14.10) of [5], $E\left[(VW)^2 |N^1(1) = k\right] \le 6\ k^2 p_1 p_2$, so that

$$E\left[(VW)^2\right] \le 6E\left[N^2\right] p_1 p_2. \qquad (67)$$

Similarly,

$$E\left[\left(V\tilde{W}\right)^2\right] = E\left[Np_1\left(1-p_1\right)\left(\left(N-\beta\right)p_2\right)^2\right], \quad E\left[\left(\tilde{V}W\right)^2\right] = E\left[Np_2\left(1-p_2\right)\left(\left(N-\beta\right)p_1\right)^2\right],$$

and

$$E[\left(\tilde{V}\tilde{W}\right)^2] = E[\left(\left(N-\beta\right)p_1\right)^2\left(\left(N-\beta\right)p_2\right)^2]. \tag{68}$$

Using Macsyma for algebraic simplification, we find first, by (65),

$$\begin{aligned} &E[\left(A_1\left(t\right)-A_1\left(t_1\right)\right)^2\left(A_1\left(t_2\right)-A_1\left(t\right)\right)^2] \\ &\quad\le \Omega \equiv 4\left(E\left[\left(VW\right)^2\right] + E\left[\left(V\tilde{W}\right)^2\right] + E\left[\left(\tilde{V}W\right)^2\right] + E\left[\left(\tilde{V}\tilde{W}\right)^2\right]\right). \end{aligned} \tag{69}$$

Then, by (66)-(68), we find that

$$\begin{aligned} \Omega &= 4\beta\left(6\gamma^3p_1p_2 + 3\beta\gamma^2p_1p_2 + 8\gamma^2p_1p_2 + 4\beta\gamma p_1p_2 + \gamma p_1p_2 + \beta p_1p_2 - p_1p_2 + 2\gamma^2p_2 + \beta\gamma p_2\right. \\ &\quad \left. +3\gamma p_2 + \beta p_2 + p_2 + 2\gamma^2p_1 + \beta\gamma p_1 + 3\gamma p_1 + \beta p_1 + p_1 + 6\gamma + 6\beta + 6\right)p_1p_2 \\ &\le 4\beta(6\gamma^3 + 3\beta\gamma^2 + 12\gamma^2 + 6\beta\gamma + 13\gamma + 9\beta + 7)p_1p_2 \\ &= 4\beta(6\gamma^3 + 3\beta\gamma^2 + 12\gamma^2 + 6\beta\gamma + 13\gamma + 9\beta + 7)\left(t-t_1\right)\left(t_2-t\right) \\ &\le 4\beta(6\gamma^3 + 3\beta\gamma^2 + 12\gamma^2 + 6\beta\gamma + 13\gamma + 9\beta + 7)\left(t_2-t_1\right)^2. \end{aligned} \tag{70}$$

Thus we conclude that (2.4) 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, Theorem 7.2.1 in [37] shows that $A_n \Rightarrow A$ where $A$ is a zero-mean Gaussian process with the same covariance kernel as $A_1$.

**Remark 7.** (*easier proof of Theorem 3 fails.*) A candidate easier proof of Theorem 3 for sums of Markov processes is provided by Theorem 7.2.2 of [37], but Conditions (2.9) and (2.10) there are not satisfied in our case. The proof of Theorem 3 shows that the expectation in (65) conditional on $N = N^1(1)$ does not have a uniform bound of the form of the right-hand side of (64). The uniform bound in (64) is obtained from the conditional expectation only after accounting for the distribution of $N$. Therefore, the candidate easier proof for sums of Markov process fails in our case.

## 8. Conclusions

In this paper we have helped expose the performance consequence on a single-server queue of a path-dependent arrival processes. Our main results are Theorem 3 and 4, which show that a superposition of $\Psi$-GPP's, which is itself a $\Psi$-GPP, converges to a $\Psi$-GMP as studied in [14]; i.e., the limit process is a Gaussian Markov process with stationary increments. Corollaries 5 and 6 provide explicit performance approximations for the queueing processes.

The $\Psi$-GPP and the limiting $\Psi$-GMP exhibit positive dependence, but the class of $\Psi$-GMP's considered in [14] also include processes exhibiting negative dependence. We close by observing that it is possible to obtain all possible $\Psi$-GMP's considered in [14] by limits like that in Theorem 3. Indeed, all possible limits are obtained by considering linear combinations of uniform empirical processes and superpositions of $\Psi$-GPP's. 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 by [19, 23]. In fact the proof of the explicit form of the distribution of the reflected $\Psi$-GMP in Corollary 5 already drew on the structure of 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 §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 Polya arrival process.

## References

[1] Arthur, W. B. (1988). Self-reinforcing mechanisms in economics. in *The Economy as an Evolving Complex System*, P. W. Anderson, K. Arrow and D. Pines (eds.). Proceedings of the Sante Fe Institute, CRC Press, Boca Raton, FL.

[2] Arthur, W. B. and Ermoliev, Yu. M. and Kanjovski, Yu. M (1987). Path-dependent processes and the emergence of macro-structure, *European Journal of Operational Research*, **30,** 294–303.

[3] Baccelli, F. and Bremaud, P. (1994). *Elements of Queueing Theorey*, Springer, New York.

[4] Bacray, E. and Delattre, S. and Hoffman, M. and Muzy, J. F. (2013). Some limit theorems for Hawkes processes and applications to finanical statistics, *Stochastic Processes and their Applications*, **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. and Badia, F. G. (2019) On a multivariate generalized Polya process without regularity property. *Probability in the Engineering and Informational Sciences*, published on line, 23 pages.

[8] Cha, J. H. (2014) Characterization of the generalized Polya process and its applications. *Advances in Applied Probability* **46** (4) 1148-1171.

[9] Debicki, K. and Kosinski, K. and Mandjes, M. (2012) Gaussian queues in light and heavy traffic. *Queueing Systems* **71** 137–149.

[10] Debicki, K. and Rolski, T. (2002) A note on transient Gaussian fluid models. *Queueing Systems* **41** 321–342.

[11] Doob, J. L. (1953). *Stochastic Processes*, Wiley, New York: Wiley.

[12] Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, Vol. I, 3rd edn. John Wiley, New York.

[13] Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. II, 2nd edn. John Wiley, New York.

[14] Fendick, K. W. (2020) Brownian motion minus the independent increments: representation and queuing application,. *Probability in the Engineering and Informational Sciences*, accepted subject to revision.

[15] 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. *Proceedings of the IEEE,* **71** (1) 171–194.

[16] Glynn, P. W. and Whitt, W. (1991) A new view of the heavy-traffic limit for infinite-server queues. *Advances in Applied Probability* **23** (1) 188–209.

[17] Gregoire, G. (1983) Negative binomial distributions for point processes. *Stochastic Processes Appl.* **16** 179–188.

[18] Hahn, M. G. (1978) Central limit theorems in $D[0,1]$. *Zeitchrift für Wahrscheinlichkeitstheorie verw. Gebiete* **44** 89-101.

[19] Hajek, B. (1994) A queue with periodic arrivals and constant service. in *Probability, Statistics and Optimization: a tribute to Peter Whittle*, F. P. Kelley (ed.), Chichester, Wiley, 147-157.

[20] Harrison, J. M. (1985). *Brownian Motion and Stochastic Flow Systems*, Wiley, New York.

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

[22] Hawkes, A. G. (1971b) Point spectra of some mutually exciting point processes. *J. Roy. Stat. Soc.* **33** (3) 438-443.

[23] Honnappa, H. and Jain, R. and Ward, A. (2015) A queueing model with independent arrials and its fluid and diffusion limits. *Queueing Systems* **80** (1) 71–103.

[24] Iglehart, D. L. and Whitt, W. (1970) Multiple channel queues in heavy traffic, II: sequences, networks, and batches. *Advances in Applied Probability* **2** (2) 355-369–194.

[25] Isaacson, D. and Madsen, R. (1976). *Markov chains: Theory and Applications*, Wiley, New York.

[26] Karatzas, I. and Shreve, S. (2000). *Brownian Motion and Stochastic Calculus*, 2nd edn. Springer, New York.

[27] Konno, T. H (2010) On the exact solution of a generalized Polya process. *Advances in Mathematical Physics* **2010** Article ID 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** (3) 261–290.

[31] Mirtchev, S. T. (2019). Study of preemptive priority single-server queue with peaked arrival flow. *Proc. X National Conference with International Participation "Electronica 2019"*, May 16 - 17, 2019, Sofia, Bulgaria.

[32] Mirtchev, S. T. and Goleva, R. (2013). New constant service time $Polya/D/n$ traffic model with peaked input stream. *Simulation Modelling Practice and Theory* **34** 200-207.

[33] Mirtchev, S. T. and Ganchev, I. (2016). Generalised Pollaczek–Khinchin formula for the $Polya/G/1$ queue. *Electronic Letters* **53** (1) 27-29.

[34] Polya, G. and Eggenberger, F. Uber die Statistik verketteter Vorgange. *Zeitschrift für angewandte Mathematische Mechanik*, **3,** 279–289.

[35] Stidham, S. A last word on $L = \lambda W$. *Operations Research* **22,** 417–421.

[36] Whitt, W. A review of $L = \lambda W$ and extensions. *Queueing Systems* **9,** 235–268.

[37] Whitt, W. (2002). *Stochastic Process Limits*, Springer, New York.

[38] Whitt, W. (2015). Stabilizing performance in a single-server queue with time-varying arrival rate. *Queueing Systems* **81** 341–378.

[39] Whitt, W. and You. W. (2018) Using robust queueing to expose the impact of dependence in single-server queues. *Operations Research* **66** (1) 184–199.

[40] Wolff, R. W. and Yao. Y. (2014) Little's law when the average waiting time is infinite. *Queueing Systems* **76** (1) 267–281.