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/

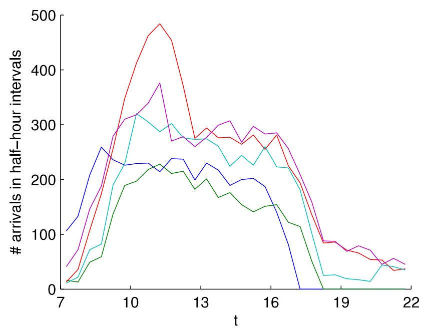

Fig. 1. Arrival counts in half-hour intervals at a call center on five Mondays during April 2001 (see [12] for details; VRI-Summit type arrival).

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.