Submitted to *Manufacturing & Service Operations Management*
manuscript (Please, provide the mansucript number!)

Are Call Center and Hospital Arrivals Well Modeled by Nonhomogeneous Poisson Processes?

Song-Hee Kim
Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027, sk3116@columbia.edu, http://www.columbia.edu/~sk3116/

Ward Whitt
Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027, ww2040@columbia.edu, http://www.columbia.edu/~ww2040/

Service systems such as call centers and hospitals typically have strongly time-varying arrivals. A natural model for such an arrival process is a nonhomogeneous Poisson process (NHPP), but that should be tested by applying appropriate statistical tests to arrival data. Assuming that the NHPP has a rate that can be regarded as approximately piecewise-constant, a Kolmogorov-Smirnov (KS) statistical test of a Poisson process (PP) can be applied to test for a NHPP, by combining data from separate subintervals, exploiting the classical conditional-uniform property. In this paper we apply KS tests to call center and hospital arrival data and show that they are consistent with the NHPP property, but only if that data is analyzed carefully. Initial testing rejected the NHPP null hypothesis, because it failed to take account of three common features of arrival data: (i) data rounding, e.g., to seconds, (ii) over-dispersion caused by combining data from multiple days that do not have the same arrival rate, and (iii) choosing subintervals over which the rate varies too much. In this paper we investigate how to address each of these three problems.

Key words: arrival processes, nonhomogeneous Poisson process, Kolmogorov-Smirnov statistical test, data rounding, over-dispersion
History: September 12, 2013

1. Introduction

Significant effort is being made to apply operations management approaches to improve the performance of service systems such as call centers and hospitals (Aksin et al. 2007, Armony et al. 2011). Since call centers and hospitals typically have strongly time-varying arrivals, when analyzing the performance to allocate resources (e.g., staffing), it is natural to model the arrival process as a nonhomogeneous Poisson process (NHPP). We usually expect these arrival processes to be well modeled by NHPP's, because the arrivals typically arise from the independent decisions of many different people, each of whom uses the service system only rarely. There is a supporting limit theorem, often called the Poisson superposition theorem or the Poisson law of rare events; e.g., see Barbour et al. (1992) and §11.2 of Daley and Vere-Jones (2008).

Nevertheless, there are phenomena that can prevent the Poisson property from occurring. For example, scheduled appointments as at a doctor's office and enforced separation in airplane landings at airports tend to make the arrival processes less variable or less bursty than Poisson. On the other hand, arrival processes tend to be more variable or more bursty than Poisson if they involve overflows from other finite-capacity systems, as occur in hospitals (Asaduzzaman et al. 2010, Litvak et al. 2008) and in requests for reservations at hotels, because the overflows tend to occur in clumps during those intervals when the first system is full. Indeed, there is a long history studying overflow systems in teletraffic engineering (Cooper 1982, Wilkinson 1956). Bursty arrival processes also occur if the arrivals directly occur in batches, as in arrivals to hospitals from accidents. In restaurants arrivals occur in groups, but the group usually can be regarded as a single customer. In contrast, in hospitals batch arrivals typically use resources as individuals.

From the extensive experience in teletraffic engineering (Cooper 1982, Wilkinson 1956), it is known that departures from the Poisson property can have a strong impact upon performance. We emphasize this key point by showing the results of simulation experiments in §3 of the online supplement.

1.1. Exploiting the Conditional Uniform Property

Hence, to study the performance of any given service system, it is appropriate to look closely at arrival data and see if an NHPP is appropriate. A statistical test of an NHPP was suggested by Brown et al. (2005). Assuming that the arrival rate can be regarded as approximately piecewise-constant (PC), they proposed applying the classical *conditional uniform* (CU) property over each interval where the rate is approximately constant. For a Poisson process (PP), the CU property states that, conditional on the number n of arrivals in any interval $[0,T]$, the n ordered arrival times, each divided by T, are distributed as the order statistics of n independent and identically distributed (i.i.d.) random variables, each uniformly distributed on the interval $[0,1]$. Thus, under the NHPP hypothesis, if we condition in that way, the arrival data over several intervals of each day and over multiple days can all be combined into one collection of i.i.d. random variables uniformly distributed over $[0,1]$.

Brown et al. (2005) suggested applying the Kolmogorov-Smirnov (KS) statistical test to see if the resulting data is consistent with an i.i.d. sequence of uniform random variables. To test for n i.i.d. random variables X_j with *cumulative distribution function* (cdf) F, the KS statistic is the uniform distance between the *empirical cdf* (ecdf)

$$\bar{F}_n(t) \equiv \frac{1}{n}\sum_{j=1}^{n} 1_{\{(X_j/T)\le t\}}, \quad 0 \le t \le 1, \tag{1}$$

and the cdf F, i.e., the *KS test statistic* is

$$D_n \equiv \sup_{0\le t\le 1} |\bar{F}_n(t) - F(t)|. \tag{2}$$

We call the KS test of a PP after applying the CU property to a PC NHPP the CU KS test; it uses (2) with the uniform cdf $F(t) = t$. The KS test compares the observed value of D_n to the *critical value*, $\delta(n, \alpha)$, and the null hypothesis is rejected at significance level α if $D_n > \delta(n, \alpha)$ where $P(D_n > \delta(n, \alpha)) = \alpha$. In this paper, we always take α to be 0.05, in which case it is known that $\delta(n, \alpha) \approx 1.36/\sqrt{n}$ for $n > 35$.

1.2. The Possibility of a Random Rate Function

It is significant that the CU property eliminates all nuisance parameters; the final representation is independent of the rate of the PP on each subinterval. That is crucial for testing a PC NHPP, because it allows us to combine data from separate intervals with different rates on each interval. The CU KS test is thus the same as if it were for a PP. However, it is important to recognize that the constant rate on each subinterval could be random; a good test result does *not* support any candidate rate or imply that the rate on each subinterval is deterministic. Thus those issues remain to be addressed. For dynamic time-varying estimation needed for staffing, that can present a challenging forecasting problem, as reviewed in Ibrahim et al. (2012) and references therein.

By applying the CU transformation to different days separately, as well as to different subintervals within each day as needed to warrant the PC rate approximation, this method accommodates the commonly occurring phenomenon of day-to-day variation, in which the rate of the Poisson process randomly varies over different days; see, e.g., Avramidis et al. (2004), Ibrahim et al. (2012), Jongbloed and Koole (2001). Indeed, the CU transformation applied in that way makes the statistical test actually be for a Cox process, i.e., for a doubly stochastic PP, where the random rate is constant over each subinterval over which the CU transformation is applied.

Indeed, even though we will not address this issue here, there is statistical evidence that the rate function often should be regarded as random. If the historical data can be regarded as an i.i.d. sample of the unknown arrival rate, it may be possible to use a data-driven staffing algorithm as suggested by Bassamboo and Zeevi (2009). It is important to note that where these more complex models with random rate function are used, as in Bassamboo and Zeevi (2009) and Ibrahim et al. (2012), it invariably is *assumed* that the arrival process is a Cox process, i.e., that it has the Poisson property with time-varying stochastic rate function. The statistical tests we consider can be used to test if that assumed model is appropriate.

1.3. An Additional Data Transformation

In fact, Brown et al. (2005) did *not* actually apply the CU KS test. Instead, they suggested applying the KS test based on the CU property *after* performing an additional logarithmic data transformation. Kim and Whitt (2013a) investigated why an additional data transformation is needed and what form it should take. They showed through extensive simulation experiments that the CU KS test of a PP has remarkably little power against alternative processes with non-exponential interarrival-time distributions. They showed that low power occurs because the CU property focuses on the arrival times instead of the interarrival times; i.e., it converts the arrival times into i.i.d. uniform random variables.

The experiments in Kim and Whitt (2013a) showed that the KS test used by Brown et al. (2005) has much greater power against alternative processes with non-exponential interarrival-time distributions. They also found that Lewis (1965) had discovered a different data transformation due to Durbin (1961) to use after the CU transformation, and that the Lewis KS test consistently has more power than the log KS test from Brown et al. (2005) (although the difference is small compared to the improvement over the CU KS test). Evidently the Lewis test is effective because it brings the focus back to the interarrival times. Indeed, the first step of the Durbin (1961) transformation is to re-order the interarrival times of the uniform random variables in ascending order. We give the full transformation in the online supplement.

Kim and Whitt (2013a) also found that the CU KS test of a PP should not be dismissed out of hand. Even though the CU KS test of a PP has remarkably little power against alternative processes with non-exponential interarrival-time distributions, the simulation experiments show that the CU KS test of a PP turns out to be relatively more effective against alternatives with dependent exponential interarrival times. The data transformations evidently make the other methods less effective in detecting dependence, because the re-ordering of the interarrival times weakens the dependence. Hence, here we concentrate on the Lewis and CU KS tests.

1.4. Remaining Issues in Applications

Unfortunately, it does not suffice to simply perform the Lewis KS test to arrival data, because there are other complications with the data. Indeed, when we first applied the Lewis KS test to call center and hospital arrival data, we found that the Lewis KS test inappropriately rejected the NHPP property. In this paper we address three further problems associated with applying the CU KS test and the Lewis refinement from Kim and Whitt (2013a) to service system arrival data. After applying these additional steps, we conclude that the arrival data we looked at are consistent with the NHPP property, but we would not draw any blanket conclusions. We think that it is appropriate to conduct statistical tests in each setting. Our analysis shows that this should be done with care.

First, we might inappropriately reject the NHPP hypothesis because of *data rounding*. Our experience indicates that it is common for arrival data to be rounded, e.g., to the nearest second. This often produces many 0-length interarrival times, which do not occur in an NHPP, and thus cause the Lewis KS test to reject the PP hypothesis. As in Brown et al. (2005), we find that inappropriate rejection can be avoided by un-rounding, which involves adding i.i.d. small uniform random variables to the rounded data. In §2 we conduct simulation experiments showing that rounding leads to rejecting the NHPP hypothesis, and that unrounding avoids it. We also conduct experiments to verify that unrounding does not change a non-NHPP into an NHPP. If the KS test rejects the NHPP hypothesis before the rounding and unrounding, we conclude that the same will be true after the rounding and unrounding.

Second, we might inappropriately reject the NHPP hypothesis because we *use inappropriate subintervals* over which the arrival rate function is to be regarded as approximately constant. In §3, we study how to choose these subintervals. As a first step, we make the assumption that the arrival-rate function can be reasonably well approximated by a piecewise-linear function. In service systems, non-constant linear arrival rates are often realistic because they can capture a rapidly rising arrival rate at the beginning of the day and a sharply decreasing arrival rate at the end of the day, as we illustrate in our call center examples. Indeed, ways to fit linear arrival rate functions have been studied in Massey et al. (1996). However, we do not make use of this estimated arrival rate function in our final statistical test; we use it only as a means to construct an appropriate PC rate function to use in the KS test. We develop simple practical guidelines for selecting the subintervals.

Third, we might inappropriately reject the NHPP hypothesis because, in an effort to obtain a larger sample size, we might *inappropriately combine data from multiple days*. We might avoid the time-of-day effect and the day-of-the-week effect by collecting data from multiple weeks, but only from the same time of day on the same day of the week. Nevertheless, as discussed in §1.2 and §4, the arrival rate may vary substantially over these time intervals over multiple weeks. We may fail to recognize that, even though we look at the same time of day and the same day of the week, that data from multiple weeks may in fact have variable arrival rate. That is, there may be over-dispersion in the arrival data. It is often not difficult to test for such over-dispersion, using standard methods, provided that we remember to do so. Even better is to use more elaborate methods, as in Ibrahim et al. (2012) and references therein. If these tests do indeed find that there is such over-dispersion, then we should not simply reject the NHPP hypothesis. Instead, the data may be consistent with i.i.d. samples of a Poisson process, but one for which the rate function should be regarded as a stochastic process.

After investigating those three causes for inappropriately rejecting the NHPP hypothesis, in §5 we illustrate these methods with call center and hospital arrival data. We draw conclusions in §6. There is a short online supplement maintained by the journal and a longer appendix available on the authors' web pages.

2. Data Rounding

A common feature of arrival data is that arrival times are rounded to the nearest second or even the nearest minute. For example, a customer who arrives at 11:15:25.04 and another customer who arrives at 11:15:25.55 may both be given the same arrival time stamp of 11:15:25. That produces batch arrivals or, equivalently, interarrival times of length 0, which do not occur in an NHPP. If we do not take account of this feature, the KS test may inappropriately reject the NHPP null hypothesis.

Of course, this rounding problem can be addressed by having accurate arrival data without rounding, but often it is not possible to get the more accurate data, e.g., because the measurements themselves are rounded. Nevertheless, as observed by Brown et al. (2005), it is not difficult to address the rounding problem in a reasonable practical way by appropriately un-rounding the rounded data. If the data has been rounded by truncating, then we can un-round by adding a random value to each observation. If the rounding truncated the fractional component of a second, then we add a random value uniformly distributed on the interval $[0,1]$ seconds. We let these random values be i.i.d.

To study rounding, we conducted simulation experiments. We simulated 1000 replications of an NHPP with constant rate $\lambda = 1000$ (an ordinary Poisson process) on the interval $[0,6]$. We then apply the CU KS test and the Lewis KS test, as described in Kim and Whitt (2013a), to three versions of the simulated arrival data: (i) raw; as they are, (2) rounded; rounded to the nearest second, and (3) unrounded; in which we first round to the nearest second and then afterwards unround by adding uniform random variables on $[0,1]$ divided by 3600 (since the units are hours and the rounding is to the nearest second) to the arrival times from (2), as was suggested in Brown et al. (2005).

Table 1 provides the test results. #P is the number of KS tests passed at significance level $\alpha = 0.05$ out of 1000 replications. It shows the average p-values under ave[p-value] and the average percentage of 0 values in the transformed sequence under ave[% 0]. The Lewis test rejects the null Poisson hypothesis when the arrival data are rounded, but the CU KS test fails to reject. In fact, it is appropriate to reject the Poisson hypothesis when the data are rounded, because the rounding produces too many 0-valued interarrival times. Since the Lewis test looks at the ordered interarrival times, all these 0-valued interarrival times are grouped together at the left end of the interval. Table 1 reports that indeed, because of the rounding, 12.7% of the interarrival times are 0. As a consequence, the Lewis test strongly rejects the Poisson hypothesis when the data are rounded. Since the CU test looks at the data in order of the initial arrival times, the 0 interarrival times are spread out throughout the data and are not detected by the CU KS test.

Table 1 Performance of the alternative KS tests with and without rounding effects.

	CU			Lewis		
Type	# P	$ave[p-value]$	$ave[\%0]$	# P	$ave[p-value]$	$ave[\%0]$
Raw	944	0.50	0.0	955	0.50	0.0
Rounded	945	0.50	0.0	0	0.00	12.7
Unrounded	945	0.50	0.0	961	0.50	0.0

Fortunately, the problem of data rounding is well addressed by unrounding. After the rounding, the Lewis KS test of a PP fails to reject the Poisson hypothesis when applied to a PP. As in Kim and Whitt (2013a), we find that plots of the empirical cdf's used in the KS tests are very revealing. Figure 1 compares the average ecdf based on 100 replications of a rate-1000 Poisson process on the time interval [0,6] with the cdf of the null hypothesis. We note that the average ecdf and its 95% confidence interval overlap (if the manuscript is seen in color, this is better detected; the average ecdf is in red and its 95% confidence interval is in green). From these plots, we clearly see that the Lewis test is very effective, whereas the CU KS test fails to detect any problem at all.

Figure 1 **Comparison of the average ecdf for a rate-1000 Poisson process. From top to bottom: CU, Lewis test. From left to right: Raw, Rounded, Unrounded.**

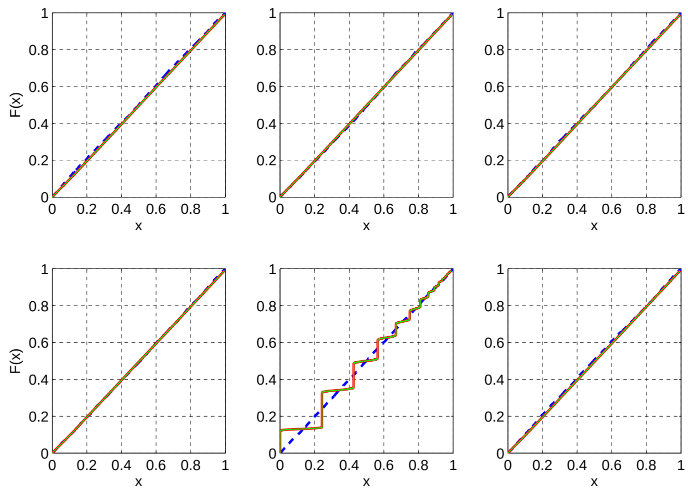
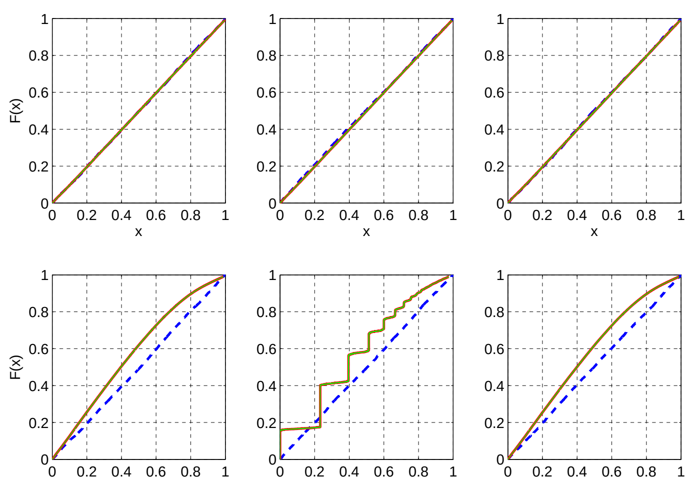


Table 2 **Performance of the alternative KS tests with and without rounding effects, applied to data with $\mathbf{H_2}$ interarrival times.**

	CU			Lewis		
Type	# P	$ave[p-value]$	$ave[\%0]$	# P	$ave[p-value]$	$ave[\%0]$
Raw	705	0.21	0.0	0	0.00	0.0
Rounded	706	0.21	0.0	0	0.00	16.2
Rounded + Uni[0,1]/3600	706	0.21	0.0	0	0.00	0.0

Figure 2 **Comparison of the average ecdf of a rate-1000 arrival process with H_2 interarrival times. From top to bottom: CU, Lewis test. From left to right: Raw, Rounded, Unrounded.**



One might naturally wonder whether unrounding makes an arrival process more like an NHPP than it was before the rounding was performed. In order to examine this issue, we simulated 1000 replications of a renewal arrival process with constant rate $\lambda = 1000$ and i.i.d. hyperexponential (H_2; a mixture of two exponentials, and hence more variable than exponential) interarrival times X_j with the squared coefficient of variation $c_X^2 = 2$ on the interval $[0,6]$. The cdf of H_2 is $P(X \le x) \equiv 1 - p_1 e^{-\lambda_1 x} - p_2 e^{-\lambda_2 x}$. We further assume balanced means for $(p_1\lambda_1^{-1} = p_2\lambda_2^{-1})$ as in (3.7) of Whitt (1982) so that $p_i = [1 \pm \sqrt{(c_X^2-1)/(c_X^2+1)}]/2$ and $\lambda_i = 2p_i$. Table 1 of Kim and Whitt (2013a) shows that the Lewis test is usually able to detect this departure from the Poisson property and to reject the Poisson hypothesis.

Table 2 shows the results of applying the CU test and the Lewis test to the renewal arrival process with H_2 interarrival times. We see that the Lewis KS test rejects the Poisson hypothesis for the raw data, as it should, but the CU KS test fails to reject. Moreover, we observe that rounding and unrounding does not eliminate the non-Poisson property. This non-Poisson property of the H_2 renewal process is detected by the Lewis test after the rounding and unrounding. Figure 2 again provides dramatic visual support as well.

3. Choosing Subintervals With Nearly Constant Rate

In order to exploit the CU property to conduct KS tests of an NHPP, we assume that the rate function is approximately piecewise-constant (PC). Since the arrival rate evidently changes relatively slowly in applications, the PC assumption should be reasonable, provided that the subintervals are chosen appropriately. However, some care is needed, as we show in this section. Before starting, we should note that there are competing interests. Using shorter intervals makes the piecewise-constant approximation more likely to be valid, but any dependence in the process from one interval to the next is lost when combining data from subintervals, so we would prefer longer subintervals unless the piecewise-constant approximation ceases to be appropriate.

As a reasonable practical first step, we propose approximating any given arrival rate function by a piecewise-linear arrival rate function with finitely many linear pieces. Ways to fit linear arrival rate functions were studied in Massey et al. (1996), and that can be extended to piecewise-linear arrival rate functions. But it is usually not necessary to have a formal estimation procedure in order to obtain a suitable rough approximation. In particular, we do not assume that we should necessarily consider the arrival rate function as fully known after this step; instead, we assume it is sufficiently well known to determine how to construct an appropriate PC approximation.

In this section we develop theory to support choosing subintervals for any given linear arrival rate function, which we *do* take as fully known. This theory leads to simple practical guidelines for evaluating whether (i) a constant approximation is appropriate for any given subinterval with linear rate and (ii) a piecewise-constant approximation is appropriate for any candidate partition of such a subinterval into further equally spaced subintervals; see §3.4 and §3.6, respectively. Equally spaced subintervals is only one choice, but the constant length is convenient to roughly judge the dependence among successive intervals.

3.1. A Call Center Example

We start by considering an example motivated by the banking call center data used in Kim and Whitt (2013b, 2012). For one 17-hour day, represented as $[6,23]$ in hours, they produced the fitted arrival rate function

$$\lambda(t) = \begin{cases} 140(t-6) & \text{on } [6,10], \\ 560 & \text{on } [10,16], \\ 560-230(t-16) & \text{on } [16,18], \\ 100-20(t-18) & \text{on } [18,23], \end{cases} \tag{3}$$

as shown in Figure 3 (taken from Kim and Whitt (2013b, 2012)). This fitted arrival rate function is actually

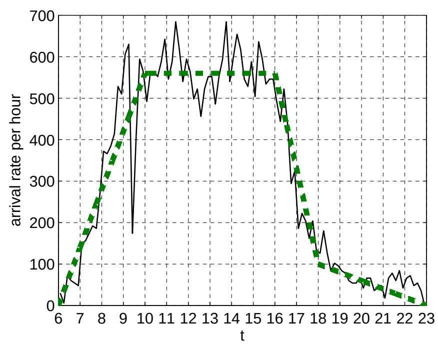


Figure 3 Fitted piecewise-linear arrival rate function for the arrivals at a banking call center.

constant in the subinterval $[10,16]$, which of course presents no difficulty. However, as in many service systems, the arrival rate is increasing at the beginning of the day, as in the subinterval $[6,10]$, and decreasing at the end of the day, as in the two intervals $[16,18]$ and $[18,23]$.

We start by considering an example motivated by Figure 3. The first interval $[6,10]$ in Figure 3 with linear increasing rate is evidently challenging. To capture the spirit of that case, we consider an NHPP with linear arrival rate function $\lambda(t) = 1000t/3$ on the interval $[0,6]$. The expected total number of arrivals over this interval is 6000. We apply simulation to study what happens when we divide the interval $[0,6]$ into $6/L$ equally spaced disjoint subintervals, each of length L, apply the CU construction to each subinterval separately and then afterwards combine all the data from the subintervals.

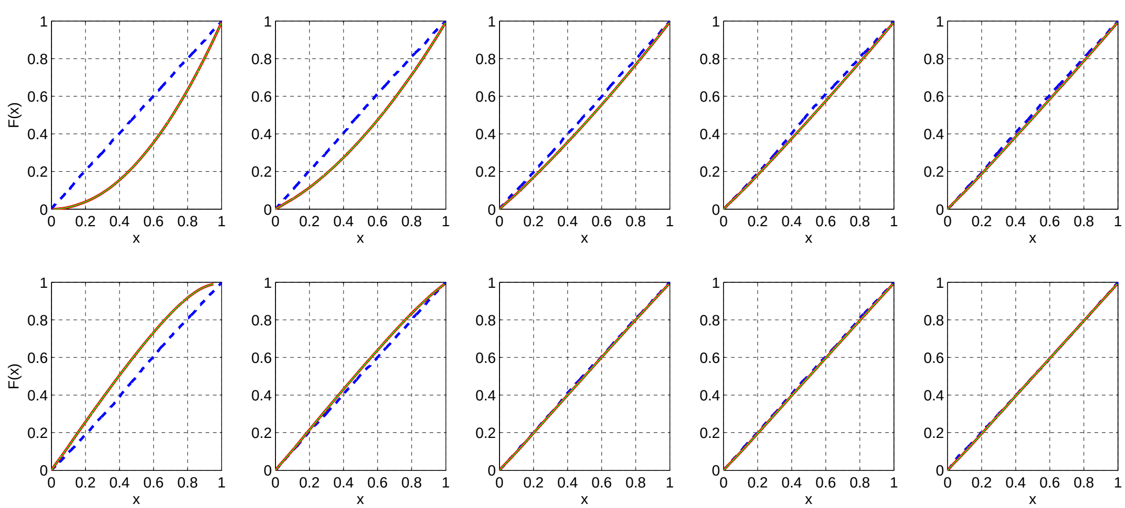
Table 3 and Figure 4 show the performance of the Lewis and CU KS tests as a function of the subinterval length. As before, #P is the number of KS tests passed at significance level $\alpha = 0.05$ out of 1000 replications. It shows the average p-values under ave[p-value] and the average percentage of 0 values in the transformed sequence under ave[% 0]. First, we see, just as in §2, that the Lewis test sees the rounding, but the CU test misses it completely. Second, we conclude that both KS tests will consistently detect this strong non-constant rate and *reject* the PP hypothesis with very high probability if we use $L = 6$ (the full interval $[0,6]$) or even if $L = 1$ or 0.5. However, the Lewis KS tests will tend *not* to reject the PP hypothesis if we divide the interval into appropriately many equally spaced subintervals.

Since we are simulating an NHPP, the actual model differs from the PP null hypothesis only through time dependence. Consistent with the observations in Kim and Whitt (2013a), we see that the CU KS test is actually *more* effective in detecting this non-constant rate than the Lewis test. The non-constant rate produces a form of dependence, for which the CU test is relatively good. However, for our tests of the actual arrival data, we will wish to test departures from the NHPP assumption. Hence, we are primarily interested in the Lewis KS test. The results indicate that we could use $L = 0.5$ for the Lewis test.

Table 3 Performance of the alternative KS test of an NHPP as a function of the subinterval length L.

		CU			Lewis		
L	Type	# P	$ave[p-value]$	$ave[\%0]$	# P	$ave[p-value]$	$ave[\%0]$
6	Raw	0	0.00	0.0	0	0.00	0.0
	Rounded	0	0.00	0.0	0	0.00	16.2
	Unrounded	0	0.00	0.0	0	0.00	0.0
3	Raw	0	0.00	0.0	0	0.00	0.0
	Rounded	0	0.00	0.0	0	0.00	16.2
	Unrounded	0	0.00	0.0	0	0.00	0.0
1	Raw	0	0.00	0.0	797	0.33	0.0
	Rounded	0	0.00	0.0	0	0.00	16.2
	Unrounded	0	0.00	0.0	815	0.33	0.0
0.5	Raw	62	0.01	0.0	946	0.47	0.0
	Rounded	69	0.01	0.1	0	0.00	16.2
	Unrounded	66	0.01	0.0	932	0.47	0.0
0.25	Raw	570	0.19	0.0	953	0.48	0.0
	Rounded	578	0.19	0.1	0	0.00	16.3
	Unrounded	563	0.19	0.0	953	0.49	0.0

Figure 4 Comparison of the average ecdf of an NHPP with different subinterval lengths. From top to bottom: CU, Lewis test. From left to right: L = 6, 3, 1, 0.5, 0.25.



In the remainder of this section we develop theory that shows how to construct piecewise-constant approximations of the rate function. We then derive explicit formulas for the conditional cdf in three cases: (i) in general (which is complicated), (ii) when the arrival rate is linear (which is relatively simple) and (iii) when the data is obtained by combining data from equally spaced subintervals of a single interval with linear rate (which remains tractable). We then apply these results to determine when a piecewise-constant approximation can be considered appropriate for KS tests.

3.2. The Conditioning Property

We first observe that a generalization of the CU method applies to show that the scaled arrival times of a general NHPP, conditional on the number observed within any interval, can be regarded as i.i.d. random variables, but with a non-uniform cdf, which we call the *conditional cdf*, depending on the rate of the NHPP over that interval. That conditional cdf then becomes the asymptotic value of the conditional-uniform Kolmogorov-Smirnov test statistic applied to the arrival data as the sample size increases, where the sample size increases by multiplying the arrival rate function by a constant.

Let $N \equiv \{N(t) : t \geq 0\}$ be an NHPP with arrival rate function λ over a time interval $[0,T]$. We assume that λ is integrable over the finite interval of interest and strictly positive except at finitely many points. Let Λ be the associated cumulative arrival rate function, defined by

$$\Lambda(t) \equiv \int_0^t \lambda(s)\, ds, \quad 0 \leq t \leq T. \tag{4}$$

We will exploit a basic conditioning property of the NHPP, which follows by the same reasoning as for the homogeneous special case. It is significant that this conditioning property is independent of scale, i.e., it is unchanged if the arrival rate function λ is multiplied by a constant. We thus later consider asymptotics in which the sample size increases in that way.

THEOREM 1. (*NHPP conditioning property*) *Let N be an NHPP with arrival rate function $c\lambda$, where c is an arbitrary positive constant. Conditional upon $N(T) = n$ for the NHPP N with arrival rate function $c\lambda$, the n ordered arrival times X_j, $1 \leq j \leq n$, when each is divided by the interval length T, are distributed as the order statistics associated with n i.i.d. random variables on the unit interval $[0,1]$, each with cumulative distribution function* (*cdf*) *F and probability density function* (*pdf*) *f, where*

$$F(t) \equiv \Lambda(tT)/\Lambda(T) \quad \textit{and} \quad f(t) \equiv T\lambda(tT)/\Lambda(T), \quad 0 \leq t \leq 1. \tag{5}$$

In particular, the cdf F is independent of c.

We call the cdf F in (5) the *conditional cdf* associated with $N \equiv N(c\lambda, T)$. Let X_j be the j^{th} ordered arrival time in N over $[0,T]$, $1 \leq j \leq n$, assuming that we have observed $n \geq 1$ points in the interval $[0,T]$. Let $\bar{F}_n(x)$ be the *empirical cdf* (ecdf) after scaling by dividing by T, defined by

$$\bar{F}_n(t) \equiv \frac{1}{n} \sum_{k=1}^{n} 1_{\{(X_j/T) \leq t\}}, \quad 0 \leq t \leq 1. \tag{6}$$

We naturally are more likely to obtain larger and larger values of n if we increase the scaling constant c.

Observe that the ecdf $\{\bar{F}_n(t) : 0 \le t \le 1\}$ is a stochastic process with

$$E[\bar{F}_n(t)] = F(t) \quad \text{for all} \quad t, \quad 0 \le t \le 1, \tag{7}$$

where F is the conditional cdf in (5). As a consequence of Lemma 1 and the Glivenko-Cantelli theorem, we immediately obtain the following asymptotic result.

THEOREM 2. (*limit for empirical cdf*) *Assuming a NHPP with arrival rate function $c\lambda$, where c is a scaling constant, the empirical cdf of the scaled order statistics in* (6)*, obtained after conditioning on observing n points in the interval $[0,T]$ and dividing by T, converges uniformly w.p.1 as $n \to \infty$* (*which may be obtained from increasing the scaling constant c*) *to the conditional cdf F in* (5)*, i.e.,*

$$\sup_{0 \le t \le 1} |\bar{F}_n(t) - F(t)| \to 0 \quad as \quad n \to \infty. \tag{8}$$

We will usually omit the scaling constant in our discussion, but with the understanding that it always can be introduced. Since we want to see how the NHPP fares in a KS test of a PP, it is natural to measure the *degree of nonhomogeneity in the NHPP* by

$$\nu(NHPP) \equiv \nu(\lambda, T) = D \equiv \sup_{0 \le t \le 1} |F(t) - t|, \tag{9}$$

where F is the conditional cdf in (5). The degree of nonhomogeneity is closely related to the CU KS test statistic for the test of a PP, which is the absolute difference between the ecdf and the uniform cdf, i.e.,

$$D_n \equiv \sup_{0 \le t \le 1} |\bar{F}_n(t) - t|; \tag{10}$$

see Marsaglia et al. (2003), Massey (1951), Miller (1956), Simard and L'Ecuyer (2011).

As a consequence of Theorem 2, we can describe the behavior of the conditional-uniform (CU) KS test of a Poisson process applied to a NHPP with general arrival rate function λ.

THEOREM 3. (*limit of the KS test of a Poisson process applied to an NHPP*) *As $n \to \infty$ in a NHPP with rate function λ over* $[0,T]$,

$$D_n \to D \equiv \sup_{0 \le t \le 1} |F(t) - t|, \tag{11}$$

where D_n is the CU KS test statistic in (10) *and D is the degree of nonhomogeneity in* (9) *involving the the conditional cdf F in* (5)*.*

COROLLARY 1. (*asymptotic rejection of the Poisson process hypothesis if NHPP is not a Poisson process*) *The probability that an NHPP with rate function $n\lambda$ will be rejected by the CU KS test for a PP converges to* 1 *as the scaling parameter $n \to \infty$ if and only if the λ is not constant w.p.1, i.e., if and only if the NHPP is not actually a PP.*

Proof. It is easy to see that the cdf F in (5) coincides with the uniform cdf t if and only if $\lambda(t)$ is constant. ■

Corollary 1 suggests that a piecewise-constant approximation of a non-PP NHPP never makes sense with enough data, but we develop a positive result exploiting appropriate subintervals, where the number of subintervals grows with the sample size n; see Theorem 6.

3.3. An NHPP with Linear Arrival Rate Function

We now consider the special case of an NHPP with linear arrival rate function

$$\lambda(t) = a + bt, \quad 0 \le t \le T, \tag{12}$$

The analysis is essentially the same for increasing and decreasing arrival rate functions, so that we will assume that the arrival rate function is increasing, i.e., $b \ge 0$. There are two cases: $a > 0$ and $a = 0$; we shall consider them both. If $a > 0$, then cumulative arrival rate function can be expressed as

$$\Lambda(t) \equiv \int_0^t \lambda(s)\, ds = at + \frac{bt^2}{2} = a\left(t + \frac{rt^2}{2}\right) \tag{13}$$

where $r \equiv b/a$ is the *relative slope*. If $a = 0$, then $\Lambda(t) = bt^2/2$.

THEOREM 4. (*aymptotic maximum absolute difference in the linear case*) *Consider an NHPP with linear arrival rate function in* (12) *observed over the interval* $[0,T]$. *If* $a > 0$, *then the conditional cdf in* (5) *assumes the form*

$$F(t) = \frac{tT + (r(tT)^2/2)}{T + (rT^2/2)}, \quad 0 \le t \le 1; \tag{14}$$

if $a = 0$, *then*

$$F(t) = t^2, \quad 0 \le t \le 1. \tag{15}$$

Thus, if $a > 0$, *then the degree of nonhomogeneity of the NHPP can be expressed explicitly as*

$$D \equiv D(rT) \equiv \sup_{0 \le t \le 1} \{|F(t) - t|\} = |F(1/2) - 1/2| = \frac{1}{2} - \frac{\left(\frac{T}{2} + \frac{rT^2}{8}\right)}{\left(T + \frac{rT^2}{2}\right)} = \frac{rT}{8 + 4rT}. \tag{16}$$

If $a = 0$, *then* $D = 1/4$ (*which agrees with* (16) *when* $r = \infty$).

Proof. For (16), observe that $|F(t) - t|$ is maximized where $f(t) = 1$, so that it is maximized at $t = 1/2$. ■

3.4. Practical Guidelines for a Single Interval

We can apply formula (16) in Theorem 4 to judge whether an NHPP with linear rate over an interval should be close enough to a PP with constant rate. (We see that should never be the case for a single interval with $a = 0$ because then $D = 1/4$.) In particular, the rate function can be regarded as approximately constant if the ratio $D/\delta(n,\alpha)$ is sufficiently small, where D is the degree of homogeneity in (16) and $\delta(n,\alpha)$ is the critical value of the KS test statistic D_n with sample size n and significance level α, which we always take to be $\alpha = 0.05$. Before looking at data, we can estimate n by the expected total number of arrivals over the interval.

We have conducted simulation experiments to determine when the ratio $D/\delta(n,\alpha)$ is sufficiently small that the KS test of a PP applied to an NHPP with that rate function consistently rejects the PP null hypothesis with probability approximately $\alpha = 0.05$. Our simulation experiments indicate that a ratio of 0.10 (0.50) should be sufficiently small for the CU (Lewis) KS test with a significance level of $\alpha = 0.05$.

Table 4 illustrates by showing the values of D, $\delta(n,\alpha)$ and $D/\delta(n,\alpha)$ along with the test results for selected subintervals of the initial example with $\lambda(t) = 1000t/3$ on the time interval $[0,6]$, just as in Table 3. (More examples appear in the appendix.)

Table 4 Judging when the rate is approximately constant: looking at the ratio $D/\delta(n,\alpha)$ for $\alpha = 0.05$

							CU		Lewis	
L	Interval	$ave[n]$	r	D	$ave[\delta(n,\alpha)]$	$D/ave[\delta(n,\alpha)]$	# P	$ave[p-value]$	# P	$ave[p-value]$
6	[0,6]	5997.3	∞	0.250	0.018	14.28	0	0.00	0	0.00
3	[0,3]	1498.8	∞	0.250	0.035	7.15	0	0.00	0	0.00
	[3,6]	4498.5	0.33	0.083	0.020	4.12	0	0.00	481	0.15
1	[0,1]	166.8	∞	0.250	0.104	2.40	0	0.00	46	0.01
	[1,2]	499.7	1.00	0.083	0.060	1.38	22	0.01	896	0.43
	[2,3]	832.4	0.50	0.050	0.047	1.07	145	0.03	928	0.48
	[3,4]	1166.9	0.33	0.036	0.040	0.90	300	0.08	931	0.49
	[4,5]	1501.0	0.25	0.028	0.035	0.79	358	0.09	949	0.49
	[5,6]	1830.6	0.20	0.023	0.032	0.72	453	0.13	948	0.49
0.5	[0,0.5]	42.0	∞	0.250	0.207	1.21	46	0.01	562	0.18
	[0.5,1]	124.8	2.00	0.083	0.121	0.69	479	0.14	918	0.48
	[1,1.5]	207.7	1.00	0.050	0.094	0.53	684	0.25	935	0.49
	[1.5,2]	292.0	0.67	0.036	0.079	0.45	766	0.29	945	0.50
	[2,2.5]	375.4	0.50	0.028	0.070	0.40	783	0.33	935	0.49
	[2.5,3]	456.9	0.40	0.023	0.063	0.36	833	0.35	960	0.51
	[3,3.5]	543.6	0.33	0.019	0.058	0.33	822	0.36	949	0.48
	[3.5,4]	623.3	0.29	0.017	0.054	0.31	865	0.38	938	0.51
	[4,3.5]	708.9	0.25	0.015	0.051	0.29	861	0.40	957	0.51
	[4.5,5]	792.1	0.22	0.013	0.048	0.27	882	0.41	936	0.50
	[5,5.5]	873.9	0.20	0.012	0.046	0.26	873	0.42	941	0.49
	[5.5,6]	956.6	0.18	0.011	0.044	0.25	893	0.42	951	0.50
0.25	[0,0.25]	10.4	∞	0.250	0.418	0.60	588	0.17	888	0.42
	[0.25,0.5]	31.6	4.00	0.083	0.239	0.35	841	0.37	946	0.49
	[0.5,0.75]	51.8	2.00	0.050	0.187	0.27	885	0.41	943	0.49
	[0.75,1]	73.0	1.33	0.036	0.157	0.23	907	0.44	947	0.50
	[1,1.25]	93.7	1.00	0.028	0.139	0.20	902	0.45	938	0.49
	[1.25,1.5]	114.0	0.80	0.023	0.126	0.18	920	0.47	951	0.50
	[1.5,1.75]	135.5	0.67	0.019	0.116	0.17	936	0.46	939	0.50
	[1.75,2]	156.5	0.57	0.017	0.108	0.15	924	0.48	940	0.49
	[2,2.25]	177.6	0.50	0.015	0.101	0.15	916	0.46	968	0.51
	[2.25,2.5]	197.9	0.44	0.013	0.096	0.14	925	0.48	939	0.48
	[2.5,2.75]	218.2	0.40	0.012	0.091	0.13	934	0.48	946	0.50
	[2.75,3]	238.7	0.36	0.011	0.087	0.12	931	0.48	956	0.50
	[3,3.25]	261.1	0.33	0.010	0.084	0.12	929	0.48	941	0.49
	[3.25,3.5]	282.6	0.31	0.009	0.080	0.12	941	0.47	948	0.49
	[3.5,3.75]	301.3	0.29	0.009	0.078	0.11	942	0.48	949	0.50
	[3.75,4]	322.0	0.27	0.008	0.075	0.11	941	0.47	946	0.50
	[4,4.25]	344.2	0.25	0.008	0.073	0.10	948	0.48	952	0.50
	[4.25,4.5]	364.7	0.24	0.007	0.071	0.10	932	0.47	957	0.52
	[4.5,4.75]	385.5	0.22	0.007	0.069	0.10	952	0.47	946	0.50
	[4.75,5]	406.7	0.21	0.006	0.067	0.10	937	0.48	953	0.50
	[5,5.25]	426.8	0.20	0.006	0.065	0.09	938	0.48	951	0.50
	[5.25,5.5]	447.2	0.19	0.006	0.064	0.09	943	0.49	942	0.48
	[5.5,5.75]	467.5	0.18	0.006	0.062	0.09	945	0.47	952	0.50
	[5.75,6]	489.2	0.17	0.005	0.061	0.09	941	0.50	943	0.50

3.5. Subintervals for an NHPP with Linear Arrival Rate

In this section we see the consequence of dividing the interval $[0,T]$ into k equal subintervals when the arrival rate function is linear over $[0,T]$ as in §3.3. As in the CU KS test discussed in Kim and Whitt (2013a), we treat each interval separately and combine all the data. An important initial observations is that the final cdf F can be expressed in terms of the cdf's F_j associated with the k subintervals. In particular, we have the following lemma.

LEMMA 1. (*combining data from equally spaced subintervals*) *If we start with a general arrival rate function and divide the interval $[0,T]$ into k subintervals of length T/k, then we obtain i.i.d. random variables with a conditional cdf that is a convex combination of the conditional cdf's for the individual intervals, i.e.,*

$$\begin{aligned} F(t) &= \sum_{j=1}^{k} p_j F_j(t), \quad 0 \le t \le 1, \quad \textit{where} \\ F_j(t) &= \frac{\Lambda_j(tT/k)}{\Lambda_j(T/k)}, \quad 0 \le t \le 1, \quad 1 \le j \le k, \\ \Lambda_j(t) &= \Lambda(((j-1)T/k)+t) - \Lambda((j-1)T/k), \quad 0 \le t \le T/k, \quad 1 \le j \le k, \\ p_j &= \frac{\Lambda(jT/k) - \Lambda((j-1)T/k)}{\Lambda(T)}, \quad 1 \le j \le k. \end{aligned} \tag{17}$$

For the special case of a linear arrival rate function as in (12) *with $a > 0$,*

$$\begin{aligned} \Lambda_j(t) &= \frac{at(k(2+rt)+2(j-1)rT)}{2k}, \quad 0 \le t \le T/k, \quad 1 \le j \le k, \\ F_j(t) &= \frac{t(2k+(2j-2+t)rT)}{2k+(2j-1)rT)}, \quad 0 \le t \le 1, \quad 1 \le j \le k, \\ p_j &= \frac{2k+(2j-1)rT}{k^2(2+rT)} \quad \textit{and} \quad r_j = \frac{b}{\lambda((j-1)T/k)} = \frac{bk}{a(k+(j-1)rT)}. \end{aligned} \tag{18}$$

For the special case of a linear arrival rate function as in (12) *with $a = 0$,*

$$\begin{aligned} \Lambda_j(t) &= \frac{bt(kt+2(j-1)T)}{2k}, \quad 0 \le t \le T/k, \quad 1 \le j \le k, \\ F_j(t) &= \frac{t(2j-2+t)}{2j-1}, \quad 0 \le t \le 1, \quad 1 \le j \le k, \\ p_j &= \frac{2j-1}{k^2} \quad \textit{and} \quad r_j = \frac{k}{(j-1)T}, \quad 1 \le j \le k. \end{aligned} \tag{19}$$

We now apply Lemma 1 to obtain a simple characterization of the maximum difference from the uniform cdf when we combine the data from all the equally spaced subintervals.

THEOREM 5. (*combining data from equally spaced subintervals*) *If we start with the linear arrival rate function in* (12) *and divide the interval $[0,T]$ into k subintervals of length T/k, and combine all the data, then we obtain*

$$D \equiv \sup_{0 \le t \le 1} \{|F(t) - t|\} = \sum_{j=1}^{k} p_j D_j = \sum_{j=1}^{k} p_j \sup_{0 \le t \le 1} \{|F_j(t) - t|\}. \tag{20}$$

If $a > 0$, then

$$D = \sum_{j=1}^{k} \frac{p_j r_j T/k}{8 + 4 r_j T/k} \le C/k \quad \textit{for all} \quad k \ge 1 \tag{21}$$

for a constant C. If $a = 0$, then

$$D = \frac{p_1}{4} + \sum_{j=2}^{k} \frac{p_j/(j-1)}{8+4/(j-1)} \le C/k \quad \textit{for all} \quad k \ge 1 \tag{22}$$

for a constant C.

Proof. By Theorem 4, by virtue of the linearity, for each $j \ge 1$, $|F_j(t) - t|$ is maximized at $t = 1/2$. Hence, the same is true for $|F(t) - t|$, where $F(t) = \sum_{j=1}^{k} p_j F_j(t)$, which gives us (20). For the final bound in (21), use $r_j \le 1 + (T/ka)$ for all j. For the final bound in (22), use $r_j = (j/(j-1))^2 \le 4$ for all $j \ge 2$ with $p_1 = 1/k^2$. ■

3.6. Practical Guidelines for Dividing an Interval into Equal Subintervals

Paralleling §3.4, assuming that the rate is strictly positive on the interval (0 at one endpoint), we can apply formula (21) ((22)) in Theorem 5 to judge whether the partition of a given interval with linear rate into equally spaced subintervals produces an appropriate PC approximation. As before, we look at the ratio $D/\delta(n, \alpha)$, requiring that it be less than 0.10 (0.50) for the CU (Lewis) KS test with significance level $\alpha = 0.05$, where now D is given by (21) or (22) and $\delta(n, \alpha)$ is again the critical value to the KS test, but now applied to all the data, combining the data after the CU transformation is applied in each subinterval. In particular, n should be the total observed sample size or the total expected number of arrivals, adding over all subintervals.

We illustrate in Table 5 by showing the values of D, $\delta(n, \alpha)$ and $D/\delta(n, \alpha)$ along with the test results for each subinterval of the initial example with $\lambda(t) = 1000t/3$ on the time interval $[0, 6]$, just as in Table 3. In all cases, $ave[n]$ is 5997.33, and hence $ave[\delta(n, \alpha)]$ values are the same and are approximately 0.0175. (Again, more examples appear in the appendix.)

Table 5 Judging when the rate is approximately constant: Looking at the ratio $D/\delta(n, \alpha)$ for equal subintervals

			CU		Lewis	
L	D	$D/ave[\delta(n,\alpha)]$	# P	$ave[p-value]$	# P	$ave[p-value]$
6	0.2500	14.278	0	0.00	0	0.00
3	0.1250	7.139	0	0.00	0	0.00
1	0.0417	2.380	0	0.00	797	0.33
0.5	0.0208	1.190	62	0.01	946	0.47
0.25	0.0104	0.595	570	0.19	953	0.48
0.1	0.0042	0.238	896	0.43	955	0.48
0.09	0.0038	0.214	902	0.43	954	0.48
0.08	0.0033	0.190	914	0.45	948	0.48
0.07	0.0029	0.167	923	0.47	960	0.49
0.06	0.0025	0.143	927	0.47	941	0.49
0.05	0.0021	0.119	941	0.50	958	0.49
0.01	0.0004	0.024	953	0.50	948	0.48
0.005	0.0002	0.012	944	0.49	943	0.48
0.001	0.00004	0.002	952	0.50	959	0.49

3.7. Asymptotic Justification of Piecewise-Constant Approximation

We now present a limit theorem that provides useful insight into the performance of the CU KS test of a NHPP with linear rate. We start with a non-constant linear arrival rate function λ as in (12) and then scale it by multiplying it by n and letting $n \to \infty$. We show that as the scale increases, with the number of subintervals increasing as the scale increases appropriately, the KS test results will behave the same as if the NHPP had constant rate. We will reject if it should and fail to reject otherwise (with probability equal to the significance level). In particular, it suffices to use k_n equally spaced subintervals, where

$$\frac{k_n}{\sqrt{n}} \to \infty \quad \text{as} \quad n \to \infty. \tag{23}$$

In order to have the sample size in each interval also grow without bound, we also require that

$$\frac{n}{k_n} \to \infty \quad \text{as} \quad n \to \infty. \tag{24}$$

For example, $k_n = n^p$ satisfies both (23) and (24) if $1/2 < p < 1$.

THEOREM 6. (*asymptotic justification of the piecewise-constant approximation of linear arrival rate functions*) *Suppose that we consider a non-constant linear arrival rate function over the fixed interval* $[0,T]$ *as above scaled by* n. *Suppose that we use the CU KS test with any specified significance level* α *based on combining data over* k_n *subintervals, each of width* T/k_n. *If conditions* (23) *and* (24) *hold, then the probability that the CU KS test of the hypothesis of a Poisson process will reject the NHPP converges to* α *as* $n \to \infty$. *On the other hand, if*

$$\frac{k_n}{\sqrt{n}} \to 0 \quad \textit{as} \quad n \to \infty, \tag{25}$$

then the probability that the CU KS test of a Poisson process will reject the NHPP converges to 1 *as* $n \to \infty$.

Proof. Recall that the critical value $\delta(n,\alpha)$ of the CU KS test statistic D_n has the form $c_\alpha/\sqrt{n}$ as $n \to \infty$, where n is the sample size (see Simard and L'Ecuyer (2011)), and here the sample size is Kn for all n, where K is some constant. Let $D^{(n)}$ be D above as a function of the parameter n. Hence, we can compare the asymptotic behavior of $\delta(n,\alpha)$ to the asymptotic behavior of $D^{(n)}$, which has been determined above. Theorem 5 shows that $D^{(n)}$ is asymptotically of the form C/k_n. Hence, it suffices to compare k_n to $\sqrt{n}$ as in (23) and (25). ■

In §5 of the online supplement we conduct a simulation experiment to illustrate Theorem 6. In §6 of the online supplement we also obtain an asymptotic result paralleling Theorem 6 for a piecewise-continuous arrival rate function; it is not strictly necessary that the arrival rate function be piecewise-linear in order for the asymptotic correctness of the piecewise-constant approximation to be true. For that result, we assume that the piecewise-continuous function not only has only finitely many discontinuities in $[0,T]$, but also is right continuous and has limits from the left at each discontinuity point. Moreover, we assume that each continuous piece is Lipschitz continuous. Those properties allows us to choose subintervals within each Lipschitz continuous piece so that the oscillations within them all are uniformly small.

4. Combining Data from Multiple Days

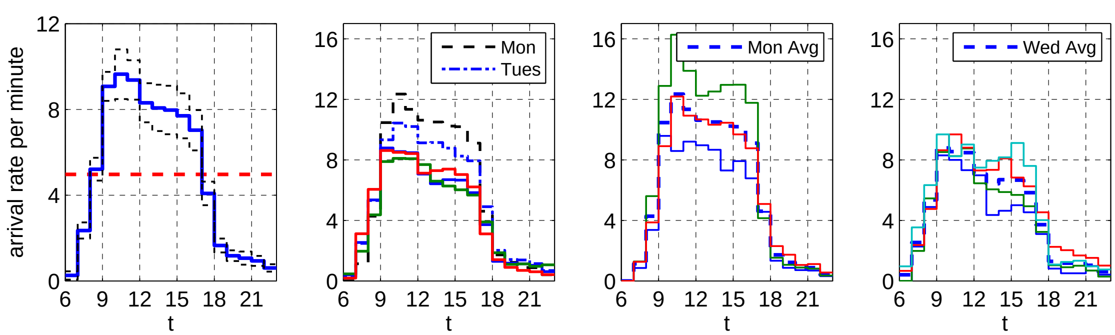
When the sample size is too small, it is natural to combine data from multiple days. For example, we may have hospital emergency department arrival data and we want to test whether the arrivals from 9am to 10am can be modeled as an NHPP. However, if there are only 10 arrivals in $[9, 10]$ on average, then data from one day alone will not be sufficient to test the PP property. A common way to address this problem is to combine data from multiple days; e.g., we can use all interarrival times in $[9, 10]$ from 20 weekdays, which will give us a sample size of about 200 interarrival times. From Kim and Whitt (2013a), we know that sample size is sufficient.

However, combining data from multiple days can be complicated because of over-dispersion, even when looking at the same time of day and the same day of the week. We first illustrate the over-dispersion problem in banking call center data used in Kim and Whitt (2013b) and §5 of this paper. Then we observe that the approach can be applied anyway if we apply the CU property to each day separately, provided that we remember that we have not yet ruled out over-dispersion. If the KS tests based on the CU property pass, then the actual process could be an NHPP with a stochastic rate function. That has not yet been ruled out. That issue can be analyzed separately.

4.1. Day-to-Day Variation in a Banking Call Center

A common feature we need to be careful about when combining data from multiple days is the day-to-day variation. Figure 5 shows illustrative plots of arrival data to a US banking call center; see §5 for detailed explanations of the data. The first plot shows that the arrival rate is highly variable from day to day, as illustrated by the wide confidence intervals for the hourly averages. Part of this day-to-day variation can be explained by day-of-week effect. The second plot shows that the average call volume on Mondays is the largest, followed by that of Tuesdays, and then the others. The third and fourth plots further illustrate day-to-day variation in the same day of week.

Figure 5 Average arrival rates. From left to right: Overall and hourly rates for 18 weekdays, arrival rates by each day of week, arrival rates on Monday, and arrival rates on Wednesday.



We can further examine whether arrival data from a given interval of the day on a given day of the week can be regarded as a sample of n i.i.d Poisson random variables with the same mean. Since the variance equals the mean for the Poisson distribution, we can directly test whether that property holds in the data. Table 6 shows the estimated values of the mean $\bar{\mu}$, variance $\bar{\sigma}^2$ with its 95% confidence interval, and the dispersion test result. It has been observed that if $x_1, x_2, ..., x_n$ are n observations from a Poisson population, then the index of dispersion $D = \sum_i (x_i - \bar{x})^2/\bar{x}$ is approximately distributed as a χ^2 statistic with $n-1$ degrees of freedom (see Kathirgamatamby (1953) and references therein). The dispersion test then uses this fact to test the null hypothesis that $x_1, x_2, ..., x_n$ are independent Poisson distributed variables with mean parameter $\bar{x}$; we report the p-value for this test. Table 6 shows that in all intervals, the estimated variance is much greater than the estimated mean (the 95% confidence interval for the variance does not include the mean in any case), and that the dispersion test at significance level 0.05 reject the null hypothesis that the counts are independent Poisson distributed variables all the time.

Table 6 Summary Statistics of the Call center arrivals: each hour on 16 Fridays.

Interval	$\bar{\mu}$	$\bar{\sigma}^2$	$\bar{D}$	p-value
[7 , 8]	180.5	12860.1, [7017.6, 30804.5]	1068.7	0.00
[8 , 9]	363.7	14078.2, [7682.3, 33722.3]	580.6	0.00
[9 , 10]	499.1	7352.8, [4012.3, 17612.5]	221.0	0.00
[10 , 11]	496.8	5411.1, [2952.8, 12961.5]	163.4	0.00
[11 , 12]	502.8	10428.9, [5690.9, 24980.8]	311.2	0.00
[12 , 13]	446.4	7782.0, [4246.5, 18640.6]	261.5	0.00
[13 , 14]	444.0	8778.0, [4790.0, 21026.4]	296.6	0.00
[14 , 15]	445.4	10669.7, [5822.3, 25557.7]	359.4	0.00
[15 , 16]	405.1	8905.1, [4859.4, 21330.9]	329.8	0.00
[16 , 17]	353.4	6723.6, [3669.0, 16105.3]	285.4	0.00
[17 , 18]	196.3	5239.9, [2859.4, 12551.5]	400.5	0.00
[18 , 19]	82.9	1778.9, [970.7, 4261.1]	322.0	0.00
[19 , 20]	57.7	752.2, [410.5, 1801.9]	195.6	0.00
[20 , 21]	52.0	1174.4, [640.9, 2813.1]	338.8	0.00
[21 , 22]	42.3	1027.9, [560.9, 2462.3]	364.9	0.00
[22 , 23]	29.2	385.4, [210.3, 923.1]	198.0	0.00

4.2. Avoiding Over-Dispersion and Testing for It

An attractive feature of the KS tests based on the CU property is that we can avoid the over-dispersion problem while testing for an NHPP. We can avoid the over-dispersion problem by applying the CU property separately to intervals from different days and then afterwards combining all the data. When the CU property is applied in this way, the observations become i.i.d. uniform random variables, even if the rates of the NHPP's are different on different days, because the CU property is independent of the rate of each interval.

However, when we apply KS tests based on the CU property in this way and conclude that the data is consistent with an NHPP, we have not yet ruled out different rates on different days, which might be modeled as a random arrival rate function over any given day. One way to test for such over-dispersion is to conduct the KS test based on the CU property by combining data from multiple days, applying the CU property by *both* (i) combining all the data before applying the CU property and (ii) applying the CU property to each day separately and then combining the data afterwards. If the data are consistent with an NHPP with fixed rate, then these two methods will give similar results. On the other hand, if there is significant over-dispersion, then the KS test will reject the NHPP hypothesis if all the data is combined before applying the CU property.

Of course, we can also directly test for the over-dispersion, as illustrated by Table 6. We have now completed our study of the three problems associated with KS tests based on the CU property, namely: (i) data rounding, (ii) choosing subintervals where the rate can be regarded as approximately constant and (iii) over-dispersion. We now illustrate these issues in data sets from call centers and hospitals.

4.3. Constructing Alternative Hypotheses with Stochastic Rate Processes

Motivated by Ibrahim et al. (2012) and related work, we may actually want to regard the arrival process model as a Poisson process with a stochastic rate process. We now observe that it is not difficult to construct alternative hypotheses related to those considered in Kim and Whitt (2013a) for testing a PP.

Let $\{\Lambda(t) : t \geq 0\}$ be an arbitrary cumulative rate stochastic process with nondecreasing sample paths, so that $\Lambda(t)$ is the random total rate over $[0,t]$, $t \geq 0$. Let $\{A(t) : t \geq 0\}$ be any rate-1 arrival counting process with stationary increments, such as one of the ones in §4.1 of Kim and Whitt (2013a), which for simplicity we take to be independent of $\{\Lambda(t) : t \geq 0\}$. Then we obtain an alternative hypothesis to a doubly-stochastic PP with a random rate function $\{\Lambda(t) : t \geq 0\}$ by letting

$$N(t) \equiv A(\Lambda(t)), \quad t \geq 0. \tag{26}$$

Since A has rate 1, $E[A(s+t) - A(s)] = t$ for all $s, t > 0$ and

$$E[N(s+t) - N(s)|\Lambda(u), \quad 0 \leq u \leq t] = E[A(\Lambda(s+t) - A(\Lambda(s))|\Lambda(u), \quad 0 \leq u \leq t] = \Lambda(s+t) - \Lambda(s),$$

so that

$$E[N(s+t) - N(s)] = E[E[N(s+t) - N(s)|\Lambda(u), \quad 0 \leq u \leq t]] = E[\Lambda(s+t) - \Lambda(s)].$$

If we can approximate the random rate function by a PC rate function, so that the random rate is constant over each subinterval, then the tests we consider are tests of the Poisson property, i.e., to test if N is a doubly stochastic NHPP. The construction here extends the study cases in Kim and Whitt (2013a) to this random-rate setting.

5. Call Center and Hospital Examples

In this section, we work with call center and hospital arrival data to show how our methods work. We first describe call center and hospital arrival datasets we have in §5.1. We start with three examples that illustrate what we need to be careful about when dealing with real data in §5.2. In §5.3, we provide summarized test results on call center and hospital arrival data we have and conclude that both can be well modeled by NHPPs. Detailed results can be found in the appendix.

5.1. Call Center and Hospital Arrival Data

We use the same **Call Center** data used in Kim and Whitt (2013b), from a telephone call center of a medium-sized American bank from the data archive of Mandelbaum (2012), collected from March 26, 2001 to October 26, 2003. This banking call center had sites in New York, Pennsylvania, Rhode Island, and Massachusetts, which were integrated to form a single virtual call center. The virtual call center had 900 - 1200 agent positions on weekdays and 200 - 500 agent positions on weekends. The center processed about 300,000 calls per day during weekdays, with about 60,000 (20%) handled by agents, with the rest being served by Voice Response Unit (VRU) technology. In this study, we focus on arrival data on April 2001. There are 4 significant entry points to the system: through VRU ~92%, Announcement ~6%, Message ~1% and Direct group (callers that directly connect to an agent) ~1%; there are a very small number of outgoing and internal calls, and we are not including them. Furthermore, among the customers that arrive to the VRU, there are five customer types: Retail ~91.4%, Premier ~1.9%, Business ~4.4%, Customer Loan ~0.3%, and Summit ~2.0%.

Hospital Emergency Department (ED) data are from one of the major teaching hospitals in South Korea, collected from September 1, 2012 to November 15, 2012; we focus on 70 days, from September 1, 2012 to November 9, 2012. There are two major entry groups, walk-ins and ambulance arrivals. On average, there are 138.5 arrivals each day with ~88% walk-ins and ~12% ambulance arrivals.

Figures 6 and 7 show average hourly arrival rate as well as individual hourly arrival rate for each arrival type on Mondays for the call center and hospital ED data, respectively. We observe strong within-day variations in call center arrivals, but not as much for the hospital ED arrivals, especially for the ambulance arrivals. Tables 7 and 8 further provide summary statistics for the number of arrivals in each day for each arrival type: estimated values of the mean $\bar{\mu}$, variance $\bar{\sigma}^2$ with its 95% confidence interval, the index of dispersion $\bar{D}$, and the p-value for the dispersion test. We can compare the index of dispersion $\bar{D}$ values to $\chi^2_{n-1,1-\alpha}$ values: for each (n, α) pair, (30, 0.05): 42.6, (21,0.05): 31.4, (70,0.05): 89.4, and (50,0.05): 66.3. We observe that the call center has significant day-to-day variation in its arrivals (the null hypothesis for Poisson distribution is rejected for every arrival type); hospital ED walk-in arrivals also have strong day-to-day variation, whereas the dispersion test for ambulance arrivals fails to reject the null hypothesis.

Table 7 Call center arrivals: summary statistics by type. (CL = Customer Loan.)

	All days (n=30)			Weekdays (n=21)		
Type	$\bar{\mu}$	$\bar{\sigma}^2$	$\bar{D}$	$\bar{\mu}$	$\bar{\sigma}^2$	$\bar{D}$
VRU-Retail	1.5×10^5	$2.1 \times 10^9, [1.3 \times 10^9, 3.8 \times 10^9]$	3.9×10^5	1.8×10^5	$1.7 \times 10^8, [9.7 \times 10^7, 3.5 \times 10^8]$	1.8×10^4
VRU-Premier	3.2×10^3	$6.3 \times 10^5, [4.0 \times 10^5, 1.2 \times 10^6]$	5.7×10^3	3.7×10^3	$7.7 \times 10^4, [4.5 \times 10^4, 1.6 \times 10^5]$	4.2×10^2
VRU-Business	7.3×10^3	$1.2 \times 10^7, [7.3 \times 10^6, 2.1 \times 10^7]$	4.6×10^4	9.4×10^3	$3.5 \times 10^5, [2.0 \times 10^5, 7.2 \times 10^5]$	7.3×10^2
VRU-CL	4.6×10^2	$2.9 \times 10^4, [1.8 \times 10^4, 5.1 \times 10^4]$	1.8×10^3	5.6×10^2	$7.0 \times 10^3, [4.1 \times 10^3, 1.2 \times 10^4]$	2.5×10^2
VRU-Summit	3.4×10^3	$2.6 \times 10^6, [1.6 \times 10^6, 4.6 \times 10^6]$	2.2×10^4	4.2×10^3	$1.0 \times 10^6, [5.9 \times 10^5, 2.1 \times 10^6]$	4.8×10^3
Business	1.5×10^3	$5.1 \times 10^5, [3.2 \times 10^5, 9.1 \times 10^5]$	9.6×10^3	2.0×10^3	$8.1 \times 10^4, [4.7 \times 10^4, 1.7 \times 10^5]$	8.2×10^2
Announcement	1.0×10^4	$1.7 \times 10^7, [1.1 \times 10^7, 3.1 \times 10^7]$	4.9×10^4	1.3×10^4	$3.1 \times 10^6, [1.8 \times 10^6, 6.5 \times 10^6]$	5.0×10^3
Message	2.6×10^3	$2.1 \times 10^6, [1.3 \times 10^6, 3.8 \times 10^6]$	2.3×10^4	3.6×10^3	$1.0 \times 10^5, [5.9 \times 10^4, 2.1 \times 10^5]$	5.7×10^2
Total	1.8×10^5	$3.2 \times 10^9, [2.0 \times 10^9, 5.8 \times 10^9]$	5.1×10^5	2.2×10^5	$1.6 \times 10^8, [9.3 \times 10^7, 3.3 \times 10^8]$	1.5×10^4

Figure 6 **Call center arrivals: average and hourly arrival rates for 5 Mondays.**

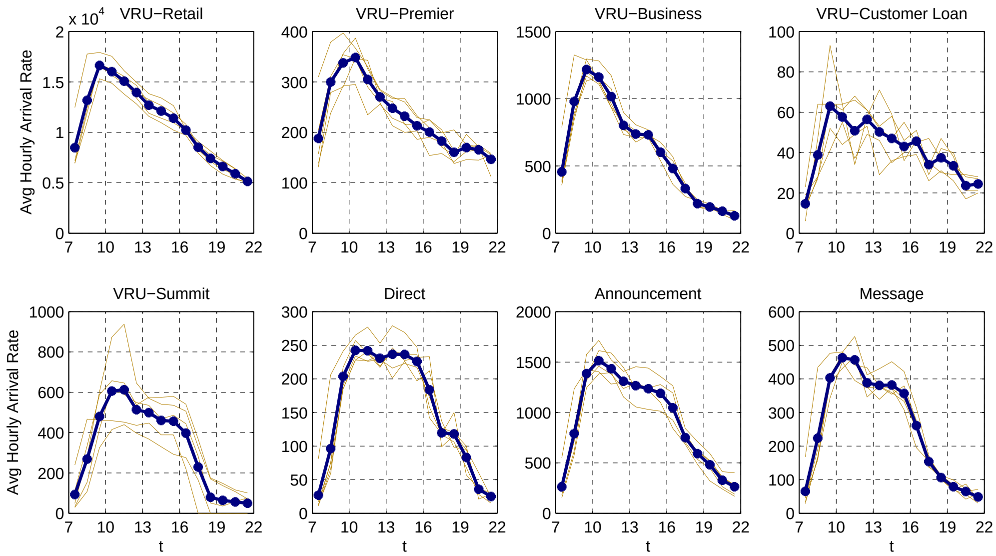
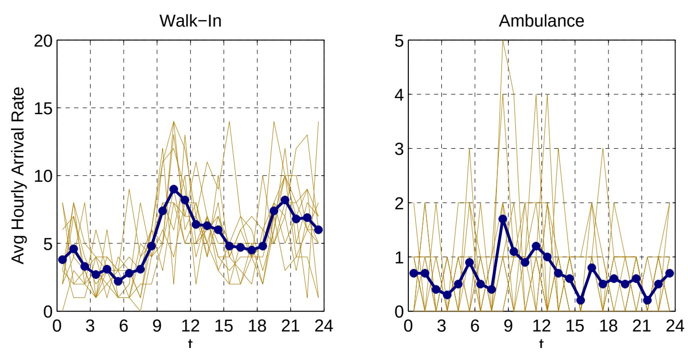


Table 8 **Hospital ED arrivals: summary statistics by type.**

	All days (n=70)				Weekdays (n=50)			
Type	$\bar{\mu}$	$\bar{\sigma}^2$	$\bar{D}$	p-value	$\bar{\mu}$	$\bar{\sigma}^2$	$\bar{D}$	p-value
Walk-in	121.9	615.8, [452.7, 886.6]	348.5	0.00	112.7	288.4, [201.3, 447.9]	125.4	0.00
Ambulance	16.6	19.2, [14.1, 27.7]	79.8	0.17	16.2	16.7, [11.7, 25.9]	50.5	0.42
Total	138.5	657.0, [483.0, 945.9]	327.2	0.00	129.0	334.7, [233.5, 519.7]	127.2	0.00

Figure 7 **Hospital ED arrivals: average and hourly arrival rates for 10 Mondays.**



5.2. Illustrative Examples

We consider three illustrative examples in this section. Suppose we are interested in testing for an NHPP in the following settings:

Case 1 We observe that the VRU - Summit arrival rate at the call center is nearly constant in the interval $[14, 15]$ (i.e., from 2pm to 3pm). We want to test whether the arrival process in $[14, 15]$ is a PP.

Case 2 We observe that the VRU - Summit arrival rate at the call center is nearly linear and increasing in the interval $[7, 10]$. We want to test whether the arrival process in $[7, 10]$ is an NHPP.

Case 3 We want to test whether the walk-in arrival process in the hospital ED data in the interval $[9, 12]$ is an NHPP.

Before proceeding to test whether the arrival data in each day come from a NHPP, we can first test whether there is over-dispersion over multiple days in the interval of interest (i.e., whether data from different days in the same interval have variable arrival rate). Table 9 provides the estimated values of mean (μ), variance σ^2 with its 95% confidence interval, and the dispersion test result for arrival data from 30 days for **Case 1** and **2** and from 70 days for **Case 3**. We observe that in three cases, there exist over-dispersion (in other words, day-to-day variation) in the arrival data. The table also illustrates how the day-of-week effect contributes to over-dispersion in each case. In the call center data, we still observe over-dispersion when the day-of-week is take into account. However, in the hospital ED data, the evidence of over-dispersion disappears in some days.

Table 9 Summary statistics for the samples of each case.

	Day of Week	# Sample	$\bar{\mu}$	$\bar{\sigma}^2$	$\bar{D}$	p-value
Case 1	Sun	5	460.6	10447.3, [3750.2, 86266.7]	90.7	0.00
	Mon	4	373.5	4083.0, [1310.3, 56762.1]	32.8	0.00
	Tues	4	371.5	11955.0, [3836.5, 166199.2]	96.5	0.00
	Wed	4	354.3	9830.3, [3154.6, 136660.8]	83.2	0.00
	Thurs	4	387.8	10204.3, [3274.7, 141860.1]	78.9	0.00
	Fri	4	139.0	2531.3, [812.3, 35190.8]	54.6	0.00
	Sat	5	81.6	270.3, [97.0, 2232.0]	13.3	0.01
	ALL	30	307.2	24583.1, [15592.2, 44426.2]	2320.9	0.00
Case 2	Sun	5	841.4	87927.8, [31562.6, 726048.2]	418.0	0.00
	Mon	4	919.8	48630.3, [15606.0, 676060.9]	158.6	0.00
	Tues	4	823.3	22344.9, [7170.7, 310640.5]	81.4	0.00
	Wed	4	769.3	1864.3, [598.3, 25916.9]	7.3	0.06
	Thurs	4	833.8	17912.3, [5748.2, 249017.3]	64.5	0.00
	Fri	4	516.3	29938.3, [9607.5, 416203.5]	174.0	0.00
	Sat	5	134.8	2222.7, [797.9, 18353.5]	66.0	0.00
	ALL	30	677.7	99470.7, [63090.7, 179761.8]	4256.7	0.00
Case 3	Sun	10	24.6	40.7, [19.3, 135.7]	14.9	0.09
	Mon	10	19.4	20.3, [9.6, 67.5]	9.4	0.40
	Tues	10	17.6	30.9, [14.6, 103.1]	15.8	0.07
	Wed	10	21.6	25.6, [12.1, 85.3]	10.7	0.30
	Thurs	10	17.9	5.9, [2.8, 19.6]	3.0	0.97
	Fri	10	20.2	56.8, [26.9, 189.5]	25.3	0.00
	Sat	10	31.3	50.2, [23.8, 167.4]	14.4	0.11
	ALL	70	21.8	50.2, [36.9, 72.3]	159.0	0.00

In tackling **Case 1**, we do not need subintervals because we observe that the arrival rate is nearly constant in the interval, so we can apply the Lewis test right away. Table 10 provides such test results under 'Before unrounding'. The average number of arrivals over 30 days was 307.2 ± 55.2. Average p-value with associated 95% confidence intervals and the number of days (out of 30 days) that passed each test at significance level $\alpha = 0.05$ are shown. We see that the arrival data pass the Lewis test on 19 days out of 30 days in April. We then find that the arrival times have been rounded to the nearest second. As discussed in Section 2 this can lead us to reject the NHPP null hypothesis more, so we *unround* the arrival data by adding uniform random variables divided by 3600 to the arrival times. The new results are under 'Unrounded' in Table 10; we see that now 29 days out of 30 days pass the Lewis test, and the average p-value has increased from 0.20 to 0.49. Thus we fail to reject the hypothesis that the arrival processes in $[14, 15]$ do come from NHPPs, with the understanding that the rates vary across different days, and should be regarded as random.

Table 10 Results for Case 1: The effect of rounding.

	Before unrounding		Unrounded	
Test	Avg p-value	# Pass	Avg p-value	# Pass
CU	0.54 ± 0.12	28	0.54 ± 0.12	28
Lewis	0.20 ± 0.08	19	0.49 ± 0.09	29

To answer our second question in **Case 2**, we first note that we want to use subintervals as discussed in §3 because the arrival rate is nearly linear and increasing in the interval $[7,10]$. Table 11 shows the result of using different subinterval lengths, L = 3, 1.5, 1, and 0.5 hours. The average number of arrivals over 30 days was 677.7 ± 111.1. We observe that more days pass the Lewis test as we decrease the subinterval lengths (and hence make the piecewise-constant approximation more appropriate in each subinterval). When we use L=0.5, all 30 days in April pass the Lewis test. We also see the importance of unrounding as discussed in **Case 1**; with L=0.5, only 18 days instead of 30 days pass the Lewis test when the arrival data are not unrounded.

Table 11 Results for Case 2: The role of subintervals.

		Before unrounding		Unrounded	
L (hours)	Test	Avg p-value	# Pass	Avg p-value	# Pass
3	CU	0.00 ± 0.00	0	0.00 ± 0.00	0
	Lewis	0.00 ± 0.01	1	0.04 ± 0.05	4
1.5	CU	0.02 ± 0.03	1	0.02 ± 0.03	1
	Lewis	0.09 ± 0.08	7	0.26 ± 0.11	18
1	CU	0.08 ± 0.04	12	0.08 ± 0.04	12
	Lewis	0.16 ± 0.08	15	0.48 ± 0.10	29
0.5	CU	0.23 ± 0.09	21	0.23 ± 0.10	21
	Lewis	0.20 ± 0.09	18	0.51 ± 0.10	30

Now we move on to **Case 3** in which we consider hospital ED walk-in arrival data. We find that the arrival times have been rounded to the nearest minute, so we unround them by adding uniform random variables divided by 60. We then apply the CU and the Lewis test to the arrival process in $[9,12]$ for each of the 70 days as in Table 12. The average number of arrivals over 70 days was 21.8 ± 1.7. When the arrival times are *unrounded*, 67 days out of 70 days pass the Lewis test. However, even when we use the raw arrival times (before unrounding), we observe that 66 days out of 70 days pass the Lewis test. The very small sample sizes evidently cause the rounding not to matter (Note that rounding matters when it produces 0 interarrival times, which do not occur in NHPPs). However, we note that the sample size is small (21.8 observations on average), suggesting that these tests have low power.

In order to obtain tests with greater power, we increase the sample size by combining data from multiple days. Table 13 shows the result of applying the CU and the Lewis tests to the seven groups of multiple days. We first apply the test assuming that the arrival rate is constant and the same in $[9,12]$ in all of the days and hence use subinterval length of $L = 30$, the entire interval. We see that four of the seven groups pass the Lewis test. However, as discussed in §4, there can be day-to-day variation which causes over-dispersion. So we use $L = 3$, which means one subinterval for each day. The test result improves and now six out of seven groups pass the Lewis test. We fail to reject our null hypothesis that the hospital emergency department walk-in arrivals in $[9,12]$ in our example are from NHPPs.

Table 12 Results for Case 3: Small sample size.

	Before unrounding		Unrounded	
Test	Avg p-value	# Pass	Avg p-value	# Pass
CU	0.43 ± 0.07	64	0.42 ± 0.07	64
Lewis	0.41 ± 0.07	66	0.48 ± 0.07	67

Table 13 Results for Case 3: Combining data over multiple days.

		Before unrounding				Unrounded			
		$L=30$		$L=3$		$L=30$		$L=3$	
Days	n	CU	$Lewis$	CU	$Lewis$	CU	$Lewis$	CU	$Lewis$
9/1/12-9/10/12	223	0.02	0.01	0.27	0.43	0.02	0.01	0.32	0.39
9/11/12-9/20/12	217	0.18	0.15	0.20	0.50	0.18	0.39	0.24	0.90
9/21/12-9/30/13	248	0.00	0.01	0.34	0.04	0.00	0.23	0.29	0.80
10/1/12-10/10/12	253	0.17	0.00	0.39	0.04	0.18	0.03	0.47	0.05
10/11/12-10/20/12	189	0.02	0.42	0.06	0.77	0.02	0.44	0.05	0.81
109/21/12-10/30/13	180	0.33	0.40	0.61	0.16	0.32	0.50	0.62	0.24
10/31/12-11/9/15	216	0.18	0.01	0.02	0.07	0.18	0.05	0.02	0.11
Average p-value	218	0.13	0.14	0.27	0.29	0.13	0.23	0.29	0.47
# Pass ($\alpha = 0.05$)		$4/7$	$3/7$	$6/7$	$5/7$	$4/7$	$5/7$	$6/7$	$6/7$

5.3. Call Center and Hospital Arrivals Can Be Well Modeled by NHPPs

Having examined what we need to account for in testing whether real data are from NHPPs, we now apply our tests to our call center data to examine whether the arrival process in each day (from 7 am to 10 pm, $[7, 22]$) is from a NHPP. Table 14 show the summarized result of applying the Lewis test using subinterval length L equal to one hour to unrounded arrival times (detailed results as well as CU test results can be found in the appendix). The average number of observations, average p-value with associated 95% confidence intervals and the number of days (out of 30 days) that passed each test at significance level $\alpha = 0.05$ are shown.

The results of the tests lead us to conclude that the arrival data from all these groups of customers are consistent with the NHPP hypothesis, with the possible exception of the VRU-Retail group. We conjecture that the greater tendency to reject the NHPP hypothesis for the VRU-Retail group is due to its much larger sample size. To test that conjecture, we reduce the sample size. We do so by further dividing the time intervals into 3-hour long subintervals. Table 14 shows that we are much less likely to reject the NHPP null hypothesis when we do this. Thus we tentatively conclude that the arrival process for each arrival type in the call center data is from a NHPP.

Table 14 Performance of the Lewis test applied to call center data.

	Avg # Obs	Avg p-value	# Pass
VRU-Retail	$1.4 \times 10^5 \pm 1.5 \times 10^4$	0.15 ± 0.09	11
VRU-Premier	$2.9 \times 10^3 \pm 2.6 \times 10^2$	0.49 ± 0.10	30
VRU-Business	$6.8 \times 10^3 \pm 1.2 \times 10^3$	0.49 ± 0.12	24
VRU-CL	$4.3 \times 10^2 \pm 5.7 \times 10^1$	0.44 ± 0.12	25
VRU-Summit	$3.3 \times 10^3 \pm 5.7 \times 10^2$	0.46 ± 0.10	28
Business	$1.5 \times 10^3 \pm 2.6 \times 10^2$	0.44 ± 0.12	25
Announcement	$9.7 \times 10^3 \pm 1.4 \times 10^3$	0.42 ± 0.13	22
Message	$2.6 \times 10^3 \pm 5.1 \times 10^2$	0.50 ± 0.11	30
VRU-Retail [7,10]	$3.5 \times 10^4 \pm 4.6 \times 10^3$	0.37 ± 0.12	22
VRU-Retail [10,13]	$3.7 \times 10^4 \pm 3.9 \times 10^3$	0.15 ± 0.08	13
VRU-Retail [13,16]	$2.8 \times 10^4 \pm 3.3 \times 10^3$	0.27 ± 0.11	20
VRU-Retail [16,19]	$2.1 \times 10^4 \pm 2.2 \times 10^3$	0.45 ± 0.11	27
VRU-Retail [19,22]	$1.4 \times 10^4 \pm 1.2 \times 10^3$	0.43 ± 0.11	27

Next, we test whether the arrivals in the hospital ED data are from NHPPs. When we apply the Lewis test with L=1 hr to the unrounded arrival data, we find that 68 and 65 days pass the test for walk-in and ambulance arrivals, respectively. Detailed test results are provided in the appendix. As reported in Table 8, the sample sizes are small; over the 70 days in the hospital ED arrival data, there are 121.9 walk-in and 16.6 ambulance arrivals on average. This can be a problem because it is much easier to pass the NHPP test with small sample size. Therefore, we also tried merging the arrival data by day of week. Because we consider data from 70 consecutive days, arrival data from 10 days are merged for each day of week. Table 15 provides the results; we observe that even with much bigger sample size, the arrival processes pass the test all the time, and we conclude that the hospital ED arrivals in our data are from NHPPs.

Table 15 Performance of alternative tests applied to hospital ED data.

			Before unrounding				Unrounded			
			$L=24$		$L=1$		$L=24$		$L=1$	
Type	Day of Week	n	CU	$Lewis$	CU	$Lewis$	CU	$Lewis$	CU	$Lewis$
Walk-In	Mon	1599	0.00	0.00	0.34	0.00	0.00	0.00	0.97	0.62
	Tues	1278	0.00	0.00	0.32	0.00	0.00	0.00	0.92	0.63
	Wed	1085	0.00	0.00	0.15	0.00	0.00	0.04	0.13	0.94
	Thurs	1063	0.00	0.00	0.58	0.00	0.00	0.02	0.68	0.36
	Fri	1122	0.00	0.00	0.03	0.00	0.00	0.00	0.03	0.54
	Sat	968	0.00	0.00	0.10	0.00	0.00	0.00	0.07	0.95
	Sun	1298	0.00	0.00	0.26	0.00	0.00	0.00	0.90	0.93
	Average	1201.9	0.00	0.00	0.25	0.00	0.00	0.01	0.53	0.71
	# Pass ($\alpha = 0.05$)		0/7	0/7	6/7	0/7	0/7	0/7	6/7	7/7
Ambulance	Mon	160	0.94	0.00	0.16	0.00	0.94	0.34	0.34	0.24
	Tues	162	0.08	0.00	0.19	0.00	0.08	0.69	0.16	0.88
	Wed	152	0.01	0.00	0.85	0.00	0.01	0.93	0.95	0.32
	Thurs	171	0.00	0.00	0.61	0.00	0.00	0.22	0.50	0.34
	Fri	169	0.03	0.00	0.00	0.00	0.03	0.71	0.01	0.28
	Sat	139	0.07	0.00	0.78	0.00	0.07	0.34	0.69	0.75
	Sun	192	0.15	0.00	0.35	0.00	0.15	0.46	0.48	0.08
	Average	163.6	0.18	0.00	0.42	0.00	0.18	0.53	0.45	0.41
	# Pass ($\alpha = 0.05$)		4/7	0/7	6/7	0/7	4/7	7/7	6/7	7/7

6. Conclusions

We examined call center and hospital arrival data and found that they are consistent with the NHPP hypothesis, i.e., that the KS tests of an NHPP applied to the data fail to reject that hypothesis, except that the rate should be regarded as random over different days in all cases except hospital ED ambulance arrivals; in that context, the test failed to reject the null hypothesis even for data from multiple days.

The analysis was not entirely straightforward. The majority of the paper was devoted to three issues that need to be addressed and showing how to do so. §2 discussed data rounding, showing that its impact can be successfully removed by unrounding. Consistent with Kim and Whitt (2013a), the Lewis test is highly sensitive to the rounding, while the CU KS test is not.

§3 discussed the problem of choosing subintervals so that the PC rate function approximation is justified. Simple practical guidelines were given for (i) evaluating any given subinterval in §3.4 and (ii) choosing an appropriate number of equally spaced subintervals in §3.6. Again consistent with Kim and Whitt (2013a), the CU KS test is more sensitive to the deviation from a constant rate function than the Lewis KS test.

Finally, §4 discussed the problem of over-dispersion caused by combining data from multiple days that do not have the same arrival rate.

References

Aksin, O. Z., M. Armony, V. Mehrotra. 2007. The modern call center: a multi-disciplinary perspective on operations management research. *Production Oper. Management* **16** 665–688.

Armony, M., S. Israelit, A. Mandelbaum, Y. Marmor, Y. Tseytlin, G. Yom-Tov. 2011. Patient flow in hospitals: a data-based queueing-science perspective. New York University, http://www.stern.nyu.edu/om/faculty/armony/.

Asaduzzaman, M., T. J. Chaussalet, N. J. Tobertson. 2010. A loss network model with overflow for capacity planning of a neonatal unit. *Annals of Operations Research* **178** 67–76.

Avramidis, A. N., A. Deslauriers, P. L'Ecuyer. 2004. Modeling daily arrivals to a telephone call center. *Management Sci.* **50** 896–908.

Barbour, A. D., L. Holst, S. Janson. 1992. *Poisson Approximation*. Oxford University Press, Oxford, U. K.

Bassamboo, A., A. Zeevi. 2009. On a data-driven method for staffing large call centers. *Operations Research* **57**(3) 714–726.

Brown, L., N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, L. Zhao. 2005. Statistical analysis of a telephone call center: a queueing-science perspective. *J. Amer. Stat. Assoc.* **100** 36–50.

Cooper, R. B. 1982. *Introduction to Queueing Theory*. 2nd ed. North Holland, Amsterdam.

Daley, D. J., D. Vere-Jones. 2008. *An Introduction to the Theory of Point Processes*, vol. II. 2nd ed. Springer, Oxford, U. K.

Durbin, J. 1961. Some methods for constructing exact tests. *Biometrika* **48**(1) 41–55.

Ibrahim, R., P. L'Ecuyer, N. Regnard, H. Shen. 2012. On the modeling and forecasting of call center arrivals. *Proceedings of the 2012 Winter Simulation Conference* **2012** 256–267.

Jongbloed, G., G. Koole. 2001. Managing uncertainty in call centers using Poisson mixtures. *Applied Stochastic Models in Business and Industry* **17** 307–318.

Kathirgamatamby, N. 1953. Note on the Poisson index of dispersion. *Biometrika* **40**(1) 225–228.

Kim, S.-H., W. Whitt. 2012. Statistical analysis with Little's law supplementary material: more on call center data. Columbia University, http://www.columbia.edu/∼ww2040/allpapers.html.

Kim, S.-H., W. Whitt. 2013a. Choosing arrival process models for service systems: Tests of a nonhomogeneous Poisson process. Columbia University, http://www.columbia.edu/∼ww2040/allpapers.html.

Kim, S.-H., W. Whitt. 2013b. Statistical analysis with Little's law. *Operations Research* **61**(4) 1030–1045.

Lewis, P. A. W. 1965. Some results on tests for Poisson processes. *Biometrika* **52**(1) 67–77.

Litvak, N., M. van Rijsbergen, R. J. Boucherie, M. van Houdenhoven. 2008. Managing the overflow of intensive care patients. *European Journal of Operational Research* **185** 998–1010.

Mandelbaum, A. 2012. Service Engineering of Stochastic Networks web page: http://iew3.technion.ac.il/serveng/.

Marsaglia, G., W. W. Tsang, J. Wang. 2003. Evaluating Kolmogorov's distribution. *Journal of Statistical Software* **8**(18) 1–4.

Massey, F. J. 1951. The Kolmgorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association* **461** 68–78.

Massey, W. A., G. A. Parker, W. Whitt. 1996. Estimating the parameters of a nonhomogeneous Poisson process with linear rate. *Telecommunication Systems* **5** 361–388.

Miller, L. H. 1956. Table of percentage points of Kolmogorov statistics. *Journal of the American Statistical Association* **51** 111–121.

Simard, R., P. L'Ecuyer. 2011. Computing the two-sided Kolmogorov-Smirnov distribution. *Journal of Statistical Software* **39**(11) 1–18.

Whitt, W. 1982. Approximating a point process by a renewal process: two basic methods. *Oper. Res.* **30** 125–147.

Wilkinson, R. 1956. Theories of toll traffic engineering in the U.S.A. *Bell System Technical Journal* **35** 421–514.