J. Appl. Prob. **45,** 963–971 (2008)
Printed in England
© *Applied Probability Trust* 2008

HOW NEARLY DO ARRIVING CUSTOMERS SEE TIME-AVERAGE BEHAVIOR?

EROL A. PEKÖZ,* *Boston University*
SHELDON M. ROSS,** *University of Southern California*
SRIDHAR SESHADRI,*** *New York University*

Abstract

Customers arriving at a queue do not usually see its time-average behavior unless arrivals occur according to a Poisson process. In this article we study how nearly customers see time-average behavior. We give total variation error bounds for comparing the distance between the time- and customer-average distributions of a queueing system in terms of properties of the interarrival distribution. Some refinements are given for special cases and numerical computations are used to demonstrate the performance of the inequalities.

Keywords: PASTA; arrivals that see time averages; arrival theorem

2000 Mathematics Subject Classification: Primary 60K25
Secondary 90B22; 60K05; 60G10

1. Introduction

Customers arriving at a queue do not necessarily see its time-average behavior. If arrivals occur in infrequent heavy bursts, for example, most customers can see a long queue even though the system is usually empty. The 'Poisson arrivals see time averages' (PASTA) property of queueing theory, attributed to Wolff (1982), says that the fraction of Poisson arrivals finding a queue in some state equals the fraction of time the queue is in that state.

There is a very large literature on the PASTA property; for some entry points to this literature, see, for example, Melamed and Whitt (1990a), Melamed and Whitt (1990b), Melamed and Yao (1995), Brèmaud *et al.* (1992), Heyman and Stidham (1980), and the references therein. There has also been development of inequalities relating the time-average distribution and the customer-average distribution; see Köning and Schmidt (1980), Niu (1984), Shanthikumar and Zazanis (1999), Peköz and Ross (2008), and the references therein. The abovementioned works have detailed conditions under which customers see either stochastically more or less than the time average. For example, Melamed and Whitt (1990a) showed that the two are related by the so-called 'covariance' formula (see Equation (8) on page 159). This formula does not appear easy to solve and does not lend itself to calculation of the total variation distance. The bounds from our main result, in contrast, are easily calculated using only properties of the interarrival distribution. In other words, we obtain the same upper bound for all queueing systems having the same arrival process.

Received 6 May 2008; revision received 21 October 2008.

* Postal address: Department of Operations and Technology Management, Boston University, 595 Commonwealth Avenue, Boston, MA 02215, USA. Email address: pekoz@bu.edu

** Postal address: Department of Industrial and System Engineering, University of Southern California, Los Angeles, CA 90089, USA.

*** Current address: Department of Information, Risk, and Operations Management, McCombs School of Business, University of Texas at Austin, 1 University Station, B6500, Austin, TX 78712, USA.

https://doi.org/10.1239/jap/1231340227 Published online by Cambridge University Press

In this article we give some results relating the total variation distance between the time- and customer-average distributions. Our purpose is to illustrate how nearly arriving customers see the time-average distribution. The organization of this article is as follows. In Section 2 we give our main result, which is to show that the total variation distance between the time- and customer-average distributions of a queueing system is less than the total variation distance between the interarrival distribution and the corresponding equilibrium distribution. In Section 3 we look at specific queueing models and incorporate further properties of the models to refine the bounds. We also give results for the superposition of renewal processes. In Section 4 we give some numerical results, where we compute the exact total variation distance and compare it with the upper bound.

2. Main result

We consider a function $Q(t)$, $-\infty < t < \infty$, of a given queueing system such that $Q(t)$ is a real-valued, left-continuous stochastic process with right-hand limits under the usual topology. We refer to this as the 'state' of our system and it could, for example, represent a quantity such as the number of customers or the total workload in the system (our results could straightforwardly be extended to more general state spaces by defining an appropriate topology). Left continuity is used so that the stochastic process at any given time tells us what will happen immediately after that time, and so it tells us what an arrival at that time would see. We suppose that customers arrive to the queueing system according to a point process (called the arrival events), where T_n denotes the time of the nth customer arrival, $-\infty < n < \infty$. We label customers according to the common convention where $\cdots T_{-2} < T_{-1} < 0 < T_1 < T_2 \cdots$. The state and arrival processes are assumed to be jointly stationary and ergodic under the probability measure P. In this article we will refer to P_0 as the *Palm transformation* of P with respect to the sequence $\{T_n\}$. Thus, we use the subscripts P_0 and E_0 to respectively denote the probabilities and expectations computed with respect to the Palm transformation of the process, as described in Baccelli and Brèmaud (1987). Intuitively, $\mathrm{P}_0(A)$ can be thought of as the conditional probability of A given that a typical customer's arrival occurs at time 0, or, in other words, according to the customer-average distribution. The interarrival time distribution is denoted generically as X with cumulative distribution function $\mathrm{P}_0(T_1 \le x) = F(x)$ and has finite mean $1/\lambda$.

Let X_{e} be a random variable having the equilibrium distribution of the interarrival time. Then

$$F_{\mathrm{e}}(x) = \lambda \int_0^x \bar{F}(y)\,\mathrm{d}y,$$

where $\bar{F}(x) = 1 - F(x)$. Note that the density function for X_{e} is

$$f_{\mathrm{e}}(x) = \lambda \bar{F}(x).$$

We say that the random variable Z has the time-average steady-state distribution if

$$\mathrm{P}(Z \in A) = \lim_{t\to\infty} \frac{1}{t} \int_0^t \mathbf{1}\{Q(s) \in A\}\,\mathrm{d}s$$

for any (measurable) set A. We say that the random variable W has the customer-average steady-state distribution if

$$\mathrm{P}(W \in A) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{Q(T_i) \in A\}.$$

https://doi.org/10.1239/jap/1231340227 Published online by Cambridge University Press

Next we similarly define a new process $Q_0(t)$ to be the state of the above queueing system at time t given that it is started at time 0 with a typical arrival and according to the Palm measure, so that $\mathrm{P}(Q_0(0) \in A) = \mathrm{P}_0(Q(0) \in A)$, but it continues without allowing any further arrivals to enter the system.

We also need the following lack of anticipation assumption similar to the one given in Shanthikumar and Zazanis (1999); ours is tailored to our particular application. This assumption ensures that the queue cannot anticipate future arrivals and rules out, for example, queues where the servers speed up shortly before new arrivals occur.

Assumption 1. *For any $t > 0$, we have*

$$\mathrm{P}_0(Q(t) \in A \mid T_1 \geq t) = \mathrm{P}_0(Q(t) \in A \mid T_1 = t) = \mathrm{P}(Q_0(t) \in A).$$

Remark 1. Intuitively, Assumption 1 states that starting when a typical customer arrives, say at time 0, T_1, the time until the next arrival, and the state of the system at time s are conditionally independent given that $T_1 > s$. This assumption holds, for example, for a queue with a renewal arrival process where future interarrivals are independent of the system state as seen by the current arrival. Also, it holds for any arrival process in which all customers presently in the system have to depart when a new arrival occurs. It does not generally hold, however, for systems with Markov modulated Poisson arrivals.

Remark 2. In contrast to much of the PASTA literature, we assume stationarity and ergodicity. Assuming Poisson arrivals and a lack of anticipation assumption or, more generally, the assumption of lack of bias (Melamed and Whitt (1990a), (1990b)), in most of the PASTA literature it is shown that if the time average over a finite horizon converges to either a limit or a random variable as the time horizon increases to ∞ then the customer average also converges to the same average or random variable. An additional problem arises when analyzing non-Poisson arrivals because there is statistical dependence between the past and future of the point process. To circumvent this problem, Niu (1984) defined the strict lack of anticipation (SLA) assumption. Assumption 1 is a weaker version of the SLA assumption. Its main purpose is to allow us to work with Palm expectations of the functions of the state of the system and, thus, allows us to work with general renewal arrival processes. It is also different from the lack of anticipation assumption, as formulated by Wolff (1982), where the entire future of the arrival process starting from time t is independent of the history of the system until time t.

We define the total variation distance between two random variables U and V as

$$d_{\mathrm{TV}}(U, V) \equiv \sup_A |\mathrm{P}(U \in A) - \mathrm{P}(V \in A)|.$$

We also need the following lemma for maximal couplings given in Ross and Peköz (2007).

Lemma 1. (Ross and Peköz (2007, Proposition 2.5).) *Given the random variables (U, V) having respective piecewise-continuous density functions $u(x)$ and $v(x)$, there exists a coupling $(\hat{U}, \hat{V})$ of (U, V), called the maximal coupling, such that*

$$d_{\mathrm{TV}}(U, V) = \mathrm{P}(\hat{U} \neq \hat{V}) = 1 - \int_{-\infty}^{\infty} \min(u(x), v(x))\, \mathrm{d}x.$$

To present our main result, recall that $Q_0(0)$ and $Q(0)$ respectively denote the customer-average distribution and the time-average distribution.

https://doi.org/10.1239/jap/1231340227 Published online by Cambridge University Press

Theorem 1. *With the above definitions,*

$$d_{\mathrm{TV}}(Q_0(0), Q(0)) \le d_{\mathrm{TV}}(X, X_{\mathrm{e}}) = 1 - \int_0^\infty \min(f(x), \lambda \bar{F}(x))\,\mathrm{d}x.$$

Proof. Let X and X_{e} be a maximal coupling of random variables having respective distributions F and F_{e}, independent of the queueing system. Given a set A, the Palm inversion formula (see Brèmaud (1993)) gives

$$\mathrm{P}(Q(0) \in A) = \lambda\,\mathrm{E}_0\biggl[\int_0^{T_1} \mathbf{1}\{Q(s) \in A\}\,\mathrm{d}s\biggr].$$

Using this, we have

$$\begin{aligned}\mathrm{P}(Q(0) \in A) &= \lambda \int_0^\infty \mathrm{E}_0[\mathbf{1}\{Q(s) \in A\}\,\mathbf{1}\{T_1 \ge s\}]\,\mathrm{d}s\\ &= \lambda \int_0^\infty \mathrm{P}_0(Q(s) \in A \mid T_1 \ge s)\bar{F}(s)\,\mathrm{d}s\\ &= \int_0^\infty \mathrm{P}(Q_0(s) \in A) f_{\mathrm{e}}(s)\,\mathrm{d}s\\ &= \mathrm{P}(Q_0(X_{\mathrm{e}}) \in A),\end{aligned}$$

where the first line uses Fubini's theorem and the third line uses the lack of anticipation assumption.

By stationarity of the process seen by the arrivals we have

$$\begin{aligned}\mathrm{P}_0(Q(0) \in A) &= \mathrm{P}_0(Q(T_1) \in A)\\ &= \int_0^\infty \mathrm{P}_0(Q(s) \in A \mid T_1 = s) f(s)\,\mathrm{d}s\\ &= \int_0^\infty \mathrm{P}(Q_0(s) \in A) f(s)\,\mathrm{d}s\\ &= \mathrm{P}(Q_0(X) \in A),\end{aligned}$$

where we have used the lack of anticipation assumption in the third line. We then have

$$\begin{aligned}|\,\mathrm{P}(Q(0) \in A) - \mathrm{P}(Q_0(0) \in A)| &= |\,\mathrm{P}(Q_0(X_{\mathrm{e}}) \in A) - \mathrm{P}(Q_0(X) \in A)|\\ &= |\,\mathrm{E}[\mathbf{1}\{Q_0(X_{\mathrm{e}}) \in A\} - \mathbf{1}\{Q_0(X) \in A\}]|\\ &\le \mathrm{P}(X \ne X_{\mathrm{e}})\\ &= 1 - \int_0^\infty \min(f(x), \lambda \bar{F}(x))\,\mathrm{d}x,\end{aligned}$$

where we have used Lemma 1 in the last line.

3. Some special cases and refinements

In this section we give some improved bounds for special queueing systems. We restrict ourselves to renewal processes for customer arrivals.

https://doi.org/10.1239/jap/1231340227 Published online by Cambridge University Press

Definition 1. A random variable $X \geq 0$ is said to be new worse than used (NWU) or new better than used (NBU) if

$$\mathrm{P}(X > t + s \mid X > t) \geq \mathrm{P}(X > s)$$

or, respectively,

$$\mathrm{P}(X > t + s \mid X > t) \leq \mathrm{P}(X > s)$$

for all $t > 0$ and $s > 0$. A random variable $X \geq 0$ is said to be new worse than used in expectation (NWUE) or new better than used in expectation (NBUE) if

$$\mathrm{E}[X - s \mid X > s] \geq \mathrm{E}[X]$$

or, respectively,

$$\mathrm{E}[X - s \mid X > s] \leq \mathrm{E}[X]$$

for all $s > 0$.

Let $f(x) = F'(x)$ be the density function of the interarrival distribution F, which we denote generically as the random variable X, and let X_{e} be the random variable with the corresponding equilibrium distribution. Let

$$p = \int_0^\infty \min(f(x), \lambda \bar{F}(x))\,\mathrm{d}x.$$

Also, let C and D be independent random variables with respective densities

$$c(x) = \frac{f(x) - \min(f(x), \lambda \bar{F}(x))}{1 - p}$$

and

$$d(x) = \frac{\lambda \bar{F}(x) - \min(f(x), \lambda \bar{F}(x))}{1 - p}.$$

The following lemma follows from the proof of Proposition 2.5 of Ross and Peköz (2007).

Lemma 2. *There is a maximal coupling* $(\hat{X}, \hat{X}_{\mathrm{e}})$ *of* (X, X_{e}) *such that*

$$p = \mathrm{P}(\hat{X} = \hat{X}_{\mathrm{e}})$$

and

$$\mathrm{P}(\hat{X} \leq x, \hat{X}_{\mathrm{e}} \leq y \mid \hat{X} \neq \hat{X}_{\mathrm{e}}) = \mathrm{P}(C \leq x)\,\mathrm{P}(D \leq y).$$

In the following propositions, suppose that the state of the queueing system $Q(t)$ measures some quantity that can only change at the times when there is an arrival or service completion, such as the number of customers in the system or the number of busy servers. We also suppose that Assumption 1 holds. Let $Q(t)$ and $Q_0(t)$ be as defined in Section 2 for the corresponding queueing systems.

Proposition 1. *Consider a* k*-server queueing system where service times at server* i *have distribution* G_i *and are mutually independent. Let the internal movements of customers in the system occur according to some arbitrary rule. Suppose that customers arrive according to a* *renewal process having interarrival cumulative distribution function F. If all the G_i are all NWU then, with $\mathfrak{C}$ being the class of all couplings of X and X_{e} and $\bar{G}_i(x) = 1 - G_i(x)$,*

$$
\begin{aligned}
d_{\mathrm{TV}}(Q_0(0), Q(0)) &\le \inf_{(\hat{X}, \hat{X}_{\mathrm{e}}) \in \mathfrak{C}} \biggl(1 - \mathrm{E}\biggl[\prod_{i=1}^{k} \bar{G}_i(|\hat{X}_{\mathrm{e}} - \hat{X}|)\biggr]\biggr) \\
&\le (1-p)\biggl(1 - \mathrm{E}\biggl[\prod_{i=1}^{k} \bar{G}_i(|C - D|)\biggr]\biggr).
\end{aligned}
$$

Proof. With Y_i, $i = 1, 2, \dots, k$, denoting independent generic service times and $(\hat{X}_{\mathrm{e}}, \hat{X})$ denoting any coupling of (X_{e}, X), we have

$$
\begin{aligned}
|\,\mathrm{P}(Q(0) \in A) - \mathrm{P}(Q_0(0) \in A)| &= |\,\mathrm{P}(Q_0(X_{\mathrm{e}}) \in A) - \mathrm{P}(Q_0(X) \in A)| \\
&\le \mathrm{P}(\text{service completion between } X \text{ and } X_{\mathrm{e}}) \\
&\le \mathrm{P}\Bigl(|\hat{X} - \hat{X}_{\mathrm{e}}| \ge \min_{1 \le i \le k} Y_i\Bigr) \\
&= 1 - \mathrm{E}[(\bar{G}(|\hat{X}_{\mathrm{e}} - \hat{X}|))^k],
\end{aligned}
$$

where the last line follows by conditioning on $|\hat{X}_{\mathrm{e}} - \hat{X}|$ and using $\mathrm{P}(\min_{1 \le i \le k} Y_i \le x) = 1 - \prod_{i=1}^{k} \bar{G}_i(x)$. The final inequality follows by using the maximal coupling of Lemma 2.

Remark. The quantity $(1-p)(1 - \mathrm{E}[\prod_{i=1}^{k} \bar{G}_i(|C - D|)])$ can be numerically evaluated in specific applications.

Next, for a cumulative distribution function F, define $F^{-1}(x) \equiv \inf\{t \colon F(t) \ge x\}$, and let U denote a uniform (0,1) random variable.

Proposition 2. *Assume that the conditions of Proposition 1 hold, and suppose that $\bar{G}_i(x) = \exp[-\mu_i x]$. Let $\mu = \sum_i \mu_i$. If customers arrive according to a renewal process whose interarrival distribution X is either NWUE or NBUE then*

$$
d_{\mathrm{TV}}(Q_0(0), Q(0)) \le 1 - \exp\biggl[-\mu\biggl(\biggl|\frac{\mathrm{E}[X^2]}{2\,\mathrm{E}[X]} - \mathrm{E}[X]\biggr|\biggr)\biggr].
$$

Proof. First suppose that X is NWUE, and define the coupling

$$
(X, X_{\mathrm{e}}) = (F^{-1}(U), F_{\mathrm{e}}^{-1}(U)).
$$

It can be shown that $X_{\mathrm{e}} \ge X$ (if X is NBUE, we have $X_{\mathrm{e}} \le X$). Then

$$
\begin{aligned}
|\,\mathrm{P}(Q_0(X) \in A) - \mathrm{P}(Q_0(X_{\mathrm{e}}) \in A)| &= |\,\mathrm{E}[\mathbf{1}\{Q_0(X) \in A\} - \mathbf{1}\{Q_0(X_{\mathrm{e}}) \in A\}]| \\
&\le \mathrm{E}[|\,\mathbf{1}\{Q_0(X) \in A\} - \mathbf{1}\{Q_0(X_{\mathrm{e}}) \in A\}|] \\
&\le \mathrm{P}(\text{at least one service completion in } (X, X_{\mathrm{e}})) \\
&\le 1 - \mathrm{E}[\exp[-\mu(X_{\mathrm{e}} - X)]] \\
&\le 1 - \exp[-\mu\,\mathrm{E}[X_{\mathrm{e}} - X]] \quad \text{(by Jensen's inequality)} \\
&= 1 - \exp\biggl[-\mu\biggl(\biggl|\frac{\mathrm{E}[X^2]}{2\,\mathrm{E}[X]} - \mathrm{E}[X]\biggr|\biggr)\biggr].
\end{aligned}
$$

When X is NBUE, the result follows similarly.

https://doi.org/10.1239/jap/1231340227 Published online by Cambridge University Press

Proposition 3. *Consider the $GI/GI/k$ queue, where service times are NWU. Let F denote the interarrival time cumulative distribution function, and let G denote the service time cumulative distribution function. Suppose that the state measures some quantity, such as the number of customers in the system or the number of busy servers, which can only change at the times when there is an arrival or service completion. Then, using the above definitions,*

(a) $d_{\mathrm{TV}}(Q_0(0), Q(0)) \le 1 - \bar{G}(b)^k$, *where* $b = \sup_x |F^{-1}(x) - F_{\mathrm{e}}^{-1}(x)|$*;*

(b) *for any* $d > 0$*, we have*

$$d_{\mathrm{TV}}(Q_0(0), Q(0)) \le 1 - \bar{G}(d)^k \sum_{i=0}^{\infty} \min(F(di+d) - F(di), F_{\mathrm{e}}(di+d) - F_{\mathrm{e}}(di)).$$

Proof. For a uniform (0,1) random variable U, part (a) follows using Proposition 2 along with the coupling $(\hat{X}, \hat{X}_{\mathrm{e}}) = (F^{-1}(U), F_{\mathrm{e}}^{-1}(U))$ of (X, X_{e}).

To prove part (b), first let $R(x) = d\lfloor x/d \rfloor$ be the value of x rounded down to the nearest multiple of d. Then first create a maximal coupling (U, V) of $(R(X_{\mathrm{e}}), R(X))$, and then let $(\hat{X}_{\mathrm{e}}, \hat{X})$ be conditionally independent of all else and distributed according to (X_{e}, X), conditional on $R(X_{\mathrm{e}}) = U$ and $R(X) = V$. Then we have

$$\begin{aligned} 1 - \mathrm{E}[(\bar{G}(|\hat{X}_{\mathrm{e}} - \hat{X}|))^k] &\le 1 - \bar{G}(d)^k\, \mathrm{P}(|\hat{X}_{\mathrm{e}} - \hat{X}| \le d) \\ &\le 1 - \bar{G}(d)^k\, \mathrm{P}(U = V) \\ &= 1 - \bar{G}(d)^k \sum_{i=0}^{\infty} \min(\mathrm{P}(U = di), \mathrm{P}(V = di)) \\ &= 1 - \bar{G}(d)^k \sum_{i=0}^{\infty} \min(F(di+d) - F(di), F_{\mathrm{e}}(di+d) - F_{\mathrm{e}}(di)), \end{aligned}$$

where we have used Proposition 2.6 of Ross and Peköz (2007) for maximally coupled discrete random variables in the second to last line. The result then follows using Proposition 2.

We now consider an application to the setting where the arrival process consists of the superposition of a number of different but independent renewal processes. Assume that there are n independent renewal arrival processes to a queueing network with interarrival times denoted generically by the random variables X^i, $i = 1, 2, \dots, n$, having cumulative distribution functions $F^i(x)$ and density functions $f^i(x)$. Let X_{e}^i, $i = 1, 2, \dots, n$, denote the corresponding equilibrium distributions. Letting $\lambda_i = 1/\mathrm{E}[X^i]$, we have the following result.

Proposition 4. *With the above definitions,*

$$\begin{aligned} d_{\mathrm{TV}}(Q_0(0), Q(0)) &\le \sum_{i=1}^{n} \frac{\lambda_i}{\lambda} d_{\mathrm{TV}}(X^i, X_{\mathrm{e}}^i) \\ &= \sum_{i=1}^{n} \frac{\lambda_i}{\lambda} \biggl(1 - \int_0^{\infty} \min(f^i(x), \lambda \bar{F}^i(x))\, \mathrm{d}x\biggr). \end{aligned}$$

Proof. Let $Q_i(0)$, $i > 0$, be the customer-average distribution from the point of view of a typical customer in the ith arrival stream—or, in other words, it is distributed according to the Palm distribution with respect to the ith arrival stream. Imagine that the ith renewal process

https://doi.org/10.1239/jap/1231340227 Published online by Cambridge University Press

is the 'real' arrival process and that all the others are considered as part of the 'system'. Then Theorem 1 gives

$$d_{\mathrm{TV}}(Q_i(0), Q(0)) \le d_{\mathrm{TV}}(X^i, X^i_{\mathrm{e}}) = 1 - \int_0^\infty \min(f^i(x), \lambda \bar{F}^i(x))\,\mathrm{d}x.$$

The proposition follows by conditioning on the type of arrival that occurs at time 0.

The next proposition shows how close the customer-average distribution is for customers in one arrival stream compared with customers in another arrival stream.

Proposition 5. *In the setting of Proposition 4 we have*

$$d_{\mathrm{TV}}(Q_i(0), Q_j(0)) \le d_{\mathrm{TV}}(X^i, X^i_{\mathrm{e}}) + d_{\mathrm{TV}}(X^j, X^j_{\mathrm{e}}).$$

Proof. The result follows immediately upon using the triangle inequality and Proposition 4 twice.

4. Some numerical examples

Consider a GI/M/1 queue, where service times are exponential(μ) and the interarrival distribution has moment generating function $M(t) = \mathrm{E}[\mathrm{e}^{tX}]$. It can be shown (see Gross and Harris (1998, Section 5.3.1)) that the customer-average steady-state distribution for the number of customers in the system is

$$\pi_j = (1-\alpha)\alpha^j, \qquad j \ge 0,$$

where α is the smallest positive root of

$$\alpha = M(\mu(\alpha - 1)).$$

It can also be shown that the time-average steady-state distribution for the number of customers in the system is

$$\mathrm{P}_j = \begin{cases} \rho(1-\alpha)\alpha^{j-1}, & j \ge 1, \\ 1-\rho, & j = 0, \end{cases}$$

where $\rho = (\mu\,\mathrm{E}[X])^{-1}$ is the traffic intensity. This means that

$$d_{\mathrm{TV}}(Q(0), Q_0(0)) = 1 - \sum_j \min(\mathrm{P}_j, \pi_j) = |\rho - \alpha|,$$

and recall also from above that

$$d_{\mathrm{TV}}(X, X_{\mathrm{e}}) = 1 - \int_0^\infty \min(f(x), f_{\mathrm{e}}(x))\,\mathrm{d}x,$$

where $f(x)$ and $f_{\mathrm{e}}(x)$ respectively denote the density function for the interarrival distribution and equilibrium distribution.

We now look at a number of numerical examples, where the interarrival distribution is gamma(n, λ). The results are shown in Table 1, where it can be seen that the upper bound we give above is usually less than twice the actual total variation distance.

https://doi.org/10.1239/jap/1231340227 Published online by Cambridge University Press

TABLE 1: Some numerical results comparing the actual distance in the upper bound between the customer- and time-average distributions for the GI/M/1 queue with gamma(n, λ) interarrivals.

n	λ	ρ	α	Actual distance $d_{\mathrm{TV}}(Q(0), Q_0(0))$	Upper bound $d_{\mathrm{TV}}(X, X_{\mathrm{e}})$
2	1.00	0.50	0.38	0.12	0.18
3	1.50	0.50	0.33	0.17	0.28
4	2.00	0.50	0.30	0.20	0.34
2	0.50	0.25	0.13	0.12	0.18
3	0.75	0.25	0.09	0.16	0.28
4	1.00	0.25	0.07	0.18	0.34
2	1.50	0.75	0.68	0.07	0.18
3	2.25	0.75	0.64	0.11	0.28
4	3.00	0.75	0.62	0.13	0.34

References

BACCELLI, F. AND BRÈMAUD, P. (1987). *Palm Probabilities and Stationary Queues* (Lecture Notes Statist. **41**). Springer, Berlin.

BRÈMAUD, P. (1993). A Swiss Army formula of Palm calculus. *J. Appl. Prob.* **30,** 40–51.

BRÈMAUD, P., KANNURPATTI, R. AND MAZUMDAR, R. (1992). Event and time averages: a review. *Adv. Appl. Prob.* **24,** 377–411.

HEYMAN, D. P. AND STIDHAM, S. JR. (1980). The relation between customer and time averages in queues. *Operat. Res.* **28,** 983–994.

GROSS, D. AND HARRIS, C. M. (1998). *Fundamentals of Queueing Theory*, 3rd edn. John Wiley, New York.

KÖNING, D. AND SCHMIDT, V. (1980). Stochastic inequalities between customer-stationary and time-stationary characteristics of queueing systems with point processes. *J. Appl. Prob.* **17,** 768–777.

KÖNING, D. AND SCHMIDT, V. (1981). Relationships between time- and customer-stationary characteristics of service systems. In *Point Processes and Queueing Problems*, eds P. Bartfai and J. Tomko, North-Holland, Amsterdam, pp. 181–225.

MELAMED, B. AND WHITT, W. (1990a). On arrivals that see time averages. *Operat. Res.* **38,** 156–172.

MELAMED, B. AND WHITT, W. (1990b). On arrivals that see time averages: a martingale approach. *J. Appl. Prob.* **27,** 376–384.

MELAMED, B. AND YAO, D. D. (1995). The ASTA property. In *Advances in Queueing*, ed. J. H. Dshalalow, CRC, Boca Raton, FL, pp. 195–224.

NIU, S. (1984). Inequalities between arrival averages and time averages in stochastic processes arising from queueing theory. *Operat. Res.* **32,** 785–795.

PEKÖZ, E. A. AND SHELDON, R. (2008). Relating time and customer averages for queues using 'forward' coupling from the past. *J. Appl. Prob.* **45,** 568–574.

ROSS, S. AND PEKÖZ, E. A. (2007). *A Second Course in Probability*. ProbabilityBookstore.com, Boston, MA.

SHANTHIKUMAR, G. AND ZAZANIS, M. (1999). Inequalities between event and time averages. *Prob. Eng. Inf. Sci.* **13,** 293–308.

WOLFF, R. (1982). Poisson arrivals see time averages. *Operat. Res.* **30,** 223–231.

https://doi.org/10.1239/jap/1231340227 Published online by Cambridge University Press