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

Song-Hee Kim\textsuperscript{a}, Ponni Vel\textsuperscript{b}, Ward Whitt\textsuperscript{b,}\textsuperscript{,*}, Won Chul Cha\textsuperscript{c}

\textsuperscript{a} Yale School of Management, Yale University, USA
\textsuperscript{b} Industrial Engineering and Operations Research, Columbia University, USA
\textsuperscript{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.

© 2015 Elsevier B.V. All rights reserved.

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
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 \( N_{a} \) 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 \( N_{a} \) and \( N_{u} \).

Fig. 2 depicts the values of \( N_{a} \), \( N_{u} \), and \( N_{a} \) during the 13-week study period. The average (standard deviation) values of \( N_{a} \), \( N_{u} \), and \( N_{a} \) are 66.1 (4.6), 22.1 (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 \( N_{a} \) and \( N_{u} \) that \( N_{a} \) necessarily has a small impact on \( N_{a} \). Also, note that \( N_{a} \) and \( N_{u} \) have low variability; we discuss and statistically test their under-dispersion in Section 4. On average, \( N_{a} \) is 95% of \( N_{a} \) in both the am and pm shifts; in particular, \( N_{a} \) ranges from 88% to 102% of \( N_{a} \) in am shifts and from 86% to 110% in pm shifts, and rarely exceeds \( N_{a} \).

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 FP.

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 \( \tilde{S}(t) \) and \( \tilde{A}(t) \) 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., \( A(t) > S(t) \) 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 \( N_{a} \) be the number of the \( N_{a} \) scheduled arrivals that do not actually arrive, which we call no-shows. Note that we have the simple conservation equation \( N_{a} = N_{a} - N_{a} + N_{a} \). Let \( X \) be the difference between an actual arrival time from its scheduled arrival time. We think of observed values of \( N_{a} / N_{a} \) 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. 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

<table>
<thead>
<tr>
<th>Interval</th>
<th>Avg # scheduled</th>
<th>% No-show</th>
<th>% Late</th>
<th>Avg($X^+$)</th>
<th>Avg($X^-$)</th>
<th>Avg($X$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM shifts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[8, 9]</td>
<td>3.8 ± 0.6</td>
<td>13.6 ± 8.3</td>
<td>25.3 ± 10.9</td>
<td>31.8 ± 22.5</td>
<td>−23.3 ± 4.5</td>
<td>−11.0 ± 6.8</td>
</tr>
<tr>
<td>[9, 10]</td>
<td>15.9 ± 0.8</td>
<td>6.3 ± 3.0</td>
<td>16.4 ± 2.8</td>
<td>30.4 ± 16.7</td>
<td>−32.6 ± 3.2</td>
<td>−22.1 ± 3.5</td>
</tr>
<tr>
<td>[10, 11]</td>
<td>16.6 ± 0.8</td>
<td>8.9 ± 2.6</td>
<td>16.4 ± 3.1</td>
<td>13.9 ± 5.2</td>
<td>−43.6 ± 4.6</td>
<td>−34.0 ± 4.7</td>
</tr>
<tr>
<td>[11, 12]</td>
<td>16.6 ± 0.5</td>
<td>7.8 ± 2.7</td>
<td>12.5 ± 3.7</td>
<td>16.5 ± 7.9</td>
<td>−57.8 ± 6.2</td>
<td>−48.3 ± 6.1</td>
</tr>
<tr>
<td>[12, 13]</td>
<td>13.1 ± 1.7</td>
<td>7.3 ± 2.7</td>
<td>7.7 ± 2.9</td>
<td>11.5 ± 5.9</td>
<td>−56.2 ± 8.3</td>
<td>−51.4 ± 8.8</td>
</tr>
<tr>
<td>[13, 14]</td>
<td>0.1 ± 0.1</td>
<td>100.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>66.1 ± 2.0</td>
<td>8.2 ± 1.6</td>
<td>14.1 ± 1.6</td>
<td>20.8 ± 5.5</td>
<td>−46.2 ± 2.9</td>
<td>−36.7 ± 3.1</td>
</tr>
<tr>
<td>PM shifts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[11, 12]</td>
<td>0.1 ± 0.2</td>
<td>50</td>
<td>0</td>
<td>−125.2</td>
<td>−125.2</td>
<td></td>
</tr>
<tr>
<td>[12, 13]</td>
<td>3.1 ± 0.7</td>
<td>6.0 ± 5.8</td>
<td>8.8 ± 9.3</td>
<td>56.7 ± 125.5</td>
<td>−67.4 ± 18.8</td>
<td>−57.8 ± 17.9</td>
</tr>
<tr>
<td>[13, 14]</td>
<td>15.5 ± 1.4</td>
<td>6.3 ± 2.8</td>
<td>10.4 ± 4.0</td>
<td>12.4 ± 5.5</td>
<td>−61.4 ± 7.1</td>
<td>−53.2 ± 6.4</td>
</tr>
<tr>
<td>[14, 15]</td>
<td>15.1 ± 0.7</td>
<td>7.2 ± 2.4</td>
<td>8.4 ± 3.2</td>
<td>21.7 ± 9.6</td>
<td>−65.3 ± 9.5</td>
<td>−58.1 ± 9.6</td>
</tr>
<tr>
<td>[15, 16]</td>
<td>15.8 ± 0.6</td>
<td>12.4 ± 3.1</td>
<td>11.4 ± 4.0</td>
<td>13.6 ± 5.8</td>
<td>−60.7 ± 10.0</td>
<td>−52.4 ± 9.3</td>
</tr>
<tr>
<td>[16, 17]</td>
<td>9.0 ± 1.4</td>
<td>8.0 ± 5.5</td>
<td>13.3 ± 5.1</td>
<td>15.4 ± 6.9</td>
<td>−59.8 ± 13.8</td>
<td>−50.6 ± 13.4</td>
</tr>
<tr>
<td>[17, 18]</td>
<td>0.2 ± 0.5</td>
<td>0</td>
<td>0</td>
<td>−34.6</td>
<td>−34.6</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>58.8 ± 2.7</td>
<td>8.4 ± 1.5</td>
<td>10.9 ± 2.2</td>
<td>17.8 ± 4.9</td>
<td>−61.9 ± 4.6</td>
<td>−53.3 ± 4.5</td>
</tr>
</tbody>
</table>
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 \( n \) conditional-uniform (CU) Kolmogorov–Smirnov (KS) test and the presence of no-shows and lateness make the actual arrival process substantially more random than the schedule.

Giventheappointments,wewouldbeinclinedtoimmediately dismiss this idea, but from Sections 3.1 and 3.2 we seethat the

Giventheappointments,wewouldbeinclinedtoimmediately

dissmissthisidea,buttromSections3.1and3.2weseethatthe

dissmissthisidea,buttromSections3.1and3.2weseethatthe

Giventheappointments,wewouldbeinclinedtoimmediately

dissmissthisidea,buttromSections3.1and3.2weseethatthe

dissmissthisidea,buttromSections3.1and3.2weseethatthe

dissmissthisidea,buttromSections3.1and3.2weseethatthe

dissmissthisidea,buttromSections3.1and3.2weseethatthe

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 ecdf of a \( U[0, 1] \) ecdf 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.

<table>
<thead>
<tr>
<th>Days</th>
<th>Scheduled arrivals</th>
<th>Actual arrivals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n )</td>
<td>( L = 1 )</td>
</tr>
<tr>
<td>July 2, 5, 9, 12, 16, 19 (T = 18)</td>
<td>279</td>
<td>0.00</td>
</tr>
<tr>
<td>July 23, 26, Aug 6, 9, 13, 16 (T = 18)</td>
<td>287</td>
<td>0.00</td>
</tr>
<tr>
<td>Aug 20, 23, 27, Sept 3, 6</td>
<td>233</td>
<td>0.00</td>
</tr>
<tr>
<td>Sept 10, 13, 17, 24, 27 (T = 15)</td>
<td>242</td>
<td>0.00</td>
</tr>
<tr>
<td>All AM shifts (T = 66)</td>
<td>1041</td>
<td>0.00</td>
</tr>
<tr>
<td>July 1, 4, 8, 11, 15, 18 (T = 18)</td>
<td>267</td>
<td>0.00</td>
</tr>
<tr>
<td>July 24, 25, Aug 5, 8, 12, 19</td>
<td>269</td>
<td>0.00</td>
</tr>
<tr>
<td>Aug 21, 22, Sept 2, 5, 9</td>
<td>225</td>
<td>0.00</td>
</tr>
<tr>
<td>Sept 12, 16, 23, 26, 30 (T = 15)</td>
<td>223</td>
<td>0.00</td>
</tr>
<tr>
<td>All PM shifts (T = 66)</td>
<td>984</td>
<td>0.00</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Days</th>
<th>Scheduled arrivals</th>
<th>Actual arrivals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n )</td>
<td>( L = 1 )</td>
</tr>
<tr>
<td>July 2, 5, 9, 12, 16, 19 (T = 18)</td>
<td>279</td>
<td>0.00</td>
</tr>
<tr>
<td>July 23, 26, Aug 6, 9, 13, 16 (T = 18)</td>
<td>287</td>
<td>0.00</td>
</tr>
<tr>
<td>Aug 20, 23, 27, Sept 3, 6</td>
<td>233</td>
<td>0.00</td>
</tr>
<tr>
<td>Sept 10, 13, 17, 24, 27 (T = 15)</td>
<td>242</td>
<td>0.00</td>
</tr>
<tr>
<td>All AM shifts (T = 66)</td>
<td>1041</td>
<td>0.00</td>
</tr>
<tr>
<td>July 1, 4, 8, 11, 15, 18 (T = 18)</td>
<td>267</td>
<td>0.00</td>
</tr>
<tr>
<td>July 24, 25, Aug 5, 8, 12, 19</td>
<td>269</td>
<td>0.00</td>
</tr>
<tr>
<td>Aug 21, 22, Sept 2, 5, 9</td>
<td>225</td>
<td>0.00</td>
</tr>
<tr>
<td>Sept 12, 16, 23, 26, 30 (T = 15)</td>
<td>223</td>
<td>0.00</td>
</tr>
<tr>
<td>All PM shifts (T = 66)</td>
<td>984</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Fig. 5. Comparison of the empirical cdfs 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.
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

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

\[
\sigma^2 = \tilde{\sigma}^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \quad \text{and} \quad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n},
\]

\[1 - \alpha \right) \) of \( D_n \) is \( \alpha \), using \( \alpha = 0.05 \). Under the null hypothesis, \( D_n \) is distributed as \( \chi^2_{n-1} \), 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, \( E[D_n|H_0] = n - 1 \), \( \text{Var}(D_n|H_0) = 2(n - 1) \) and \( (\chi^2_{n-1} - n)/\sqrt{2n} \) converges to the standard normal as \( n \) increases. Thus \( \delta(n, 0.95) = \chi^2_{n-1,0.05} \) the 5th percentile of the \( \chi^2_{n-1} \) distribution.

When we apply the dispersion test to the data, we find evidence of under-dispersion in \( N_0 \) of am shifts. Since the sample size for am shifts is \( n = 22 \), \( E[D_n|H_0] = 21 \) and \( \text{Var}(D_n|H_0) = 42 \). The 1st and 5th percentiles of the \( \chi^2_{21} \) distribution are, respectively, 8.9 and 11.6. The \( D_n \) of \( N_0 \) in 22 am shifts is 5.8, well below the 1st percentile of the chi-squared distribution. On the other hand, \( D_n \) of \( N_0 \) 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 \( N_0 \) of scheduled arrivals and the number \( N_{k,j} \), 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) distribution as the cdf \( F_i \), 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 \( N_0 \) 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