Journal of Simulation (2011) 5, 3–8

www.palgrave-journals.com/jos/

Modelling and simulating non-stationary arrival processes to facilitate analysis

BL Nelson[1]* and I Gerhardt[2]

[1]*Northwestern University, Evanston, USA;* [2]*Manhattan College, Riverdale, USA*

This paper introduces a method to model and simulate non-stationary, non-renewal arrival processes that depends only on the analyst setting intuitive and easily controllable parameters. Thus, it is suitable for assessing the impact of non-stationary, non-exponential, and non-independent arrivals on simulated performance when they are suspected. A specific implementation of the method is also described and provided for download.

Journal of Simulation (2011) **5,** 3–8. doi:10.1057/jos.2010.21; published online 15 October 2010

Keywords: simulation; statistics; stochastic processes

1. Introduction

Arrival processes are one of the basic drivers of many stochastic simulation models, including, but not limited to, queueing and supply chain simulations. The stationary Poisson arrival process—implying interarrival times that are independent and identically distributed (iid) and exponentially distributed—is well known, and often justified because it represents 'arrivals from a large customer population making independent decisions about when to arrive'. However, interarrival times are frequently more variable (eg, telecommunications) or more regular (eg, manufacturing orders) than Poisson. To handle this, general stationary renewal arrival processes—with general iid interarrival times—are a feature of every commercial simulation product.

Arrival processes with a time-varying arrival rate representing, for instance, peak and off-peak load are also prevalent in practice. As a result, some of the same software products include the capability to generate arrivals from a non-stationary Poisson process (NSPP). The problem of fitting renewal processes or NSPPs to data has been well studied and there are practically useful tools available (eg, Leemis, 2006; Law, 2007 and references therein).

Of course, stationary renewal processes and NSPPs do not address all of the possible departures from 'Poissonness', which lead Gerhardt and Nelson (2009) to consider non-stationary, non-Poisson (NSNP) arrival processes; NSNP processes are generalizations of stationary renewal processes that allow a time-varying arrival rate. Their work provides methods for fitting and simulating NSNP processes.

*Correspondence: *BL Nelson, Department of Industrial Engineering* and Management Science, Northwestern University, Evanston, IL 60208-3119, USA.

The purpose of this paper is two-fold: From a basic theory perspective, we extend one of Gerhardt and Nelson's results to facilitate generation of non-stationary, non-renewal (NSNR) arrivals, which, in a sense, addresses the final remaining departure from Poisson arrival characteristics (dependent interarrival times) and includes NSPP and NSNP processes as special cases. However, rather than focusing on fitting NSNR processes to data, as Gerhardt and Nelson (2009) do, we provide a specific method designed to allow a user to easily and intuitively define NSNR processes without data. This facilitates assessing the impact of non-stationary, non-exponential, and dependent arrival processes on simulation results when no or only partial information on the arrival processes is available.

We believe that this situation is very common in practice: The modeller is aware that the arrivals are not well represented as Poisson, but has neither sufficient data nor enough information to fully specify the alternative. Therefore, the goal for the modeller—if it is easy enough to do—is to see how much these non-Poisson features matter. A central premise of this work is that modellers will analyse what they can readily model. Thus, it is more important to be able to incorporate non-Poisson features than it is to represent them perfectly.

We also argue that it *is* important to model deviations from stationary, Poisson arrivals. It is well known in queueing models of service systems that replacing a time-varying arrival rate by, say, a constant arrival rate set to the maximum or average value can lead to systems being badly under-staffed or over-staffed (eg, Whitt, 2007). Additionally, a number of highly accurate approximations for stationary non-Markovian queues show that congestion measures are increasing functions of arrival process variability (eg, Whitt, 1981). Less well studied is the impact of dependent interarrival times, although some queueing approximations attempt to represent the impact of dependence through an increase in variability. We will show by example that variability and dependence have distinct effects and both need to be modelled to accurately estimate queueing system performance.

The paper is organized as follows: In the following section we present our method for representing and simulating NSNR arrival processes and prove its basic properties. We then describe a specific implementation of this method and introduce a tool (which is available for download) for generating NSNR arrivals. We use a queueing example to illustrate the dangers of blindly using Poisson arrivals when they are not appropriate. Conclusions are offered at the end.

2. Theory

Our goal is to define and simulate a sequence of interarrival times $\{W_n,\ n \geqslant 1\}$ such that the arrival counting process $I(t) = \max\{n \geqslant 0\colon V_n \leqslant t\}$ (where $V_n = \sum_{i=1}^{n} W_i$) is non-stationary and non-renewal in easily controllable and understandable ways. In particular, the situation we address is when there is a desired time-varying arrival rate and a possibility of dependence between arrivals. We will define, by construction, a stochastic process that exactly matches the desired arrival rate, and gives the user some control of the marginal interarrival-time variance and autocorrelation. We do not assume, or even expect, that the desired arrival process is an instance of our constructed family; rather we seek to match or approximate some important characteristics of their process.

We begin with a set of stationary non-negative interarrival times $\{X_n,\ n \geqslant 1\}$, and let S_n denote the time of the nth arrival; that is, $S_0 = 0$ and $S_n = \sum_{i=1}^{n} X_i$, for $n = 1, 2, \ldots$. Let $N(t)$ denote the number of arrivals that have occurred on or before time t; that is, $N(t) = \max\{n \geqslant 0\colon S_n \leqslant t\}$, for $t \geqslant 0$. We assume that $N(t)$ is initialized in equilibrium, so that, in particular, $\mathrm{E}\{N(t)\} = rt$, for all $t \geqslant 0$, for some fixed arrival rate $r > 0$, and $X_2, X_3, \ldots$ are identically distributed (while X_1 has the associated equilibrium distribution).

The index of dispersion for counts (eg, Sriram and Whitt, 1986) for this process is

$$\mathrm{IDC} = \lim_{t\to\infty} \frac{\mathrm{Var}\{N(t)\}}{\mathrm{E}\{N(t)\}} \tag{1}$$

which we assume exists (more discussion of this assumption follows). For a Poisson process IDC = 1; for an equilibrium renewal process $\mathrm{IDC} = \mathrm{cv}^2$, the squared coefficient of variation of X_2. Notice that (1) implies that for large t, $\mathrm{Var}\{N(t)\} \approx \mathrm{IDC} \cdot \mathrm{E}\{N(t)\}$. From here on we will assume $r = 1$.

The IDC is not an intuitively understandable measure of variability and dependence. However, for many stationary arrival processes it is equal to the index of dispersion for intervals (IDI, Gusella, 1991)

$$\mathrm{IDC} = \mathrm{IDI} \equiv \lim_{n\to\infty} \frac{\mathrm{Var}\{S_n\}}{n\mathrm{E}^2\{X_2\}} = \mathrm{cv}^2\left(1 + 2\sum_{j=1}^{\infty} \rho_j\right)$$

where ρ_j is the lag-j autocorrelation of the stationary interarrival times $X_2, X_3, \ldots$. Therefore, IDC captures both the variability (via cv^2) and dependence (via $1 + 2\sum_{j=1}^{\infty}\rho_j$) in a stationary arrival process.

For IDI = IDC, it is clear that the autocorrelation structure of the interarrival times must be summable, ruling out certain types of long-range dependence (eg, see Leland *et al*, 1994 and references therein). More precisely, Theorem 7.3.1 of Whitt (2002) implies that the IDI and IDC will exist and be equal if the arrival times S_n of the stationary arrival-counting process $N(t)$ satisfy a Central Limit Theorem of the form

$$\frac{1}{\sqrt{n}}(S_n - n\mu) \xrightarrow{\mathcal{D}} N(0,\ \tau^2).$$

Now suppose that $r(t)$, $t \geqslant 0$, is the desired, integrable non-negative arrival rate for $I(t)$, and let $R(t) = \int_0^t r(s)\mathrm{d}s$. Therefore, the 'arrival rate' $r(t)$ is the instantaneous rate of change of the number of arrivals of non-stationary arrival process $I(t)$ at time t. For $s \in \Re^+$, define $R^{-1}(s) \equiv \inf\{t\colon R(t) \geqslant s\}$. Then we have the following algorithm for generating NSNR processes.

Algorithm 1
The Inversion Method for NSNR Processes

1. Set $V_0 = 0$, index counter $n = 1$. Generate S_1. Set $V_1 = R^{-1}(S_1)$.
2. Return interarrival time $W_n = V_n - V_{n-1}$.
3. Set $n = n + 1$. Generate X_n. Set $S_n = S_{n-1} + X_n$ and $V_n = R^{-1}(S_n)$.
4. Go to Step 2.

This algorithm generalizes Algorithm 2.1 of Gerhardt and Nelson (2009) to stationary non-renewal base processes.

Figure 1 illustrates the inversion method when $r(t) = 2t$ customers/time, so that $R(t) = t^2$; in words, the arrival rate is linearly increasing over time. The circles on the vertical axis are arrival times in the rate-1 base process $N(t)$, while the arrows on the horizontal axis are the arrival times in the non-stationary arrival process $I(t)$.

We have the following properties of $I(t)$:

Theorem 1
$\mathrm{E}\{I(t)\} = R(t)$, *for all* $t \geqslant 0$, *and* $\mathrm{Var}\{I(t)\} \approx \mathrm{IDC} \cdot R(t)$, *for large* t.

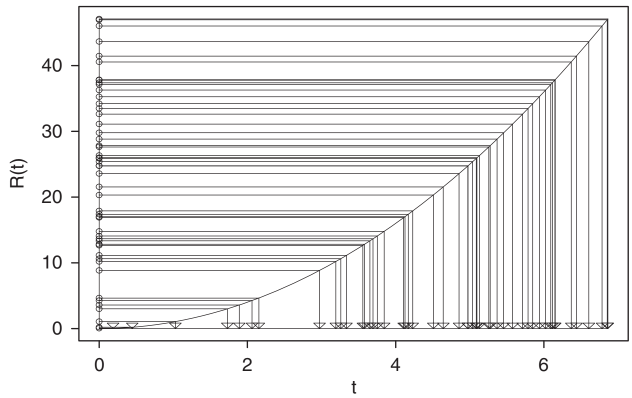

Figure 1 Illustration of the inversion method when $r(t) = 2t$.

Proof

Since $N(t)$ is an equilibrium arrival process with $r=1$, we have $\mathrm{E}\{N(t)\}=t$, for all $t \geqslant 0$, while $\mathrm{Var}\{N(t)\} \approx \mathrm{IDC} \cdot t$, for large t. Thus,

$$\begin{aligned}\mathrm{E}\{I(t)\} &= \mathrm{E}\{\mathrm{E}[I(t)|N(R(t))]\} \\ &= \mathrm{E}\{N(R(t))\} \\ &= R(t)\end{aligned}$$

for all $t \geqslant 0$, while

$$\begin{aligned}\mathrm{Var}\{I(t)\} &= \mathrm{E}\{\mathrm{Var}[I(t)|N(R(t))]\} \\ &\quad + \mathrm{Var}\{\mathrm{E}[I(t)|N(R(t))]\} \\ &= 0 + \mathrm{Var}\{N(R(t))\} \\ &\approx \mathrm{IDC} \cdot R(t)\end{aligned}$$

for large t. □

Thus, $I(t)$ has the desired arrival rate, while preserving the IDC of the stationary base arrival process $N(t)$ from which it was derived. When $N(t)$ is a rate-1 Poisson process, this is the well-known inversion method for generating an NSPP (see, for instance, Çinlar, 1975). Gerhardt and Nelson (2009) extended this method (along with the so-called 'thinning' method) to non-stationary, non-Poisson processes (but still a renewal base process).

In summary, the inversion method attains the desired arrival rate while transferring the IDC of the base process to the NSNR arrival process. In the next section we describe a particular implementation of this result that facilitates analysis.

3. Modelling arrival processes for analysis

The inversion method provides a basis for constructing NSNR arrival processes with control over the arrival rate $r(t)$, marginal variability of the interarrival times cv^2, and dependence among the interarrival times $1+2\sum_{j=1}^{\infty}\rho_j$. In this section we describe a specific implementation that is highly suitable for analysis.

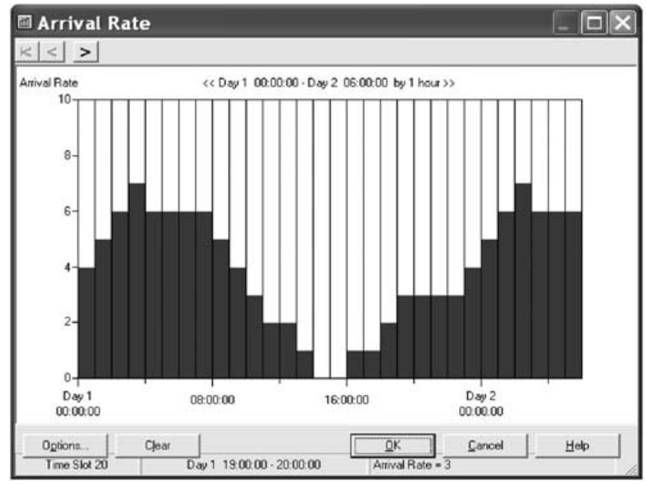

Figure 2 Graphical interface for specifying a piecewise constant arrival rate in Arena (Rockwell Software, Inc.).

4. Arrival rate

The desired arrival rate $r(t)$ should be specified in an intuitive manner that also facilitates inversion of $R(t)$. A piecewise-constant arrival rate fills this need, since $R(t)$ is then piecewise-linear and therefore easily inverted. It is also natural for the modeller to think in terms of the hourly, daily, weekly, etc arrival rate, and if data on arrivals are available, a piecewise-constant rate function is easily estimated (eg, Law, 2007). Figure 2 shows the point-and-click graphical interface used in the commercial simulation software Arena to specify a piecewise-constant arrival rate function hour-by-hour.

5. Base process

For the base arrival process $N(t)$, we suggest the Markov-MECO process of Johnson (1998). The Markov-MECO is a particular case of a Markovian arrival process (MAP); MAPs represent interarrival times as the time to absorption of a continuous-time Markov chain (CTMC) where the initial state of the next interarrival time depends upon which absorbing state the previous interarrival time entered. The Markov-MECO is based on the MECO (Mixture of Erlangs of Common Order) renewal process that can capture any feasible first three moments (equivalently mean, variance, and skewness) of the interrenewal time (Johnson and Taaffe, 1989). The Markov-MECO extends the MECO to non-renewal arrivals by providing a way to control the dependence between interarrival times (described more fully below).

Figure 3 shows one representation of how a Markov-MECO works. The current interarrival time has either an Erlang (k, λ_1) distribution or an Erlang(k, λ_2) distribution (the 'common order' is k); which Erlang distribution provides the next interarrival time is governed by a discrete-time Markov chain with transition probabilities p_{ij}.

As discussed in Gerhardt and Nelson (2009), a key benefit of using a MAP base process is that it is easy to initialize in equilibrium, requiring only that the distribution of the current state of the CTMC in equilibrium be computed; given the current state, the remaining time in that state is always exponentially distributed.

Since the arrival rate for the base Markov-MECO must be 1, this leaves three additional parameters for the user: cv(where $\text{cv}=\sqrt{\text{cv}^2}$), the third moment or skewness of the interarrival-time distribution, and some measure of dependence between interarrival times. Skewness is not a parameter that is easily selected by intuition, so we do not ask the user to provide it and instead use an implied third moment obtained in the following way: We select a Markovian distribution that is fully specified by knowing only its mean and cv, choose its parameters to match our desired mean of 1 and cv, and then use its third moment as the third moment for our Markov-MECO. Specifically, we do the following:

1. If $\text{cv}<1$, then we use a MECon distribution (Mixture of Erlangs of consecutive order, see for instance Tijms, 1994) and extract its implied third moment.
2. If $\text{cv}\geqslant 1$, then we use a balanced hyperexponential distribution (see for instance Sauer and Chandy, 1975) and extract its implied third moment.

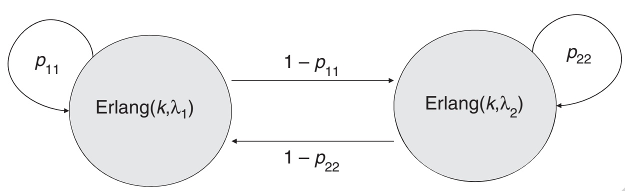

Figure 3 The Markov chain that describes Markov-MECO interarrival times.

For a Markov-MECO, the dependence can be specified either as ρ_1 or as $1+2\sum_{j=1}^{\infty}\rho_j$; these two are equivalent as the Markov-MECO has geometrically decreasing autocorrelations (ie, $\rho_j=\rho_1 v^j$, where v is a function of the Markov-MECO parameters). In our implementation, the user specifies ρ_1.

Figure 4 shows the interface to our tool for allowing users to easily specify and modify an NSNR arrival process. The user is asked for a piecewise-constant arrival-rate function, a simulation end time, a number of replications, and a desired cv^2 and ρ_1. The software—which is written in VBA for Excel—then produces a spreadsheet of arrival or interarrival times, with one replication per column, that could be read into a simulation program. Notice that specifying a constant arrival rate with $\text{cv}^2=1$ and $\rho_1=0$ gives a Poisson arrival

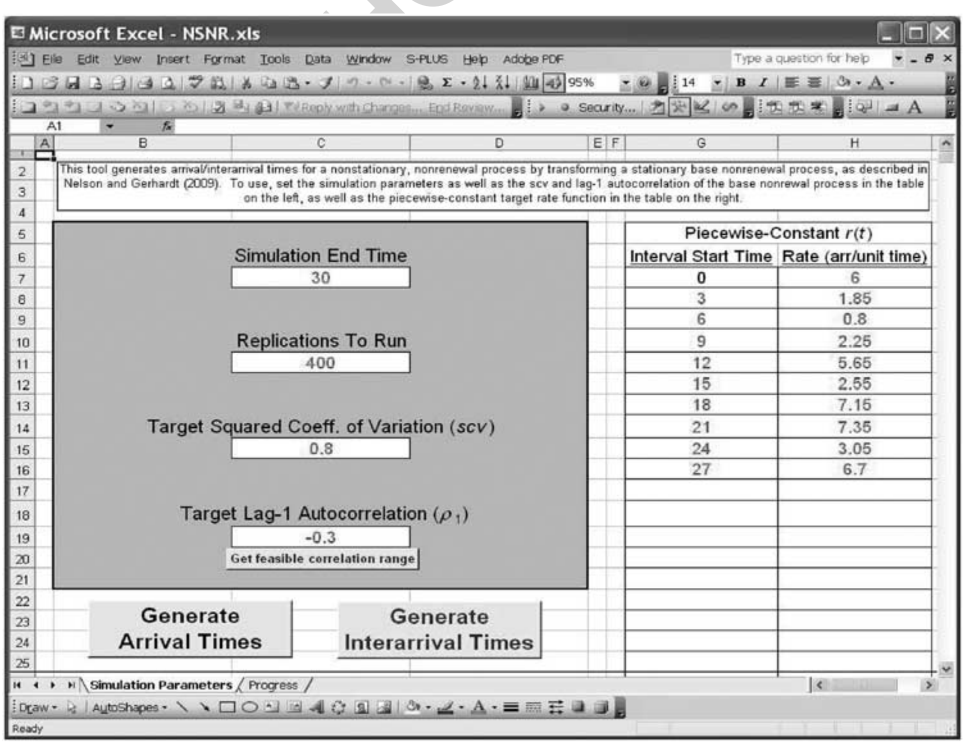

Figure 4 Markov-MECO-based tool for generating NSNR arrival processes.

process, and the same specifications with a time-varying arrival rate provide an NSPP. The spreadsheet is available for download at users.iems.northwestern.edu/$\sim$nelsonb/NSNR.xls.

6. Illustration

It is intuitively clear that if the load on a service system varies significantly over time then ignoring the time-dependent arrival rate may mask a significant aspect of system performance. This is why the NSPP is so widely applied. However, it is less well known that deviations from 'Poissonness' also matter.

Many queueing approximations account for the impact of correlation among interarrival times by adjusting the variance of a renewal arrival process rather than actually incorporating dependence (eg, Whitt, 1981). However, dependence can have an effect that is distinct from variability, as we illustrate in this section. Therefore, it is important to be able to control both variability and dependence in arrival processes.

Consider an arrival process with the piecewise-constant arrival rate given in Table 1. If the arrival process is an NSPP, then this fully characterizes it. Suppose that arrivals join a single-server, first-come-first-served queue with exponentially distributed service times (with mean 1/6), and we are interested in the mean and standard deviation of the number of customers in the queue over time.

If the arrival process is not Poisson, then our method allows control of $\mathrm{Var}\{I(t)\}/\mathrm{E}\{I(t)\}$ through the IDI; notice that if variability and dependence were interchangeable when it comes to queueing performance, then only the IDI would matter and not cv^2 and $1+2\sum_{j=1}^{\infty}\rho_j$ individually.

To illustrate that this is not the case, we feed the queue with two arrival processes that have the same $\mathrm{IDI}=524$ but are obtained via different combinations of cv^2 and $1+2\sum_{j=1}^{\infty}\rho_j$. In the first queue, the base process is a MECO renewal process with $\mathrm{cv}^2=524$ and $\rho_1=0$, yielding what Gerhardt and Nelson (2009) call an NSNP arrival process. In the second queue, the base process is a Markov-MECO with $\mathrm{cv}^2=2$ and $\rho_1=0.9$ (also yielding $\mathrm{IDI}=524$), giving NSNR arrivals.

Table 1 A piecewise-constant arrival rate function

Interval start time	*Rate (arrivals/time)*
0	6
5	1.85
10	0.8
15	2.25
20	5.65
25	2.55
30	7.15
35	7.35
40	3.05
45	3.2

Plots of the time-dependent mean and standard deviation of the queue size are provided in Figures 5 and 6. For comparison purposes, results with NSPP arrivals having the same arrival rate are included as well. For each queue the plot was produced by simulating the queue and averaging across 1000 replications; the standard errors in both plots are roughly 3% of the estimated values.

The effect of variability in the arrival process is apparent, as both moments in the non-Poisson queues significantly dwarf the corresponding moments in the NSPP queue. However, the effect of dependence is also quite noticeable, as the mean and standard deviation of queue size in the NSNP queue are typically twice as large as that of the NSNR queue. This indicates that the highly positively correlated interarrival times in the NSNR queue lead to lower variability in the queue size than in the corresponding

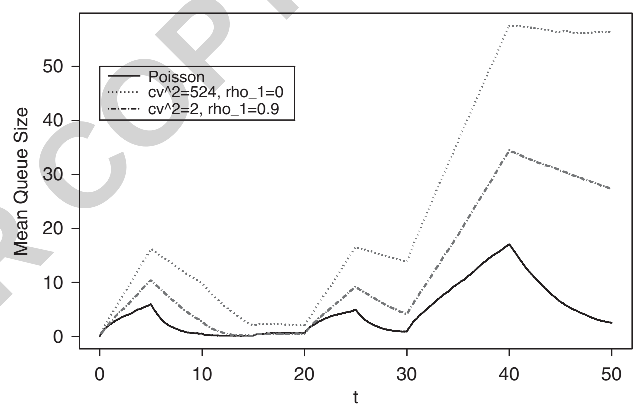

Figure 5 Time-dependent mean queue size of a $\cdots/M/1$ queue, with arrivals from an NSPP process, an NSNP process with $\mathrm{cv}^2=524$ and $\rho_1=0$, and an NSNR process with $\mathrm{cv}^2=2$ and $\rho_1=0.9$.

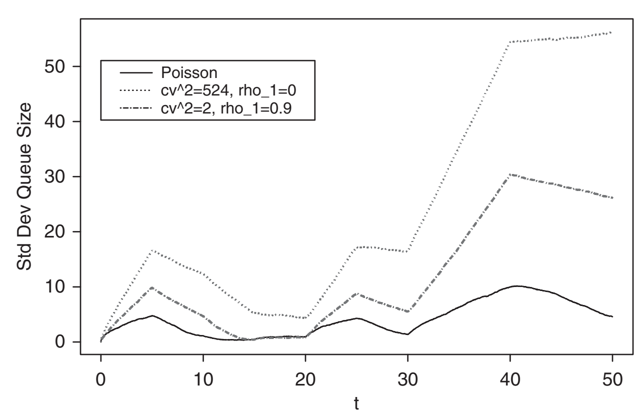

Figure 6 Time-dependent standard deviation of queue size of a $\cdots/M/1$ queue, with arrivals from an NSPP process, an NSNP process with $\mathrm{cv}^2=524$ and $\rho_1=0$, and an NSNR process with $\mathrm{cv}^2=2$ and $\rho_1=0.9$.

NSNP queue, even though both queues have arrival processes with the same arrival rate and IDI.

7. Conclusions

We have presented a basic relationship between a stationary, rate-1 base arrival process and its transformation via the inverse integrated rate function $R(t)$: the arrival rate $r(t) = \mathrm{d}R(t)/\mathrm{d}t$ is attained, and certain properties of the marginal variance and dependence structure of the base process are preserved by the transformation. Using this result we constructed a tool for defining and generating non-stationary, non-renewal arrival processes for simulation that only requires the user to provide a desired piecewise-constant arrival rate, cv^2, and lag-1 autocorrelation of the base process. With this tool the modeller can easily evaluate the impact of departures from Poissonness on the conclusions of a simulation study. An example illustrated the importance of capturing both variability and dependence in an arrival process.

The problem of estimating base-process parameters from data—which was solved for renewal base processes by Gerhardt and Nelson (2009)—is still open when the interarrival times exhibit dependence.

Acknowledgements— This work was supported by National Science Foundation Grant DMII-0521857.

References

Çinlar E (1975). *Introduction to Stochastic Processes*. Prentice-Hall: Englewood Cliffs, NJ.

Gerhardt I and Nelson BL (2009). Transforming renewal processes for simulation of non-stationary arrival processes. *INFORMS J Comput* **21**: 630–640.

Gusella R (1991). Characterizing the variability of arrival processes with indexes of dispersion. *IEEE J Select Areas Commun* **9**: 203–211.

Johnson MA (1998). Markov-MECO: A simple Markovian model for approximating non-renewal arrival processes. *Commun Statist-Stochast Models* **14**: 419–442.

Johnson MA and Taaffe MR (1989). Matching moments to phase distributions: Mixtures of Erlang distributions of common order. *Commun Statist-Stochast Models* **5**: 711–743.

Law AM (2007). *Simulation Modeling and Analysis*. 4th edn. McGraw Hill: New York.

Leemis LM (2006). Arrival processes, random lifetimes and random objects. In: Henderson SG and Nelson BL (eds). *Handbooks in Operations Research and Management Science*. Chapter 6. North-Holland: New York.

Leland WE, Taqqu MS, Willinger W and Wilson DV (1994). On the self-similar nature of ethernet traffic. *IEEE/ACM Trans Networking* **21**: 1–15.

Sauer C and Chandy K (1975). Approximate analysis of central server models. *IBM J Res Dev* **19**: 301–313.

Sriram K and Whitt W (1986). Characterizing superposition arrival processes in packet multiplexers for voice and data. *IEEE J Select Areas Commun* **4**: 833–846.

Tijms HC (1994). *Stochastic Models: An Algorithmic Approach*. Wiley: New York.

Whitt W (1981). Approximating a point process by a renewal process: The view through a queue, an indirect approach. *Mngt Sci* **27**: 619–634.

Whitt W (2002). *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*. Springer: New York.

Whitt W (2007). What you should know about queueing models to set staffing requirements in service systems. *Nav Res Logist* **54**: 476–484.

Received 30 September 2009; accepted 9 August 2010 after one revision