Operations Research Letters 43 (2015) 247–253

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Poisson and non-Poisson properties in appointment-generated arrival processes: The case of an endocrinology clinic

Song-Hee Kim [a], Ponni Vel [b], Ward Whitt [b,*], Won Chul Cha [c]

[a] Yale School of Management, Yale University, USA

[b] Industrial Engineering and Operations Research, Columbia University, USA

[c] Department of Emergency Medicine, Samsung Medical Center, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 26 December 2014
Received in revised form
21 February 2015
Accepted 22 February 2015
Available online 2 March 2015

Keywords:
Fitting queueing models to data
Queues with scheduled arrivals
Appointments
Statistical tests
Poisson processes
Dispersion

ABSTRACT

Previous statistical tests showed that call center arrival data were consistent with a non-homogeneous Poisson process (NHPP) within each day, but exhibit over-dispersion over multiple days. These tests are applied to arrival data from an endocrinology clinic, where arrivals are by appointment. The clinic data are also consistent with an NHPP within each day, but exhibit under-dispersion over multiple days. This analysis supports a new Gaussian-uniform arrival process model, with Gaussian daily totals and uniformly distributed arrivals given the totals.

1. Introduction

When building stochastic models to help improve the performance of service systems, it is important to have an appropriate arrival process model. Since the arrival rate typically varies strongly over the day, the most common arrival process model is a nonhomogeneous Poisson process (NHPP). The Poisson property is mathematically supported when arrivals come from the independent decisions of many different users who use the service system infrequently [2].

There is growing interest in testing the usual NHPP assumption for arrival processes [1,3,7,8,13,12,21]. Kim and Whitt [12] applied statistical tests to call center arrival data and found that (i) the data are consistent with an NHPP within each day, but (ii) the daily totals are more variable than Poisson; i.e., there is significant *over-dispersion* over multiple days. Fig. 1 shows the arrival counts over half hours. A casual glance shows no problem, but careful analysis exposes the over-dispersion: The number of arrivals in each half-hour interval is vastly different on five different Mondays on the same month. In the interval [11, 11.5], the sample mean number of arrivals is 317.8, with sample variance 12699.2 and variance-to-mean ratio 40.0. All of the half-hour intervals have variance-to-mean ratios greater than 1, with minimum of 5.8 in the interval [13, 13.5].

In this paper, we apply the statistical tests in [13,12] to arrival data from an endocrinology clinic, where all arrivals are by appointment for individual doctors. Despite the strongly deterministic framework, we show that, because of (i) randomness in the schedule, (ii) patient no-shows and (ii) early/late arrivals, the actual arrivals are distributed approximately as a Poisson process (PP, NHPP with constant rate) within each shift. However, the variance of the daily totals is significantly less than would be the case for Poisson random variables; i.e., we provide evidence of *under-dispersion* over multiple days. Based on this analysis, we propose a new two-time-scale Gaussian-uniform arrival process model for long-term planning for appointment-generated arrivals (which is to be examined in future work).

We note that there is extensive literature on appointment scheduling; see [4,6] for detailed reviews. While most of the early models assume a simple deterministic arrival pattern, new models are increasingly incorporating no-shows and non-punctuality, e.g., see [16,9] and references therein. There are also studies that show empirical evidence of patient no-shows and non-punctual arrivals. The estimated no-show rates vary across different services and

* Correspondence to: MailCode 4704, S. W. Mudd Building, 500 West 120th Street, New York, NY 10027-6699, USA.

E-mail addresses: hailey.kim@yale.edu (S.-H. Kim), pvv2001@columbia.edu (P. Vel), ww2040@columbia.edu (W. Whitt), docchaster@gmail.com (W.C. Cha).

http://dx.doi.org/10.1016/j.orl.2015.02.008
0167-6377/

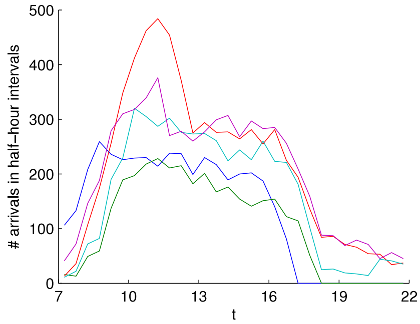

Fig. 1. Arrival counts in half-hour intervals at a call center on five Mondays during April 2001 (see [12] for details; VRI-Summit type arrival).

patient populations; the reported no-show rates are as low as 4.2% at a general practice outpatient clinic in United Kingdom [18] and as high as 31% at a family practice clinic [17].

Here is how the rest of this paper is organized: In Section 2 we introduce our study data from an endocrinology outpatient clinic. In Section 3 we compare scheduled arrivals and actual arrivals, show the presence of no-shows and early and late arrivals, and conduct statistical tests that show the arrivals are consistent with a PP within shifts. In Section 4 we statistically substantiate under-dispersion over multiple days. In Section 5 we propose stochastic arrival process models based on our data analysis.

2. The study data

The appointment arrival data are from an endocrinology outpatient clinic of a major teaching hospital in South Korea, collected over a 13-week period from July 2013 to September 2013. Sixteen doctors work in this clinic and patients arrive to the clinic knowing which doctor they will meet; hence, each doctor operates as a single-server system. Each doctor works in a subset of available days and shifts. There are three shifts: morning (am) shifts, roughly from 8:30 am to 12:30 pm, afternoon (pm) shifts, roughly from 12:30 pm to 4:30 pm, and full-day shifts. During the weekdays of the 13-week study period, the 16 doctors worked for a total of 228 am shifts, 220 pm shifts, 25 full-day shifts. The shifts are not evenly distributed among the doctors; the numbers ranged from 11 to 46.

In this paper, we primarily focus on patient arrivals to one doctor, called doctor 9 in our longer more detailed study [11]; doctor 9 was selected because of the relatively high volume and even distribution between the am and the pm shifts. Analysis of all doctors is in [11]. During our study period, doctor 9 worked for a total of 22 am shifts (12 on Tuesdays and 10 on Fridays) and 22 pm shifts (11 on Mondays, 2 on Wednesdays, and 9 on Thursdays).

We first consider the number of daily scheduled and actual arrivals. Patients make appointments for a specific time slot (available in 10 min intervals and each slot can have multiple patients). The schedule fills up over time (cancellations are allowed), and we see in the data that patients book appointments as early as a year before the appointment date. In this paper, we do not consider the booking date and examine only whether each patient has an appointment at the end of the previous day. We then differentiate between the number of scheduled (scheduled by the night before) patients (N_S) and the number of unscheduled (scheduled and arrived on the same day) patients (N_U). The number of patients who show up on their appointment date (N_A) is always less than or equal to the sum of N_S and N_U.

Fig. 2 depicts the values of N_S, N_U, and N_A during the 13-week study period. The average (standard deviation) values of N_S, N_U, and N_A are 66.1 (4.6), 2.2 (1.7), and 62.6 (4.2), respectively, in am shifts and 58.8 (6.0), 2.1 (1.7), and 55.7 (7.0), respectively, in pm shifts. Note that N_U is so small relative to N_S and N_A that N_U necessarily has a small impact on N_A. Also, note that N_S and N_A have low variability; we discuss and statistically test their under-dispersion in Section 4. On average, N_A is 95% of N_S in both the am and pm shifts; in particular, N_A ranges from 88% to 102% of N_S in am shifts and from 86% to 110% in pm shifts, and rarely exceeds N_S.

3. Arrivals within each shift

We now examine the arrival data within each shift (am or pm) on a single day. We start by estimating the cumulative arrival rate and instantaneous arrival rate functions for both the scheduled and actual arrivals. We then analyze no-shows and the lateness (or earliness), which explain why the actual arrival process is more variable than the scheduled arrival process. Afterwards, we test whether the arrival data within shifts are consistent with an NHPP or even a PP.

3.1. Estimated arrival rate functions

Patients are scheduled to arrive in 10-min intervals over each shift. Since about 66 patients arrive in each shift, each slot has on average 2.6 patients scheduled. Let $S(t)$ ($A(t)$) be the numbers of patients within a shift scheduled to arrive (that actually arrive) by time t, starting from the beginning of the day. Fig. 3 shows (at the left) the 22 observed functions $S(t)$ and $A(t)$ for the am shifts (top) and pm shifts (bottom). Moving to the right, Fig. 3 then shows that averages $\bar{S}(t)$ and $\bar{A}(t)$ and the associated histogram over 30-min subintervals.

We draw two conclusions from Fig. 3. First, on average the patients tend to arrive early, i.e., $\bar{A}(t) > \bar{S}(t)$ except at the end of the shift. Second, from the plots, we can see that there is much more variability in the actual arrivals than in the scheduled arrivals. In particular, the plots of $S(t)$ are step functions, whereas the plots of $A(t)$ are not.

3.2. No-shows and lateness

Let N_{no} be the number of the N_S scheduled arrivals that do not actually arrive, which we call no-shows. Note that we have the simple conservation equation $N_A = N_S - N_{no} + N_U$. Let X be the difference between an actual arrival time from its scheduled arrival time. We think of observed values of N_{no}/N_S and X as estimates of a no-show probability and a random deviation X, with associated lateness cumulative distribution function (cdf) F, both of which might depend on the scheduled arrival time. We examine deviations in more detail by looking at the proportion of arrivals that are late ($P(X > 0)$) and the average of the earliness among those that arrive early (X^-) and of the lateness among those that arrive late (X^+), as well as the overall average lateness or deviation (X). Table 1 shows the details for the scheduled patients in each hour of the am and pm shifts. A similar analysis of the other 15 doctors appears in [11].

Table 1 supports the following conclusions: (i) the proportion of no-shows is consistently about 8%, with the hourly values falling between 6% and 8% except for a rise at the ends of the day, (ii) the proportion of lateness is about 14% in the am and 11% in the pm, but otherwise roughly stable over time, (iii) the average lateness (X^+) is quite steady at just under 20 min, except for an increase to 30 min at the beginning of the day, (iv) the average earliness increases at the beginning of the day, soon approaching a steady-state value of about 60 min. The low initial earliness is evidently due a fixed start time. Our data are consistent with previous empirical evidence that patients arrive early more often than late [14,15].

Fig. 4 shows the lateness empirical cdf's (ecdf's) that are estimates of the lateness cdf F for each hour of the day. Consistent with the order of the averages seen in Table 1, Fig. 4 shows that the ecdf's are stochastically ordered (Section 9.1 of [19]), with the least earliness (lowest ecdf) in the first hour.

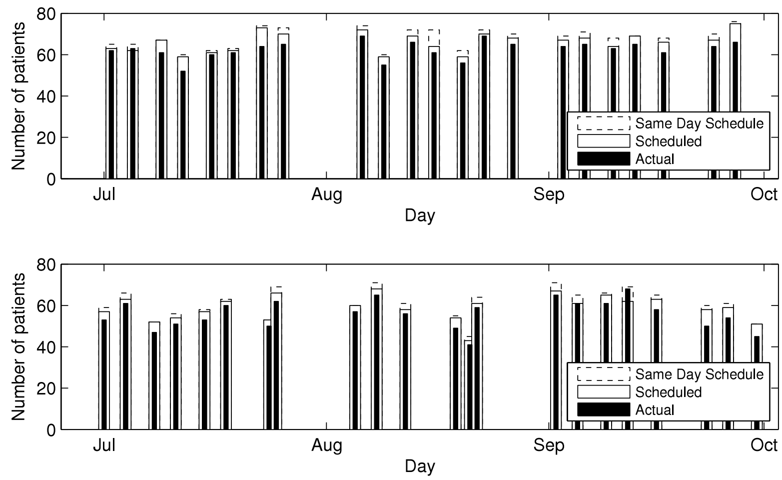

Fig. 2. Daily totals of N_S, N_U and N_A for am (top) and pm (bottom) shifts.

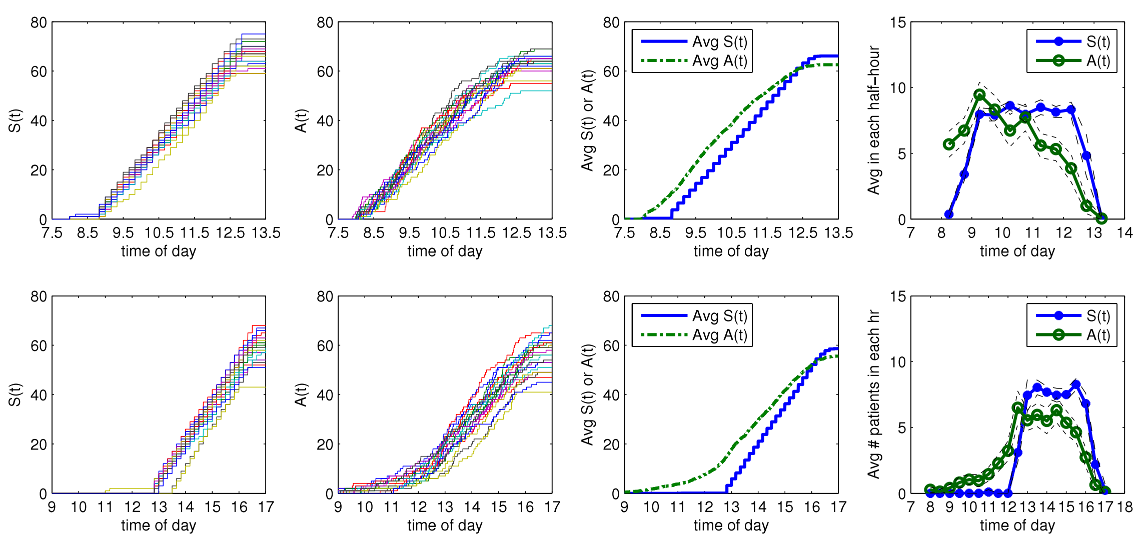

Fig. 3. Plots of the 22 scheduled arrival functions ($S(t)$) and actual arrival functions ($A(t)$) during am (top) and pm (bottom) shifts, followed by the direct averages and averages within 30-min intervals.

Table 1
Average numbers of scheduled arrivals for each hour, proportions of no-shows and lateness, and the average earliness (X^-), lateness (X^+) and overall deviation (X), among the 22 am and 22 pm shifts, plus 95% confidence intervals.

Interval	Avg # scheduled	% No-show	% Late	Avg(X^+)	Avg(X^-)	Avg(X)
AM shifts						
[8, 9]	3.8 ± 0.6	13.6 ± 8.3	25.3 ± 10.9	31.8 ± 22.5	-23.3 ± 4.5	-11.0 ± 6.8
[9, 10]	15.9 ± 0.8	6.3 ± 3.0	16.4 ± 2.8	30.4 ± 16.7	-32.6 ± 3.2	-22.1 ± 3.5
[10, 11]	16.6 ± 0.8	8.9 ± 2.6	16.4 ± 3.1	13.9 ± 5.2	-43.6 ± 4.6	-34.0 ± 4.7
[11, 12]	16.6 ± 0.5	7.8 ± 2.7	12.5 ± 3.7	16.5 ± 7.9	-57.8 ± 6.2	-48.3 ± 6.1
[12, 13]	13.1 ± 1.7	7.3 ± 2.7	7.7 ± 2.9	11.5 ± 5.9	-56.2 ± 8.3	-51.4 ± 8.8
[13.14]	0.1 ± 0.1	100.0				
Total	66.1 ± 2.0	8.2 ± 1.6	14.1 ± 1.6	20.8 ± 5.5	-46.2 ± 2.9	-36.7 ± 3.1
PM shifts						
[11, 12]	0.1 ± 0.2	50	0		-125.2	-125.2
[12, 13]	3.1 ± 0.7	6.0 ± 5.8	8.8 ± 9.3	56.7 ± 125.5	-67.4 ± 18.8	-57.8 ± 17.9
[13.14]	15.5 ± 1.4	6.3 ± 2.8	10.4 ± 4.0	12.4 ± 5.5	-61.4 ± 7.1	-53.2 ± 6.4
[14, 15]	15.1 ± 0.7	7.2 ± 2.4	8.4 ± 3.2	21.7 ± 9.6	-65.3 ± 9.5	-58.1 ± 9.6
[15, 16]	15.8 ± 0.6	12.4 ± 3.1	11.4 ± 4.0	13.6 ± 5.8	-60.7 ± 10.0	-52.4 ± 9.3
[16, 17]	9.0 ± 1.4	8.0 ± 5.5	13.3 ± 5.1	15.4 ± 6.9	-59.8 ± 13.8	-50.6 ± 13.4
[17, 18]	0.2 ± 0.5	0	0		-34.6	-34.6
Total	58.8 ± 2.7	8.4 ± 1.5	10.9 ± 2.2	17.8 ± 4.9	-61.9 ± 4.6	-53.3 ± 4.5

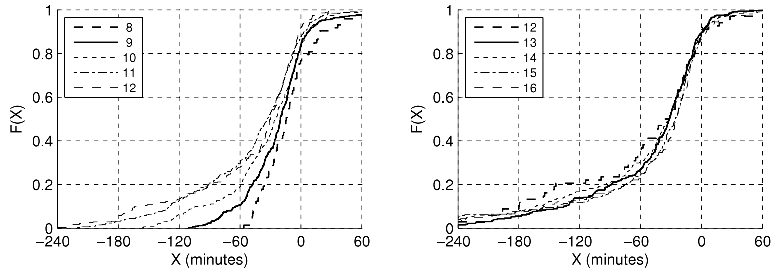

Fig. 4. The lateness empirical cdf's by hour for the am (left) and pm (right) shifts.

3.3. *Testing an NHPP within shifts*

We now test whether or not the actual arrival process within each shift can reasonably be regarded as an NHPP or even a PP. Given the appointments, we would be inclined to immediately dismiss this idea, but from Sections 3.1 and 3.2 we see that the presence of no-shows and lateness make the actual arrival process within the day substantially more random than the schedule.

To perform our statistical test of an NHPP, we use the conditional-uniform (CU) Kolmogorov–Smirnov (KS) test and the Lewis KS test from [13,12]; see those papers for a full development. The KS test determines if n observations can be regarded as a sample from a sequence of i.i.d. random variables $\{Y_n : n \geq 1\}$, each distributed as a random variable Y with a specified continuous cumulative distribution function (cdf) $G(x) \equiv P(Y \leq x), x \in \mathbb{R}$. Just like many of the other related tests, the KS test is based on the difference between the *empirical cdf* (ecdf)

$$G_n(x) \equiv n^{-1} \sum_{k=1}^{n} 1_{\{Y_k \leq x\}}, \quad x \in \mathbb{R}, \tag{1}$$

and the underlying cdf G, where 1_A is an indicator function, equal to 1 if the event A occurs, and equal to 0 otherwise. The KS test focuses on the maximum difference

$$D_n \equiv \sup_x \{|G_n(x) - G(x)|\}, \tag{2}$$

which has a distribution that is independent of the cdf G (provided that the cdf is continuous), e.g., see [20].

Both the CU and Lewis KS tests apply the classical CU property over each interval where the rate is approximately constant. For a 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.) $U[0, 1]$ random variables, each uniformly distributed on the interval $[0, 1]$. With this transformation, the observations become i.i.d. $U[0, 1]$ random variables, even if the rates of the NHPP's are different on different intervals of the day, because the CU property is independent of the rate of each interval. Then the KS test is used to test whether the data comes from an i.i.d. $U[0, 1]$ sequence.

This general approach to NHPP tests follows Brown et al. [3], which used a logarithmic data transformation after the CU transformation. Kim and Whitt [13] found that the Brown [3] KS test has significant power against alternative processes with non-exponential interarrival-time distributions, but that the Lewis KS test, which is based on the Durbin [5] transformation, consistently has more power. Also [13] showed that the direct CU KS test has especially low power against alternative processes with non-exponential interarrival-time distributions, which we attributed to the fact that the CU property focuses on the arrival times instead of the interarrival times, whereas the Durbin [5] transformation re-orders the interarrival times of the uniform random variables in ascending order.

In this section, we present the results of both the CU and Lewis KS tests, because [13] also showed that the CU KS test turns out to be relatively more effective against alternatives with dependent exponential interarrival times. (The re-ordering of the interarrival times by the data transformations evidently make the other methods less effective in detecting dependence, because the re-ordering weakens the dependence.) We note that the percentage of unscheduled arrivals among arrivals is so small in our application (on average 2.2 unscheduled arrivals out of 62.6 arrivals or 3.5%—see Section 2 and Fig. 2 for details) that it should have no bearing on the results.

In each day, we consider only the intervals $[9, 12]$ for am shifts and $[13, 16]$ for pm shifts. Because we have around 60 patient arrivals in each shift, if we apply the KS test to each day, the power of the test is weak because of the sample size. A common way to address this problem is to combine data from multiple days. We use arrival times in $[9, 12]$ from 5–6 am shifts and arrival times in $[13, 16]$ from 5–6 pm shifts to make sample sizes of about 200–300 interarrival times. From [13], we know that a sample size of 200–300 is sufficient to have reasonable power. For $L = 1$, we apply the CU property to each 1-hr interval; in other words, we allow each of the 1-hour intervals to have a different arrival rate. Similarly, when we set $L = 3$, we require each shift to have constant arrival rate but allow different arrival rates over different shifts, and $L = T$ means that we require the arrival rate to be constant throughout all of the shifts that are merged to give 200–300 interarrival times.

Table 2 shows the performance of the Lewis KS test as a function of the subinterval length L, represented by the p-values. In particular, the p-value is the probability of such a large deviation under the null hypothesis. We compare the p-value to the significance level of the test, which we take to be $\alpha = 0.05$. Consequently, the test dictates rejecting the NHPP hypothesis if the p-value is less than $\alpha = 0.05$. The smaller the p-value, the less likely the data came from an NHPP.

We apply these KS tests to both the scheduled arrivals and the actual arrivals. First, we see that the Lewis test consistently rejects the NHPP hypothesis for the scheduled arrivals for all values of L. In contrast, for the actual arrivals, no matter what value of L we use, the Lewis KS test consistently fails to reject the Poisson property. Just as in [13,12], the plots of the ecdf's used in the Lewis KS tests in the left two columns of Fig. 5 for am shifts dramatically support these results. (We note that the plots look very similar for pm shifts.) Recall that the cdf of a $U[0, 1]$ cdf is a line with slope 1 on $[0, 1]$.

The fact that the Lewis tests fails to reject the Poisson null hypothesis for $L = T$ supports the notion that the arrival data are not only consistent with an NHPP, but are also consistent with a PP (with constant rate over the shift). That is not surprising, because the appointment system serves to stabilize the arrival rate over time.

Table 3 and the right two columns of Fig. 5 provide the counterparts for the CU KS test to the results for the Lewis KS test in Table 2

Table 2
P-values of the Lewis KS test of an NHPP.

Days	Scheduled arrivals				Actual arrivals			
	n	$L = 1$	$L = 3$	$L = T$	n	$L = 1$	$L = 3$	$L = T$
July 2, 5, 9, 12, 16, 19 ($T = 18$)	279	0.00	0.00	0.00	265	0.35	0.35	0.52
July 23, 26, Aug 6, 9, 13, 16 ($T = 18$)	287	0.00	0.00	0.00	283	0.77	0.76	0.87
Aug 20, 23, 27, Sept 3, 6	233	0.00	0.00	0.00	220	0.15	0.71	0.96
Sept 10, 13, 17, 24, 27($T = 15$)	242	0.00	0.00	0.00	229	0.21	0.39	0.38
All AM shifts ($T = 66$)	1041	0.00	0.00	0.00	997	0.12	0.60	0.41
July 1, 4, 8, 11, 15, 18 ($T = 18$)	267	0.00	0.00	0.00	203	0.95	0.99	0.58
July 24, 25, Aug 5, 8, 12, 19	269	0.00	0.00	0.00	219	0.94	0.83	0.77
Aug 21, 22, Sept 2, 5, 9	225	0.00	0.00	0.00	186	0.49	0.62	0.87
Sept 12, 16, 23, 26, 30 ($T = 15$)	223	0.00	0.00	0.00	168	0.28	0.33	0.63
All PM shifts ($T = 66$)	984	0.00	0.00	0.00	776	0.15	0.49	0.80

Table 3
P-values of the CU KS test of an NHPP.

Days	Scheduled arrivals				Actual arrivals			
	n	$L = 1$	$L = 3$	$L = T$	n	$L = 1$	$L = 3$	$L = T$
July 2, 5, 9, 12, 16, 19 ($T = 18$)	279	0.00	0.34	0.97	265	0.14	0.00	0.09
July 23, 26, Aug 6, 9, 13, 16 ($T = 18$)	287	0.00	0.21	0.99	283	0.69	0.00	0.53
Aug 20, 23, 27, Sept 3, 6	233	0.00	0.45	1.00	220	0.44	0.00	0.26
Sept 10, 13, 17, 24, 27($T = 15$)	242	0.00	0.19	0.99	229	0.57	0.07	0.76
All AM shifts ($T = 66$)	1041	0.00	0.00	1.00	997	0.11	0.00	0.93
July 1, 4, 8, 11, 15, 18 ($T = 18$)	267	0.00	0.37	1.00	203	0.33	0.17	0.91
July 24, 25, Aug 5, 8, 12, 19	269	0.00	0.36	0.96	219	0.14	0.06	0.20
Aug 21, 22, Sept 2, 5, 9	225	0.00	0.47	0.98	186	0.29	0.81	0.90
Sept 12, 16, 23, 26, 30 ($T = 15$)	223	0.00	0.48	0.94	168	0.34	0.59	0.16
All PM shifts ($T = 66$)	984	0.00	0.00	1.00	776	0.05	0.05	0.81

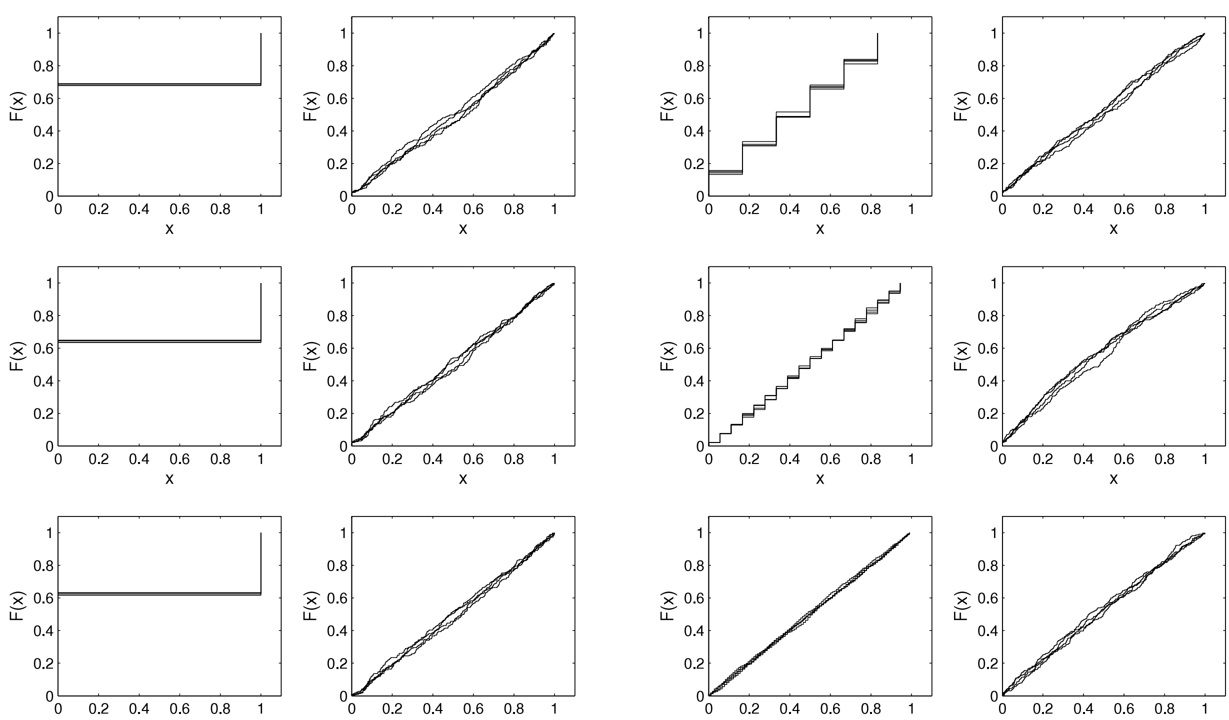

Fig. 5. Comparison of the empirical cdf's of the scheduled arrivals ($S(t)$) and actual arrivals ($A(t)$) in the final step of the Lewis KS test (after both the CU and the Lewis transformations are applied) and the CU KS test (after only the CU transformation is applied) to am shifts. The two columns on the left are after the Lewis KS test, while the two in the right are after the CU KS test. From top to bottom: $L = 1, 3, T$.

and the left two columns of Fig. 5. (Again, the plots look very similar for pm shifts.) For scheduled arrivals, we observe that the CU KS test completely misses the non-Poisson property when we use $L = 3$ or $L = T$; The right two columns of Fig. 5 help explain why: When $L = 1$, the almost uniform spacing between appointments is emphasized more. For actual arrivals, the CU KS test fails to reject the null hypothesis of an NHPP when $L = 1$ or $L = T$.

On the other hand, for the actual arrivals in the am shifts when $L = 3$ (meaning that the CU transformation is applied to each shift separately), the CU KS test rejects the null hypothesis of an NHPP.

Table 4
Sample mean and variance, and their ratio, of the number of scheduled arrivals (appointments made before the appointment day) and actual arrivals over am shifts (left) and pm shifts (right) on different days for all 16 clinic doctors.

Doc	AM Shift							PM Shift						
	n	Scheduled			Actual			n	Scheduled			Actual		
		$\bar{\mu}$	$\bar{\sigma}^2$	ratio	$\bar{\mu}$	$\bar{\sigma}^2$	ratio		$\bar{\mu}$	$\bar{\sigma}^2$	ratio	$\bar{\mu}$	$\bar{\sigma}^2$	ratio
1	17	76.9	65.1	0.8	73.5	48.5	0.7	19	72.9	50.8	0.7	70.6	54.7	0.8
2	21	53.9	33.6	0.6	50.8	38.6	0.8	18	47.8	58.3	1.2	45.4	23.2	0.5
3	12	51.4	26.6	0.5	50.1	29.7	0.6	11	40.3	46.4	1.2	37.7	30.2	0.8
4	34	65.4	27.9	0.4	63.4	19.6	0.3	12	64.6	16.4	0.3	61.1	14.6	0.2
5	23	105.5	133.8	1.3	101.8	134.2	1.3	19	100.9	70.4	0.7	97.2	75.5	0.8
6	17	102.8	557.4	5.4	96.1	493.4	5.1	4	86.3	42.9	0.5	76.5	73.0	1.0
7	13	90.3	155.6	1.7	86.7	146.1	1.7	8	83.6	75.1	0.9	79.6	82.6	1.0
8	10	56.8	198.2	3.5	57.4	95.4	1.7	19	50.8	180.4	3.6	48.8	103.6	2.1
9	22	66.1	21.3	0.3	62.6	17.4	0.3	22	58.8	35.9	0.6	55.7	49.5	0.9
10	2	19.0	8.0	0.4	24.5	24.5	1.0	9	29.2	16.9	0.6	28.4	11.5	0.4
11	10	46.6	17.2	0.4	44.4	24.7	0.6	12	41.5	22.3	0.5	40.3	17.8	0.4
12	12	42.5	18.6	0.4	39.3	33.8	0.9	10	34.9	51.2	1.5	33.0	10.9	0.3
13	12	40.8	19.1	0.5	38.3	8.6	0.2	13	31.1	23.7	0.8	32.0	19.7	0.6
14	1	25.0			28.0			23	34.4	38.2	1.1	34.1	22.8	0.7
15	10	38.0	15.8	0.4	36.9	13.7	0.4	11	34.5	26.1	0.8	33.6	24.9	0.7
16	12	37.8	23.8	0.6	35.1	17.5	0.5	10	34.6	24.9	0.7	33.3	26.0	0.8

In these cases, the CU KS test evidently detects dependence among the interarrival times, in a way that the Lewis test does not, which is evidently due to the time-varying arrival rate over the shift, which is evident from Fig. 3. Overall, we conclude that the arrival data within the shifts of each day are quite consistent with an NHPP.

4. Under-dispersion over multiple days

We now show that there is strong evidence of under-dispersion in the arrival process $A(t)$ over multiple days. We also show that this under-dispersion is primarily due to the anticipated under-dispersion in the scheduled arrival process $S(t)$.

4.1. Low variability of the shift totals

The under-dispersion of the scheduled and actual arrivals over multiple shifts is easily seen by looking at the ratio of the sample variance to the sample mean of the shift totals. To emphasize this point, we show these statistics for all 16 doctors in Table 4.

Table 4 shows consistent variance-to-mean ratios between 0.3 and 0.8, with the exception of doctors 5–8. For these, we see that: (i) the much higher variability is already present in the scheduled arrivals and (ii) that a closer examination reveals that there was a systematic change in the target schedule during the data collection period. Thus, we conclude that these exceptions should be considered anomalies and should be ignored.

Following Section 4.1 of [12], we can also apply the dispersion test to statistically test whether or not the dispersion in the arrival data are consistent with the Poisson property. We apply the dispersion test to doctor 9 alone, so that the sample size is not large, as before. The null hypothesis is that the shift arrival counts constitute i.i.d. Poisson random variables with unknown mean. The dispersion test uses the statistic

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

$$\bar{\sigma}^2 \equiv \bar{\sigma}_n^2 \equiv \frac{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}{n-1} \quad \text{and} \quad \bar{x} \equiv \bar{x}_n \equiv \frac{\sum_{i=1}^{n} x_i}{n}; \qquad (3)$$

e.g., see [10].

Since we are concerned with low variability, we consider the one-sided test and reject if $\bar{D}_n < \delta(n, 1-\alpha)$ where $P(\bar{D}_n < \delta(n, 1-\alpha)|H_0) = \alpha$, 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, $E[\bar{D}_n|H_0] = n-1$, $\mathrm{Var}(\bar{D}_n|H_0) = 2(n-1)$ and $(\chi^2_n - n)/\sqrt{2n}$ converges to the standard normal as n increases. Thus $\delta(n, 0.95) = \chi^2_{n-1,0.05}$, the 5th percentile of the χ^2_{n-1} distribution.

When we apply the dispersion test to the data, we find evidence of under-dispersion in N_A of am shifts. Since the sample size for am shifts is $n = 22$, $E[\bar{D}_n|H_0] = 21$ and $\mathrm{Var}(\bar{D}_n|H_0) = 42$. The 1st and 5th percentiles of the χ^2_{21} distribution are, respectively, 8.9 and 11.6. The $\bar{D}_n$ of N_A in 22 am shifts is 5.8, well below the 1st percentile of the chi-squared distribution. On the other hand, $\bar{D}_n$ of N_A in 22 pm shifts is 18.7, so we cannot reject the null hypothesis that the daily arrival counts of the pm shifts over the study period constitute independent Poisson random variables with the same mean. Thus, from the daily totals for doctor 9 alone, we have weak evidence of under-dispersion over multiple shifts, but when we consider all doctors, as in Table 4, the statistical evidence of under-dispersion is overwhelming.

5. Models for short-term and long-term planning

Our analysis suggests stochastic arrival process models, which can be applied in simulation planning tools. For short-term planning, i.e., given the schedule at the end of the previous day, the total number N_S of scheduled arrivals and the number $N_{S,j}$ scheduled in time slot j for all j are known. The model then assumes that no-shows are independent with probability p_j and, given an actual arrival in time slot j, there is an independent lateness (or earliness) distributed as the cdf F_j, where (p_j, F_j) are estimated from the data as in Section 3.2. We can incorporate the unscheduled arrivals by letting there be a Poisson random number N_U of unscheduled arrivals, distributed as i.i.d. uniform random variables over the shift.

For long-term planning, e.g., a week or a month in advance, our statistical analysis of the data at least partly supports a new parsimonious stylized *two-time-scale binomial-uniform arrival process model*: The number of actual arrivals during each shift (a longer time scale) can be assigned a binomial distribution with probability mass function $b(k; n, p)$, where the parameters n and p are chosen so that the mean np and variance $np(1-p)$ match the estimated values, assuming that the ratio of the variance to the mean is less than one, as was observed in Section 4. Then the arrivals throughout the shift (shorter time scale) can be distributed as i.i.d. uniform random variables over the shift. Over many shifts, the various binomial random variables can be dependent. For any single shift, this binomial-uniform model is consistent with the KS tests that exploit the CU transformation in Section 3.3.

More generally, we propose an associated *two-time-scale Gaussian-uniform arrival process model* for *both* call center arrivals and appointment-generated arrivals. We let the number of actual arrivals in period j of day d be $N_{A,d,j}$, where each period is a designated time interval. We let the stochastic process $\{N_{A,d,j} : d \geq 1, 1 \leq j \leq p\}$ be a Gaussian process, where the mean and variance of $N_{A,d,j}$ are chosen to match the sample mean and sample variance, just as for the binomial-uniform model. (The models are related by using the Gaussian approximation for the binomial distribution.) We then assume that the $N_{A,d,j}$ arrivals in period j arrive as i.i.d. uniform random variables over interval (d, j). The Gaussian-uniform model allows dependence among the $N_{A,d,j}$ variables for different pairs (d, j), which is parsimoniously characterized via the covariances. We think that the Gaussian-uniform model may provide a basis for forecasting, as in [7] and references therein, and then staffing.

Acknowledgments

This research was conducted while the first author was a doctoral student and the second author was an undergraduate at Columbia University. Support was received from NSF grants CMMI 1066372 and 1265070. The authors thank Dr. Sang-Man Jin in the Division of Endocrinology and Metabolism at Samsung Medical Center for providing advice.

References

[1] A.N. Avramidis, A. Deslauriers, P. L'Ecuyer, Modeling daily arrivals to a telephone call center, Manage. Sci. 50 (7) (2004) 896–908.
[2] A.D. Barbour, L. Holst, S. Janson, Poisson Approximation, Oxford University Press, Oxford, UK, 1992.
[3] L. Brown, N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, L. Zhao, Statistical analysis of a telephone call center: a queueing-science perspective, J. Amer. Statist. Assoc. 100 (2005) 36–50.
[4] T. Cayirli, E. Veral, Outpatient scheduling in health care: a review of literature, Prod. Oper. Manage. 12 (4) (2003) 519–549.
[5] J. Durbin, Some methods for constructing exact tests, Biometrika 48 (1) (1961) 41–55.
[6] D. Gupta, B. Denton, Appointment scheduling in health care: challenges and opportunities, IIE Trans. 40 (9) (2008) 800–819.
[7] R. Ibrahim, P. L'Ecuyer, N. Regnard, H. Shen, On the modeling and forecasting of call center arrivals, in: Proceedings of the 2012 Winter Simulation Conference 2012, 2012, pp. 256–267.
[8] G. Jongbloed, G. Koole, Managing uncertainty in call centres using Poisson mixtures, Appl. Stoch. Models Bus. Ind. 17 (4) (2001) 307–318.
[9] O. Jouini, S. Benjaafar, Queueing systems with appointment-driven arrivals, non-punctual customers, and no-shows, Tech. Rep., Working Paper, 2012.
[10] N. Kathirgamatamby, Note on the Poisson index of dispersion, Biometrika 40 (1) (1953) 225–228.
[11] S.-H. Kim, P. Vel, W. Whitt, W.C. Cha, Analysis of arrival data from an endocrinology clinic, Columbia University, 2015. http://www.columbia.edu/~ww2040/allpapers.html.
[12] S.-H. Kim, W. Whitt, Are call center and hospital arrivals well modeled by nonhomogeneous Poisson processes? Manuf. Serv. Oper. Manage. 16 (3) (2014) 464–480.
[13] S.-H. Kim, W. Whitt, Choosing arrival process models for service systems: tests of a nonhomogeneous Poisson process, Nav. Res. Logist. 61 (1) (2014) 66–90.
[14] K.J. Klassen, T.R. Rohleder, Scheduling outpatient appointments in a dynamic environment, J. Oper. Manage. 14 (2) (1996) 83–101.
[15] B. Lehaney, S. Clarke, R. Paul, A case of an intervention in an outpatients department, J. Oper. Res. Soc. (1999) 877–891.
[16] N. Liu, S. Ziya, V.G. Kulkarni, Dynamic scheduling of outpatient appointments under patient no-shows and cancellations, Manuf. Serv. Oper. Manage. 12 (2) (2010) 347–364.
[17] C.G. Moore, P. Wilson-Witherspoon, J.C. Probst, Time and money: effects of no-shows at a family practice residency clinic, Fam. Med. 33 (7) (2001) 522–527.
[18] R.D. Neal, D.A. Lawlor, V. Allgar, M. Colledge, S. Ali, A. Hassey, C. Portz, A. Wilson, Missed appointments in general practice: retrospective data analysis from four practices, Br. J. Gen. Pract. 51 (471) (2001) 830–832.
[19] S.M. Ross, Stochastic Processes, second ed., Wiley, New York, 1996.
[20] R. Simard, P. L'Ecuyer, Computing the two-sided Kolmogorov–Smirnov distribution, J. Stat. Softw. 39 (11) (2011) 1–18.
[21] X. Zhang, L. Hong, P.W. Glynn, Timescales in modeling call center arrivals, Working Paper, Stanford University, 2014.