MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 16, No. 3, Summer 2014, pp. 464–480
ISSN 1523-4614 (print) | ISSN 1526-5498 (online)
http://dx.doi.org/10.1287/msom.2014.0490

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

Song-Hee Kim, Ward Whitt
Department of Industrial Engineering and Operations Research, Columbia University, New York, New York 10027
{sk3116@columbia.edu, ww2040@columbia.edu}

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 banking call center and hospital emergency department 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 account for three common features of arrival data: (i) data rounding, e.g., to seconds; (ii) choosing subintervals over which the rate varies too much; and (iii) overdispersion caused by combining data from fixed hours on a fixed day of the week over multiple weeks that do not have the same arrival rate. In this paper, we investigate how to address each of these three problems.

Keywords: arrival processes; nonhomogeneous Poisson process; Kolmogorov–Smirnov statistical test; data rounding; overdispersion
History: Received: June 17, 2013; accepted: March 12, 2014. Published online in *Articles in Advance* June 2, 2014.

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). Because 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 infrequently. Mathematical support is provided by the Poisson superposition theorem; e.g., see Barbour et al. (1992), §11.2 of Daley and Vere-Jones (2008) and §9.8 of Whitt (2002).

Nevertheless, there are phenomena that can prevent the Poisson property from occurring. For example, scheduled appointments 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 occurs in hospitals (Asaduzzaman et al. 2010, Litvak et al. 2008) and with requests for reservations at hotels, because the overflows tend to occur in clumps during the intervals when the first system is full. Indeed, there is a long history of studying overflow systems in teletraffic engineering (Cooper 1982, Wilkinson 1956). Bursty arrival processes also occur if the arrivals occur in batches, as in arrivals to hospitals after accidents. Restaurant arrivals occur in groups, but these groups usually can be regarded as single customers. In contrast, batch arrivals in hospitals typically use resources as individuals. From the extensive experience in teletraffic engineering, it is known that departures from the Poisson property can strongly impact performance; that is supported by recent work in Li and Whitt (2013), Pang and Whitt (2012). We emphasize this key point by showing the results of simulation experiments in §3 of the online supplement (available at http://dx.doi.org/10.1287/msom.2014.0490).

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 directly 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 with the *critical value*, $\delta(n, \alpha)$; the PP null hypothesis H_0 is rejected at significance level α if $D_n > \delta(n, \alpha)$, where $P(D_n > \delta(n, \alpha) \mid H_0) = \alpha$. 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 $n > 35$; see Simard and L'Ecuyer (2011) and the references therein.

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 helps 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, if the CU transformation is applied in that way (by combining the data over multiple days treated separately), then the statistical test should be regarded as a test of a Cox process, i.e., for a doubly stochastic PP, where the rate is random over the days but is constant over each subinterval over which the CU transformation is applied.

Indeed, even though we will not address that issue here, there is statistical evidence that the rate function often should be regarded as random over successive days, even for the same day of the week. 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 (2014) investigated why an additional data transformation is needed and what form it should take. They showed through large-sample asymptotic analysis and extensive simulation experiments that the CU KS test of a PP has remarkably little power against alternative processes with nonexponential 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 (2014) showed that the KS test used by Brown et al. (2005) has much greater power against alternative processes with nonexponential interarrival-time distributions. Kim and Whitt (2014) also found that Lewis (1965) had discovered a different data transformation in 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 with 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 reorder the interarrival times of the uniform random variables in ascending order. We display the full transformation in the online supplement.

Kim and Whitt (2014) 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 nonexponential 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 reordering of the interarrival times weakens the dependence. Hence, here we concentrate on the Lewis and CU KS tests. For applications, we recommend applying both of these KS tests.

1.4. Remaining Issues in Applications

Unfortunately, it does not suffice to simply perform these KS tests on 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 (2014) 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 unrounding, which involves adding i.i.d. small uniform random variables to the rounded data. In §2 we conduct simulation experiments showing that rounding a PP leads to rejecting the PP hypothesis and that unrounding avoids it. We also conduct experiments to verify that unrounding does not change a non-PP into a PP, provided that the rounding is not too coarse. If the KS test rejects the PP hypothesis before the rounding and unrounding, and if the rounding is not too coarse, then 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, nonconstant 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. (It is important to note that some fundamental smoothness in the arrival rate function is being assumed; see §3.7 for more discussion.) 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, the arrival rate may vary substantially over these time intervals over multiple weeks. We may fail to recognize that data from multiple weeks may in fact have variable arrival rates even though we look at the same time of day and the same day of the week. That is, there may be overdispersion in the arrival data. It is often not difficult to test for such overdispersion, 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 the references therein. If these tests do indeed find that there is such overdispersion, 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 random over different days (and thus a stochastic process). In §4 we review how to test for overdispersion in the arrival data.

After investigating those three causes for inappropriately rejecting the NHPP hypothesis in §§5 and 6, we illustrate these methods with call center and hospital emergency department arrival data. We draw conclusions in §7. Supporting material is provided in the online supplement (available at http://dx.doi.org/10.1287/msom.2014.0490) and in the online appendix (available at http://www.columbia.edu/~ww2040/allpapers.html).

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 (rounding to seconds). 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.

The rounding problem can be addressed by having accurate arrival data without rounding, but often that is not possible, e.g., because the rounding is done in the measurement process. Nevertheless, as observed by Brown et al. (2005), it is not difficult to address the rounding problem in a reasonable practical way by appropriately unrounding the rounded data. If the data has been rounded by truncating, then we can unround 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. It usually is straightforward to check if rounding has been done, and we would only unround to undo the rounding that we see.

2.1. The Need for Unrouding

To study rounding and unrounding, we conducted simulation experiments. We first simulated 1,000 replications of an NHPP with constant rate $\lambda = 1{,}000$ (an ordinary PP) on the interval $[0, 6]$, with time measured in hours, so that a mean interarrival time is 3.6 seconds. We then apply the CU KS test and the Lewis KS test, as described in Kim and Whitt (2014), to three versions of the simulated arrival data: (i) raw, as they are; (ii) rounded, rounded to the nearest second; and (iii) unrounded, in which we first round to the nearest second and then afterward unround by adding uniform random variables on $[0, 1]$ divided by 3,600 (since the units are hours and the rounding is to the nearest second) to the arrival times from (ii), as was suggested in Brown et al. (2005).

Table 1 summarizes the results of the 1,000 experiments. For each of the three forms of the data (raw, rounded, and unrounded) and two KS tests (CU and Lewis), we display the number of the 1,000 KS tests that fail to reject the PP hypothesis at significance level $\alpha = 0.05$ (#Pass), the average p-values (Ave p-value), and the average percentage of 0 values (Ave %0). The Lewis test consistently rejects the PP null hypothesis when the arrival data are rounded, but the CU KS test fails to do so. In fact, it is clearly appropriate to reject the PP null hypothesis when the data are rounded because the rounding produces too many 0-valued interarrival times. Table 1 shows that the rounding turns 12.7% of the interarrival times into 0.

Table 1 Results of the Two KS Tests with Rounding and Unrounding: Poisson Data

	CU			Lewis		
Type	#Pass	Ave p-value	Ave %0	#Pass	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

These test results illustrate the advantage of the Lewis KS test over the CU KS test. Because 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. As a consequence, the Lewis test strongly rejects the Poisson hypothesis when the data are rounded. The CU test looks at the data in order of the initial arrival times, so the 0 interarrival times are spread out throughout the data and are not detected by the CU KS test. 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.

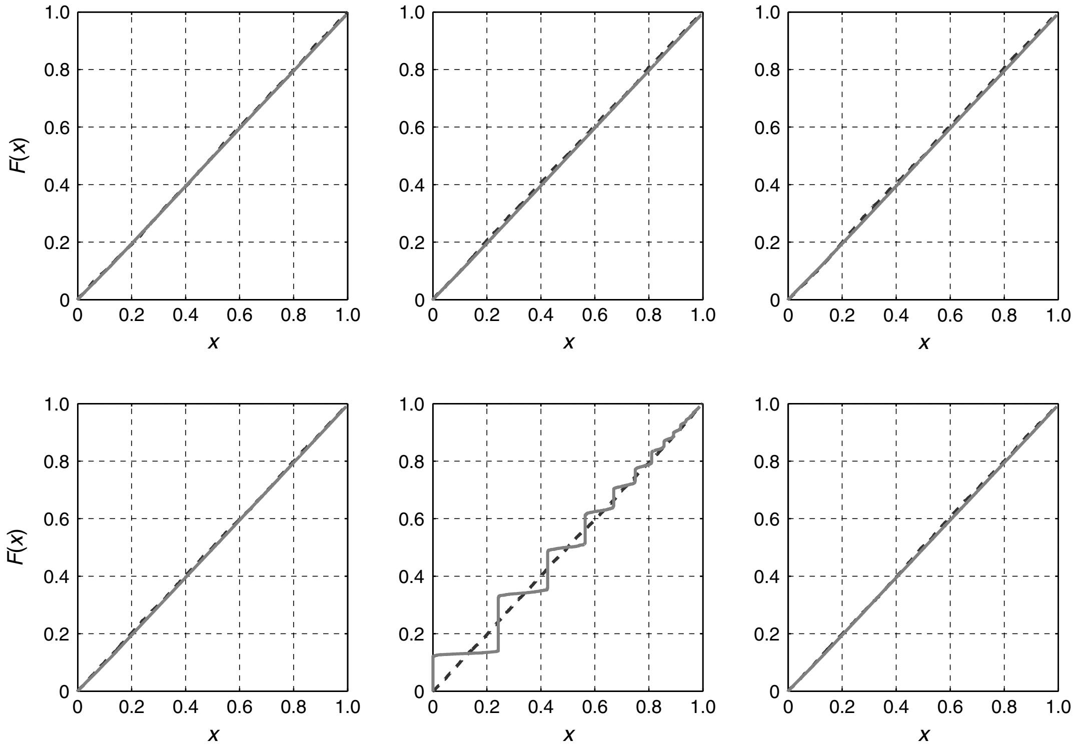
As in Kim and Whitt (2014), we find that plots of the empirical cdf's used in the KS tests are revealing. Figure 1 compares the average ecdf based on 100 replications of a rate-1,000 Poisson process on the time interval $[0, 6]$ with the cdf of the null hypothesis. We do not show the 95% confidence intervals for the average ecdf's as they overlap with the average ecdf's. From these plots, we clearly see that the Lewis test is effective, whereas the CU KS test fails to detect any problem at all.

2.2. The Possible Loss of Power

Evidently, the rounding and subsequent unrounding makes an arrival process more like an NHPP than it was before the rounding was performed. It is thus natural to wonder if the rounding and subsequent unrounding causes a serious loss of power. To examine that issue, we performed additional experiments. First, we simulated 1,000 replications of a renewal arrival process with a constant rate $\lambda = 1{,}000$ 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 (2014) shows that the Lewis test is usually able to detect this departure from the Poisson property and to reject the Poisson hypothesis.

Table 2 here 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 consistently rejects the Poisson hypothesis for the raw data, as it should, but the CU KS test fails to reject in 70% of the cases. 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. (The ecdf's from the CU test look similar to the ones provided in Figure 1.)

Figure 1 **Comparison of the Average ecdf for a Rate-1,000 Poisson Process**



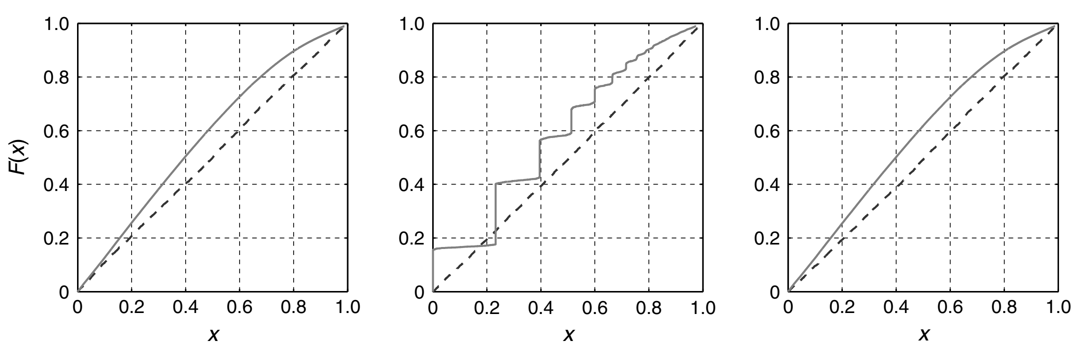
Notes. From top to bottom: CU, Lewis test. From left to right: raw, rounded, unrounded.

Table 2 **Results of the Two KS Tests with Rounding and Unrounding: H_2 Interarrival Times**

	CU			Lewis		
Type	#Pass	Ave *p*-value	Ave %0	#Pass	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
Unrounded	706	0.21	0.0	0	0.00	0.0

We also conducted other experiments of the same kind to show that the unrounding does not inappropriately cause the Lewis KS test to fail to reject a non-PP, provided that the unrounding is not done too coarsely. Among the more interesting cases are two forms of batch Poisson processes. The first is a rate-1,000 renewal process in which the interarrival times are 0 with a probability p and an exponential random variable with a probability $1-p$. The second is a modification of a PP in which every kth arrival occurs in batches of size 2; the arrival rate is reduced to $1{,}000k/(k+1)$, so the overall arrival rate is again 1,000. Assuming that the rounding is done to the nearest seconds as in the PP and H_2 examples above, the unrounding consistently detects the deviation from the PP when p is not too small (e.g., when $p \geq 0.05$) in the renewal process example and when k is not too large (e.g., when $k \leq 9$) in the second modification of a PP with batches.

Figure 2 **Comparison of the Average ecdf of a Rate-1,000 Arrival Process with H_2 Interarrival Times**



Notes. Lewis test only. From left to right: raw, rounded, unrounded.

As long as the rounding is not too coarse, the story for these examples is just like the H_2 renewal process we have already considered. The unrounding removes all 0 interarrival times, but it still leaves too many very short interarrival times, so the PP hypothesis is still rejected. The details appear in the online supplement.

On the other hand, if the rounding is too coarse, as in the batch-Poisson examples above when the rounding is to the nearest minute instead of the nearest second, then the unrounding *can* hide the non-PP character of the original process and thus reduce the power of the KS test. We also illustrate this phenomenon in the online supplement. Overall, rounding should not matter, so unrounding is unnecessary if the rounding is fine, e.g., to less than 0.01 mean service time, whereas there is a danger of a loss of power if the rounding is too coarse, e.g., to more than a mean service time. We recommend using a simulation to investigate in specific instances, as we have done here. See the online supplement and the online appendix for more discussion.

3. Choosing Subintervals with a Nearly Constant Rate

To exploit the CU property to conduct KS tests of an NHPP, we assume that the rate function is approximately PC. Because 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 PC approximation more likely to be valid, but interarrival times are necessarily truncated at boundary points and 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 PC 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 a finite number of linear pieces. Ways to fit linear arrival rate functions were studied in Massey et al. (1996), which can be extended to piecewise-linear arrival rate functions (e.g., by choosing roughly appropriate boundary times and applying the least-squares methods there over each subinterval with the endpoint values constrained). However, it usually should not be necessary to have a formal estimation procedure 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 (1) a constant approximation is appropriate for any given subinterval with a linear rate and (2) a PC 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 are 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 (2013a, b). For one 17-hour day, represented as $[6, 23]$ in hours, they produced the fitted arrival rate function

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

as shown in Figure 3 (taken from Kim and Whitt 2013a, b). This fitted arrival rate function is actually 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 a linear arrival rate function $\lambda(t) = 1{,}000t/3$ on the interval $[0, 6]$. The expected total number of arrivals

Figure 3 Fitted Piecewise-Linear Arrival Rate Function for the Arrivals at a Banking Call Center

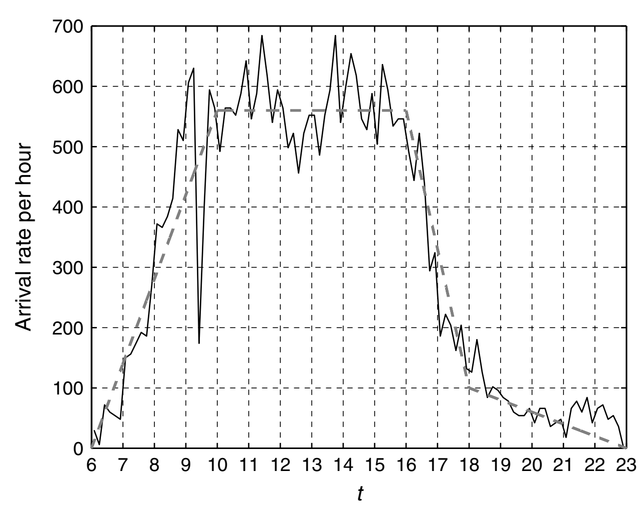


Table 3 Performance of the Alternative KS Test of an NHPP as a Function of the Subinterval Length L

		CU			Lewis		
L	Type	#Pass	Ave p-value	Ave %0	#Pass	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

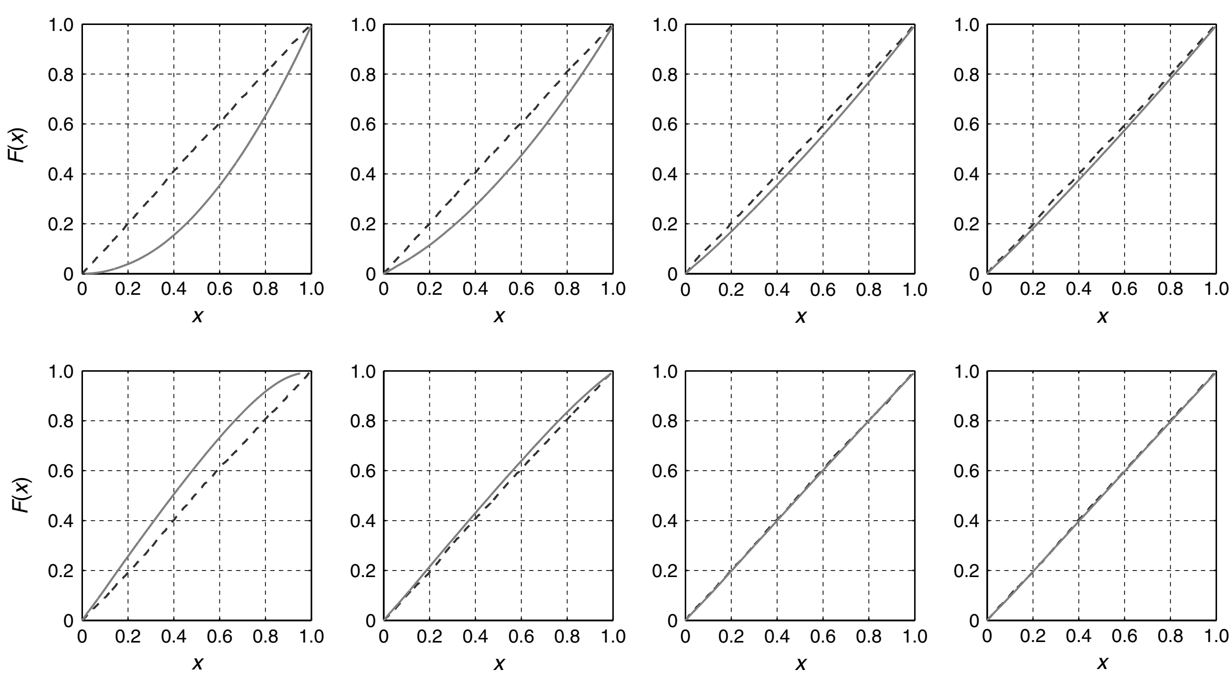
over this interval is 6,000. We apply simulation to study what happens when we divide the interval $[0, 6]$ into $6/L$ equally spaced disjoint subintervals, each of length L; apply the CU construction to each subinterval separately; and then afterward combine all the data from the subintervals.

Table 3 and Figure 4 show the performance of the Lewis and CU KS tests as a function of the subinterval length. As before, #Pass is the number of KS tests passed at significance level $\alpha = 0.05$ out of 1,000 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 nonconstant rate and *reject* the PP hypothesis with a 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.

Because 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 (2014), we see that the CU KS test is actually *more* effective in detecting this nonconstant rate than the Lewis test. The nonconstant 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 in §§3.3–3.6 below indicate that we could use $L = 0.5$ for the Lewis test.

In the remainder of this section, we develop a theory that shows how to construct PC approximations of the rate function. We then derive explicit formulas for the conditional cdf in three cases: (1) in general (which is complicated), (2) when the arrival rate is linear (which is relatively simple), and (3) when the data is obtained by combining data from equally spaced subintervals of a single interval with linear rate (which

Figure 4 Comparison of the Average ecdf of an NHPP with Different Subinterval Lengths



Notes. From top to bottom: CU, Lewis test. From left to right: $L = 6, 3, 1, 0.5$.

remains tractable). We then apply these results to determine when a PC 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 nonuniform 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 a finite number of 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 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 the 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 jth 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 $\{\bar{F}_n(t): 0 \leq t \leq 1\}$ is a stochastic process with

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

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

Theorem 2 (Limit for Empirical cdf). *Assuming a NHPP with an 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 by increasing the scaling constant c) to the conditional cdf F in (5), i.e.,*

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

We will usually omit the scaling constant in our discussion but do so with the understanding that it always can be introduced. Becasue 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

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

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.,

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

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 a 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* $[0, T]$,

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

where D_n is the CU KS test statistic in (10) and D is the degree of nonhomogeneity in (9) involving the 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 PC 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 a 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 we will assume that the arrival rate function is increasing, i.e., $b \ge 0$. There are two cases, $a > 0$ and $a = 0$, and we 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 a 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*

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

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 a linear rate over an interval should be close enough to a PP with a 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 shows 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 $\lambda(t) = 1{,}000t/3$ on the time interval $[0, 6]$. (The full table with all intervals and other examples appear in the online appendix.)

3.5. Subintervals for an NHPP with a 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 (2014), 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.,*

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

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

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

Table 4 **Judging When the Rate Is Approximately Constant: The Ratio $D/\delta(n,\alpha)$ for Single Subintervals with $\alpha = 0.05$**

							CU		Lewis	
L	Interval	Ave[n]	r	D	Ave[$\delta(n,\alpha)$]	D/Ave[$\delta(n,\alpha)$]	#Pass	Ave p-value	#Pass	Ave p-value
6	[0, 6]	5,997.3	∞	0.250	0.018	14.28	0	0.00	0	0.00
3	[0, 3]	1,498.8	∞	0.250	0.035	7.15	0	0.00	0	0.00
	[3, 6]	4,498.5	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]	1,166.9	0.33	0.036	0.040	0.90	300	0.08	931	0.49
	[4, 5]	1,501.0	0.25	0.028	0.035	0.79	358	0.09	949	0.49
	[5, 6]	1,830.6	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
	[1.5, 2]	292.0	0.67	0.036	0.079	0.45	766	0.29	945	0.50
	[2.5, 3]	456.9	0.40	0.023	0.063	0.36	833	0.35	960	0.51
	[3.5, 4]	623.3	0.29	0.017	0.054	0.31	865	0.38	938	0.51
	[4.5, 5]	792.1	0.22	0.013	0.048	0.27	882	0.41	936	0.50
	[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
	[1.75, 2]	156.5	0.57	0.017	0.108	0.15	924	0.48	940	0.49
	[2.75, 3]	238.7	0.36	0.011	0.087	0.12	931	0.48	956	0.50
	[3.75, 4]	322.0	0.27	0.008	0.075	0.11	941	0.47	946	0.50
	[4.75, 5]	406.7	0.21	0.006	0.067	0.10	937	0.48	953	0.50
	[5.75, 6]	489.2	0.17	0.005	0.061	0.09	941	0.50	943	0.50

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

$$\Lambda_j(t) = \frac{bt(kt + 2(j-1)T)}{2k}, \quad 0 \le t \le T/k,\ 1 \le j \le k,$$

$$F_j(t) = \frac{t(2j-2+t)}{2j-1}, \quad 0 \le t \le 1,\ 1 \le j \le k,$$

$$p_j = \frac{2j-1}{k^2} \quad \text{and} \quad r_j = \frac{k}{(j-1)T},\ 1 \le j \le k. \tag{19}$$

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), 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

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

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

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

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, if the rate is strictly positive on the interval (or if it is 0 at one endpoint), then we can apply formula (21) (respectively, (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

Table 5 Judging If a PC Approximation Is Good for an Interval Divided Into Equal Subintervals: The Ratio $D/\text{Ave}[\delta(n,\alpha)]$

			CU		Lewis	
L	D	$D/\text{Ave}[\delta(n,\alpha)]$	#Pass	Ave p-value	#Pass	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

size or the total expected number of arrivals, adding over all subintervals.

We illustrate in Table 5 by showing the values of D and $D/\delta(n,\alpha)$ along with the test results for each subinterval of the initial example with $\lambda(t)=1{,}000t/3$ on the time interval $[0,6]$, just as in Table 3. In all cases, Ave$[n]$ is 5,997.33, and hence the *ave*$[\delta(n,\alpha)]$ values are the same and are approximately 0.0175. (Again, more examples appear in the online appendix.)

In summary, we present the following algorithm for choosing an appropriate subinterval length to enable a PC approximation of a linear arrival rate over some interval.

1. Given an interval T with a fitted arrival rate function that is linear ($\lambda(t)=a+bt$), let n be the number of arrivals in that interval.
2. Compute the critical value of the KS test, $\delta(n,\alpha)$. It can be approximated as $\delta(n,\alpha)\approx 1.36/\sqrt{n}$ if $n>35$ and when we choose $\alpha=0.05$. (See Simard and L'Ecuyer 2011 and references therein for other values of n and α.)
3. Start with subinterval length $L=T/2$. Given L, compute *the degree of nonhomogeneity of the NHPP D* using (21) if $a>0$ and using (22) if $a=0$.
4. Compute $D/\delta(n,\alpha)$. Use bisection method to find the value of L that gives the ratio $D/\delta(n,\alpha)$ less than 0.10 (0.50) for the CU (Lewis) KS test.

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 a linear rate. We start with a nonconstant 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 a 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

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

To enable the sample size in each interval to also grow without bound, we also require that

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

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

Theorem 6 (Asymptotic Justification of the PC Approximation of Linear Arrival Rate Functions). *Suppose that we consider a nonconstant 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*

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

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 $\delta(n,\alpha)$ with 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 with $\sqrt{n}$ 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 where each piece is Lipschitz continuous.

4. Combining Data from Multiple Days: Possible Overdispersion

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 9 a.m. to 10 a.m. 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 (2014), we know that that sample size is sufficient.

In call centers, as in many other service systems, there is typically significant variation in the arrival rate over the hours of each day and even over different days of the week. It is thus common to estimate the arrival rate for each hour of the day and day of the week by looking at arrival data for specified hours and days of the week, using data from several successive weeks. The natural null hypothesis is that those counts over successive weeks are i.i.d. Poisson random variables. However, that null hypothesis should not be taken for granted. Indeed, experience indicates that there is often excessive variability over successive weeks. When that is found, we say that there is *overdispersion* in the arrival data.

In some cases, overdispersion can be explained by special holidays and/or seasonal trends in the arrival rate. The seasonal trends often can be identified by applying time-series methods. However, it can be the case that the observed overdispersion is far greater than can be explained in those systematic ways, as we will illustrate for the arrival data from a banking call center in §5.

4.1. Directly Testing for Overdispersion

In the spirit of the rest of this paper, we recommend directly testing whether or not there is overdispersion in arrival data. The null hypothesis is that the hourly arrival counts at fixed hours on fixed days of week over a succession of weeks constitute independent Poisson random variables with the same mean. A commonly used way to test if n observations $x_1, \ldots, x_n$ can be regarded as a sample from n i.i.d. Poisson random variables is the *dispersion test*, involving the statistic

$$\bar{D} \equiv \bar{D}_n \equiv \frac{(n-1)\bar{\sigma}_n^2}{\bar{x}_n} = \frac{\sum_{i=1}^n (x_i - \bar{x}_n)^2}{\bar{x}_n}, \quad \text{where}$$

$$\bar{\sigma}^2 \equiv \bar{\sigma}_n^2 \equiv \frac{\sum_{i=1}^n (x_i - \bar{x}_n)^2}{n-1} \quad \text{and} \quad \bar{x} \equiv \bar{x}_n \equiv \frac{\sum_{i=1}^n x_i}{n}; \qquad (26)$$

e.g., see Kathirgamatamby (1953). Because we are concerned with excessive variability, we consider the one-sided test and reject if $\bar{D}_n > \delta(n, \alpha)$ where $P(\bar{D}_n > \delta(n, \alpha) \mid H_0) = \alpha$, again using $\alpha = 0.05$. Under the null hypothesis, $\bar{D}_n$ is distributed as χ^2_{n-1}, a chi-squared random variable with $n-1$ degrees of freedom, which in turn is distributed as the sum of squares of $n-1$ standard normal random variables. Thus, under the null hypothesis, $E[\bar{D}_n \mid H_0] = n-1$, $\mathrm{Var}(\bar{D}_n \mid H_0) = 2(n-1)$ and $(\chi^2_n - n)/\sqrt{2n}$ converges to the standard normal as n increases. Thus, $\delta(n, 0.05) = \chi^2_{n-1, 0.95}$, the 95th percentile of the χ^2_{n-1} distribution.

See Brown and Zhao (2002) for a discussion and comparison of several tests of the Poisson hypothesis. The dispersion test above is called the conditional chi-squared test in §3.3 there; it is shown to perform well along with a new test that they introduce, which is based on the statistic

$$\bar{D}^{bz} \equiv \bar{D}^{bz}_n \equiv 4\sum_{i=1}^n (y_i - \bar{y}_n)^2 \quad \text{where } y_i \equiv \sqrt{x_i + (3/8)}, \qquad (27)$$

where $\bar{y}_n = n^{-1}\sum_{i=1}^n y_i$. Under the null hypothesis, $\bar{D}_n^{(bz)}$ is distributed as χ^2_{n-1} as well. We used both tests for overdispersion and found that the results were similar, so we only discuss $\bar{D}_n$ in (26); see the online appendix for details.

4.2. Avoiding Overdispersion and Testing for It with KS Tests

An attractive feature of the KS tests based on the CU property is that we can avoid the overdispersion problem while testing for an NHPP. We can avoid the overdispersion 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. Of course, when we apply KS tests based on the CU property in that 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 over any given day.

One way to test for such overdispersion is to conduct the KS test based on the CU property by combining data from multiple days, by *both* (1) combining all the data before applying the CU property and (2) applying the CU property to each day separately and then combining the data afterward. 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 overdispersion, then the KS test will reject the NHPP hypothesis if all the data is combined before applying the CU property. By conducting both KS tests of a PP, we can distinguish among three alternatives: (1) PP with a fixed rate, (2) PP with a random rate, and (3) neither of those.

5. Banking Call Center Arrival Data

We now consider arrival data from service systems, first a banking call center and then, in the next section, a hospital emergency department. We use the same call center data used in Kim and Whitt (2013a, b) 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 900–1,200 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 and with the rest being served by voice response unit (VRU) technology. In this study, we focus on arrival data during April 2001. There are four 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 do not include 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%.

5.1. Variation in the Arrival Rate Function

Figure 5 shows average hourly arrival rate and individual hourly arrival rate for each arrival type on Mondays. As usual, Figure 5 shows strong within-day variation. The variation over days of the week and over successive weeks in this call center data can be visualized by looking at four plots shown in Figure 6. The first plot on the left shows the average hourly arrival rate (per minute) over 18 weekdays (solid line) along with the daily average (horizontal dashed line). The average for each hour is the average of the arrival counts over that hour divided by 60 to get the average arrival rate per minute. The first plot also shows 95% confidence intervals about the hourly averages, which are wide in the middle of the day.

Part of the variability seen in the first plot can be attributed to the day-of-the-week effect. This is shown by the second plot, which displays the hourly averages for the five weekdays. From this second plot, we see that the arrival rates on Mondays are the highest, followed by Tuesdays and then the others. Finally, the third plot focuses directly on the overdispersion by displaying the hourly average rates by a specific day of the week, three Mondays. Even when restricting attention to a single day of the week, we see considerable variation.

However, we have yet to quantify the variation. When we applied the dispersion test to the call center data, we found overwhelming evidence of overdispersion. That is illustrated by the test results for 16 hours on 16 Fridays. Because the sample size for each hour was $n = 16$, $E[\bar{D}_n \mid H_0] = 15$, and $\mathrm{Var}(\bar{D}_n \mid H_0) = 30$. The 95th and 99th percentiles of the χ^2_{15} distribution are, respectively, 25.0 and 30.6. However, the 16 observed values of $\bar{D}_n$ corresponding to the 16 hours on these Fridays ranged from 163.4 to 1,068.7, with an average of 356. The average value of $\bar{D}_n$ exceeds the 99th percentile of the chi-squared distribution by a factor of 10. Moreover, the sample sizes were not small; the average hourly counts ranged from 29 to 503.

Figure 5 Call Center Arrivals: Average and Hourly Arrival Rates for Five Mondays

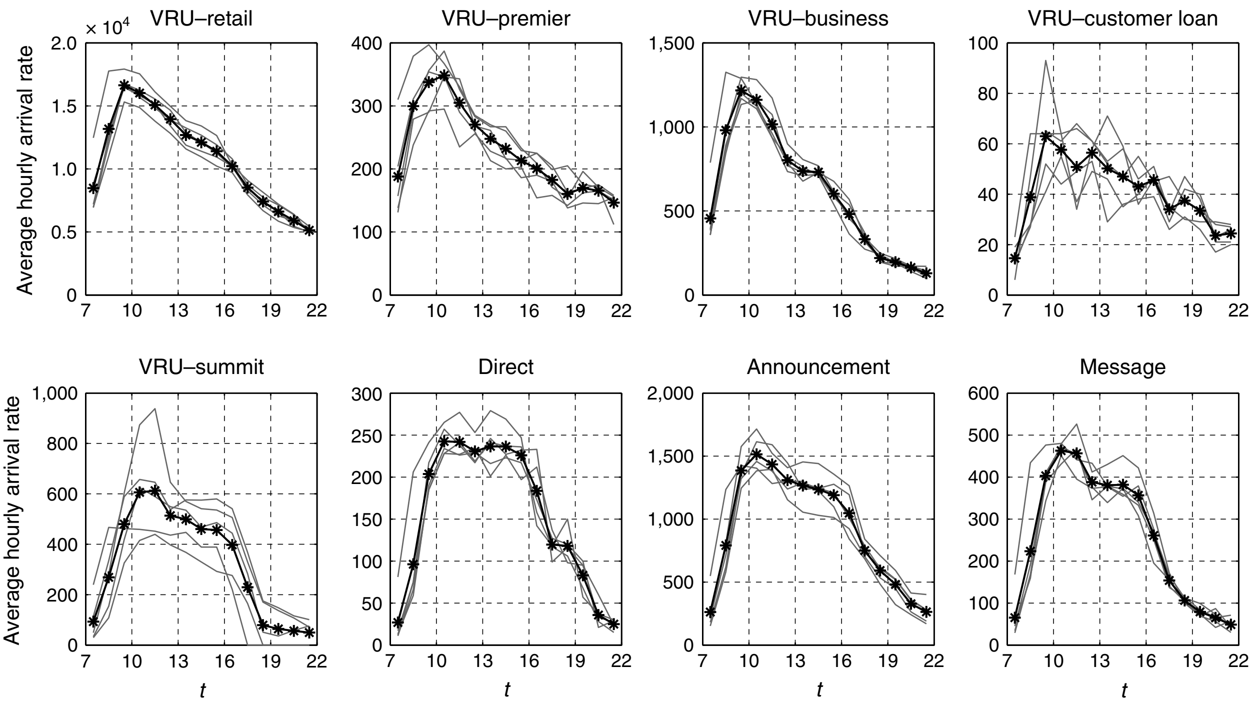
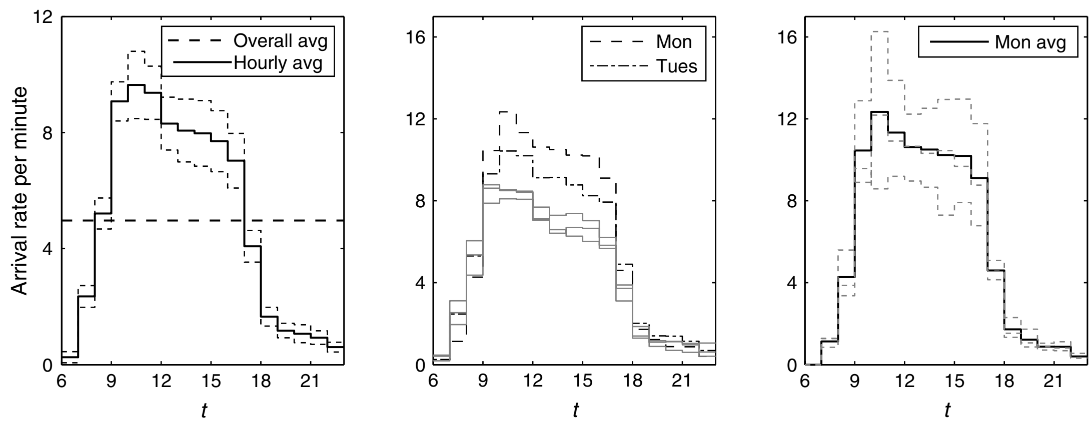


Figure 6 Average Arrival Rates



Note. From left to right: Overall and hourly average arrival rates for 18 weekdays, average arrival rates by each day of week, and arrival rates on three Mondays with its average.

5.2. One Interval with a Nearly Constant Arrival Rate

Figure 5 shows that the VRU-summit arrival rate at the call center is nearly constant in the interval [14, 15] (i.e., from 2 P.M. to 3 P.M.). We want to test whether the arrival process in [14, 15] can be regarded as a PP. Consistent with the observations above, we see that there is a strong day-of-the-week effect. When we applied the dispersion test to all 30 days, we obtained $\bar{D}_n = 2{,}320$, whereas $\delta(30, 0.05) = 43.8$. When we considered individual days of the week, we had four samples for each weekday and five for each weekend day. For Saturday we had $\bar{D}_n = 13.3$, while $\delta(5, 0.05) = 9.5$ and $\delta(5, 0.01) = 13.3$, showing that the p-value is 0.01, but in the other cases the $\bar{D}_n$ values ranged from 32.8 to 90.7, so the arrival data for the time interval [14, 15] on a fixed day of the week exhibits strong overdispersion.

Because the arrival rate is approximately constant over [14, 15], we do not need to consider subintervals to have a PC rate approximation. We can directly test for the PP, treating the data from separate days separately (and thus avoiding the day-of-the-week effect and the overdispersion over successive weeks). First, we note that the arrival data were rounded to the nearest second. The results of the CU and Lewis KS tests, with and without unrounding, are shown in Table 6. This table shows that the Lewis test fails to reject the PP hypothesis in 29 of 30 cases after unrounding, but in only 19 before unrounding. Just as in §2, the CU KS test fails to detect any problem caused by the rounding. Except for the overdispersion, this analysis supports the PP hypothesis for the arrival data in the single interval [14, 15].

Table 6 Results of KS Tests of PP for the Interval [14, 15]

	Before unrounding		Unrounded	
Test	Ave p-value	#Pass	Ave 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

5.3. One Interval with Increasing Arrival Rate

Figure 5 shows 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] can be regarded as an NHPP.

Just as in the previous example, we see that there is a strong day-of-the-week effect. When we applied to dispersion test to all 30 days, we obtained $\bar{D}_n = 4{,}257$, whereas $\delta(30, 0.05) = 43.8$. When we considered individual days of the week, we again had four samples for each weekday and five for each weekend day. For Wednesday we had $\bar{D}_n = 7.3$, while $\delta(4, 0.05) = 7.8$ and $\delta(4, 0.01) = 11.3$, and the p-value is 0.06, but for the other days of the week the $\bar{D}_n$ values ranged from 64.5 to 418, so the arrival data for the time interval [7, 10] on a fixed day of the week exhibits strong overdispersion.

The arrival rate is nearly linear and increasing over [7, 10], so we need to use subintervals, as discussed in §3. Table 7 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 PC 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; with $L = 0.5$, only 18 days instead of 30 days pass the Lewis test when the arrival data are not unrounded.

5.4. The KS Test of All the Call Center Arrival Data

We now consider all the call center arrival data. Table 8 shows the result of applying the Lewis test to all the

Table 7 Results of KS Tests of NHPP for the Interval [7, 10]

		Before unrounding		Unrounded	
L (hours)	Test	Ave p-value	#Pass	Ave 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

Table 8 Lewis KS Test Applied to the Call Center Data by Type with $L = 1$ and Unrounding

	Average no. of obs.	Ave 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

call center data by call type 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 online appendix.) We avoid the overdispersion by applying the CU transformation to all hours separately and then combining the data. 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 8 shows that we are much less likely to reject the NHPP null hypothesis when we do this.

In conclusion, we find significant overdispersion in the call center data—i.e., variation over successive weeks in counts for the same hour on the same day of the week. Otherwise, we conclude that the arrival data is consistent with the NHPP hypothesis. However, failure to reject the NHPP null hypothesis depends critically on (1) unrounding, (2) properly choosing subintervals over which the rate can be regarded as approximately constant, and (3) avoiding the overdispersion by applying the CU transformation to different hours separately.

6. Hospital Emergency Department Arrival Data

The emergency department (ED) arrival 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. Figure 7 shows the average hourly arrival rates for each arrival type on 10 Mondays. We observe less within-day variation among the ED arrivals than among the call center arrivals.

Figure 7 Hospital ED Arrivals: Average and Hourly Arrival Rates for 10 Mondays

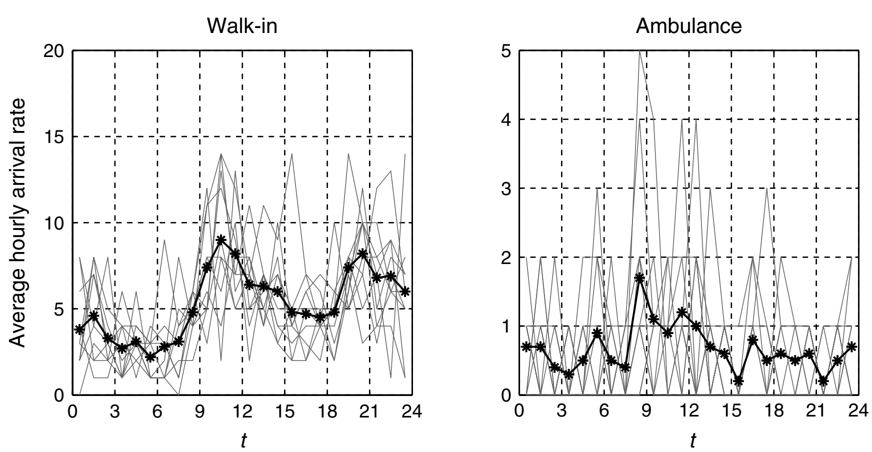


Table 9 KS Tests of NHPP for the Hospital ED Data

			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	1,599	0.00	0.00	0.34	0.00	0.00	0.00	0.97	0.62
	Tues	1,278	0.00	0.00	0.32	0.00	0.00	0.00	0.92	0.63
	Wed	1,085	0.00	0.00	0.15	0.00	0.00	0.04	0.13	0.94
	Thurs	1,063	0.00	0.00	0.58	0.00	0.00	0.02	0.68	0.36
	Fri	1,122	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	1,298	0.00	0.00	0.26	0.00	0.00	0.00	0.90	0.93
	Average	1,201.9	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

We first apply the dispersion test to test the Poisson hypothesis for daily counts for all days ($n = 70$) and all weekdays ($n = 50$), for all arrivals and by the two types. We can compare the dispersion statistic $\bar{D}$ values to $\chi^2_{n-1,1-\alpha}$ values for each (n, α) pair: $(70, 0.05)$: 89.4, and $(50, 0.05)$: 66.3. The dispersion test rejects the Poisson hypothesis for the walk-in arrivals and the daily totals, with $448 \le \bar{D}_n \le 520$ in the 4 cases, but it does not reject for the ambulance arrivals, with $\bar{D}_n = 79.8$ for $n = 70$ (p-value 0.17) and $\bar{D}_n = 50.5$ for $n = 50$ (p-value 0.42).

However, in the analysis above we have not yet considered the day-of-the-week effect. When we analyze the walk-in arrivals by day of the week, we obtain $n = 10$ and $\chi^2_{n-1,1-\alpha} = 16.9$ for $(n, \alpha) = (10, 0.05)$. The observed daily values of $\bar{D}_n$ on the 7 days of the week, starting with Sunday, were 14.9, 9.4, 15.8, 10.7, 3.0, 25.3, and 14.4. Hence, we would reject the Poisson hypothesis only on the single day Friday. The associated p-values were 0.09, 0.40, 0.07, 0.30, 0.97, 0.00, and 0.11. While we might want to examine Fridays more closely, we tentatively conclude that there is no overdispersion in the ED arrival data. We do not reject the Poisson hypothesis for ambulance arrivals on all days and walk-in arrivals by day of the week.

Next, we apply the CU and Lewis KS tests of an NHPP to the ED arrival data. First, based on the dispersion test results, we combine the data over the 10 weeks for each type and day of the week. We consider two cases for L: $L = 24$ (the entire day) and $L = 1$ using single hours as subintervals. Table 9 shows that, with unrounding and subintervals of length $L = 1$, the Lewis test never rejects the PP hypothesis, while the CU test rejects only once (Fridays). As before, using unrounding and subintervals is critical to these conclusions.

7. 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 significant overdispersion was found in the call center data. In particular, (1) variation in the arrival rate over the hours of each day was strong for the call center data and significant for the ED data; (2) variation in the arrival rate over the days of the week was significant for both the call center and ED data, except for ambulance arrivals; and (3) variation in the arrival rate over successive weeks for the same time of day and day of week (overdispersion) was significant for the call center data but not the ED data.

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. Section 2 discussed data rounding, showing that its impact can be successfully removed by unrounding. Consistent with Kim and Whitt (2014), the Lewis test is highly sensitive to the rounding, while the CU KS test is not. Section 3 discussed the problem of choosing subintervals so that the PC rate function approximation is justified. Simple practical guidelines were given for (1) evaluating any given subinterval in §3.4 and (2) choosing an appropriate number of equally spaced subintervals in §3.6. Again consistent with Kim and Whitt (2014), 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 overdispersion caused by combining data from multiple days that do not have the same arrival rate. These three issues played an important role for both sets of data.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/msom.2014.0490.

Acknowledgments

The authors thank Avishai Mandelbaum, Galit Yom-Tov, Ella Nadjharov, and the Center for Service Enterprise Engineering (SEE) at Technion–Israel Institute of Technology for access to the SEE call center data and advice about its use. They also thank Won-Chul Cha and the Samsung Medical Center in South Korea for providing the hospital emergency department data and advice about its use. Lastly, they thank the Samsung Foundation and the National Science Foundation for support [Grants CMMI 1066372 and 1265070].

References

Aksin OZ, Armony M, Mehrotra V (2007) The modern call center: A multi-disciplinary perspective on operations management research. *Production Oper. Management* 16(6):665–688.

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

Asaduzzaman M, Chaussalet TJ, Tobertson NJ (2010) A loss network model with overflow for capacity planning of a neonatal unit. *Ann. Oper. Res.* 178:67–76.

Avramidis AN, Deslauriers A, L'Ecuyer P (2004) Modeling daily arrivals to a telephone call center. *Management Sci.* 50(7):896–908.

Barbour AD, Holst L, Janson S (1992) *Poisson Approximation* (Oxford University Press, Oxford, UK).

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

Brown L, Gans N, Mandelbaum A, Sakov A, Shen H, Zeltyn S, Zhao L (2005) Statistical analysis of a telephone call center: A queueing-science perspective. *J. Amer. Statist. Assoc.* 100(469):36–50.

Brown L, Zhao LH (2002) A test for the Poisson distribution. *Sankhya: Indian J. Statist.* 64(A-3):611–625.

Cooper RB (1982) *Introduction to Queueing Theory*, 2nd ed. (North-Holland, Amsterdam).

Daley DJ, Vere-Jones D (2008) *An Introduction to the Theory of Point Processes*, Vol. II, 2nd ed. (Springer, Oxford, UK).

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

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

Jongbloed G, Koole G (2001) Managing uncertainty in call centers using Poisson mixtures. *Appl. Stochastic Models Bus. Indust.* 17(4):307–318.

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

Kim S-H, Whitt W (2013a) Statistical analysis with Little's law. *Oper. Res.* 61(4):1030–1045.

Kim S-H, Whitt W (2013b) Statistical analysis with Little's law, supplementary material: More on call center data. Columbia University, New York. http://www.columbia.edu/~ww2040/allpapers.html.

Kim S-H, Whitt W (2014) Choosing arrival process models for service systems: Tests of a nonhomogeneous Poisson process. *Naval Res. Logist.* 61(1):66–90.

Lewis PAW (1965) Some results on tests for Poisson processes. *Biometrika* 52(1):67–77.

Li AA, Whitt W (2013) Approximate blocking probabilities for loss models with independence and distribution assumptions relaxed. *Performance Evaluation*, ePub ahead of print August 13, http://dx.doi.org/10.1016/j.peva.2013.08.004.

Litvak N, van Rijsbergen M, Boucherie RJ, van Houdenhoven M (2008) Managing the overflow of intensive care patients. *Eur. J. Oper. Res.* 185(3):998–1010.

Mandelbaum A (2012) Service engineering of stochastic networks webpage. Last accessed May 27, 2014, http://iew3.technion.ac.il/serveng/.

Marsaglia G, Tsang WW, Wang J (2003) Evaluating Kolmogorov's distribution. *J. Statist. Software* 8(18):1–4.

Massey FJ Jr (1951) The Kolmgorov-Smirnov test for goodness of fit. *J. Amer. Statist. Assoc.* 46(253):68–78.

Massey WA, Parker GA, Whitt W (1996) Estimating the parameters of a nonhomogeneous Poisson process with linear rate. *Telecomm. Systems* 5(253):361–388.

Miller LH (1956) Table of percentage points of Kolmogorov statistics. *J. Amer. Statist. Assoc.* 51(273):111–121.

Pang G, Whitt W (2012) The impact of dependent service times on large-scale service systems. *Manufacturing Service Oper. Management* 14(2):262–278.

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

Whitt W (1982) Approximating a point process by a renewal process, I: Two basic methods. *Oper. Res.* 30(1):125–147.

Whitt W (2002) *Stochastic-Process Limits* (Springer, New York).

Wilkinson R (1956) Theories of toll traffic engineering in the U.S.A. *Bell System Technical J.* 35(2):421–514.