A Data-Driven Model of an Emergency Department

Ward Whitt [1], Xiaopei Zhang [2]

Abstract

This paper develops an aggregate stochastic model of an emergency department (ED) based on a careful study of data on individual patient arrival times and length of stay in the ED of the Rambam Hospital in Israel, which was used in a large-scale exploratory data analysis by Armony et al. (2015). This data set is of special interest because it has been made publicly available, so that the experiments are reproducible. Our analysis confirms the previous conclusions about the time-varying arrival rate and its consequences, but we also find that the admission probability and the patient length-of-stay distribution should be time varying as well. Our analysis culminates in a new time-varying infinite-server aggregate stochastic model of the ED, where both the length-of-stay distribution and the arrival rate are periodic over a week.

Keywords: emergency departments, nonstationary stochastic models, queueing models, nonhomogeneous Poisson process, time-varying length-of-stay distribution, two-time-scale arrival process model

1. Introduction

There is a long history of operations research studies aimed at improving the quality and efficiency of healthcare, as illustrated by the early study [1] and the recent surveys [2, 3, 4, 5]. Nevertheless, as emphasized in [6], there remains a great need for further improvement. Much of this improvement is likely to come from vastly improved data collection, storage, retrieval and analysis. The power of data analysis is illustrated by extensive exploratory

[1]Industrial Engineering and Operations Research, Columbia University, Email: ww2040@columbia.edu, Correspondence to: MailCode 4704, S. W. Mudd Building, 500 West 120th Street, New York NY 10027-6699, U.S.A.

[2]Industrial Engineering and Operations Research, Columbia University, Email: xz2363@columbia.edu

data analysis of the patient flow in the large Rambam Hospital in Israel from a queueing science perspective conducted by Armony et al. [7]. In addition to their own analysis of the patient flow data at the level of individual patients, they arranged to make their data publicly available, thus facilitating reproducible studies aimed at generating general conclusions of widespread applicability. In this paper, we respond by analyzing a portion of the patient flow data provided by [7]. In particular, we focus on the emergency department (ED), just as in §3 of [7].

Among the many OR studies in healthcare, many have already focused on the emergency department, e.g., [8, 9, 10, 11, 12]. As those papers illustrate, the customary goal is to improve system design and operations. In contrast, in this paper, we focus solely on analyzing the patient flow data to determine what is a good aggregate stochastic model of the emergency department. This careful analysis is justified because emergency departments are complicated. The results here are intended to make it possible to more quickly build better stochastic models that can be used to improve healthcare design and operations.

As others have discovered before, e.g., see §6.3 of [13], the authors in §3 of [7] observe significant time dependence in the arrival rate, departure rate and average occupancy levels of the ED. We confirm those observations here, but we go further by pointing out significant time dependence in (i) the admission probability of an arriving patient and (ii) the length-of-stay (LoS) distribution of arriving patients. Time dependence in LoS was also a major theme in the recent study of a Singapore hospital in [14].

The available ED patient flow data is powerful in that it includes arrival and departure times of individual patients, but it is also limited in that it does not contain a detailed account of all the steps and processes that take place during a patient's stay. Thus, given the available data, we are only able to construct a relatively rough aggregate stochastic model, but even that can be useful and is not easy. Our model has only three components: (i) an arrival process model, (ii) an admission probability model, and (iii) a LoS model. All three are complicated, because we find that all three should be regarded as time-varying. Given those components, our aggregate stochastic model for system occupancy is a $G_t/GI_t/\infty$ time-varying infinite-server queue, which is much more tractable than the notation suggests. (G_t denotes a general (non-renewal, non-Markov) time-varying arrival process, while GI_t denotes mutually independent service times, independent of the arrival process, but with a time-dependent distribution.)

Consistent with §3 of [7], we find that the ED arrival rate should be time varying, but we emphasize that the proper view is over a week as opposed to the common daily view. In particular, we think that the arrival rate can be regarded as periodic over a week. As in [7] and [11] before, we find that there is moderate overdispersion compared to a non-homogeneous Poisson process (NHPP). We conclude that it might be reasonable to use an NHPP arrival process model, but in fact we suggest instead a two-time-scale arrival process model. We suggest first modeling the daily totals as a discrete-time Gaussian process and then letting the arrivals during the day, given the daily total, be distributed as an NHPP. The conditional NHPP means that the arrival times (not interarrival times!) of the daily total number of arrivals are treated as i.i.d. random variables on the entire day with a probability density function (pdf) proportional to the arrival rate function for that day, as discussed in [15, 16]. This arrival process model is a variant of the model proposed by [15]. We find that the model is supported by statistical tests in [16, 17]; see §3.4.

The two-time-scale model is convenient because it supports focusing on arrivals over successive days and arrivals within days separately. There is precedent for two-time-scale healthcare models in [14, 18], but these are very different. The first [18] focuses on the hospital plus the ED, observing that the Internal Wards (IW's) operate on the slower time scale of days, whereas the ED operates on the faster time scale of hours. The paper [18] proposes and analyzes a Markov chain (MC) model of that system, using a discrete-time MC for the days and a continuous-time MC for the transitions within days. On the other hand, in §3.2 of [14] the authors propose a two-time-scale model of the LoS. The general thrust of of [14] is consistent with our time-varying LoS distribution, which we discuss next.

We present strong evidence that the LoS distribution needs to be regarded as time-varying, and find that it suffices to make it time-varying over hours. Figure 8 here shows that the average occupancy level and departure rate are not predicted properly by a using the overall LoS distribution. We can make the connection by applying the theory of infinite-server queues as in [19] or, equivalently, the time-varying Little's law [20, 21]. We show how the time-varying LoS can be efficiently and effectively analyzed by exploiting a discrete-time model in the time scale of hours.

Here is how the rest of this paper is organized. In §2 we briefly describe the Rambam hospital and our data source, referring to [7] for more details. We analyze and model the ED arrival process in §3; we also discuss the admission probability there. In §4, we analyze and model the LoS. In §5, we examine the departure process, showing that it can be useful to view the departure process in reverse time. In §6 we compare our model to simulation. Finally, we draw conclusions in §7. Supplementary material is provided in an online appendix [22].

2. The Rambam Emergency Department and the Data

As in [7], we study the Rambam hospital, a large 1000-bed hospital with 45 medical units in Haifa, Israel. In particular, as in §3 of [7], we focus on the emergency internal medicine unit (EIMU), which is the largest unit in a comprehensive emergency department (ED). That focus is justified because the different units of the ED are physically separate and share few resources. About 60% of all new patients enter the hospital through the ED and the majority of those enter through the EIMU, which we henceforth simply call the ED. After being examined and treated within the ED, patients are either admitted to one of the internal wards (IW's) or released, as depicted in Figure 2 of [7]. About 40% of arrivals to the ED are admitted.

As directed in Appendix 2 of [7], we obtained the data from the SEELab data-based research laboratory at the Technion. The available hospital data was collected from January 2004 to October 2007. We only focus on the 25-week period from December 2004 to May 2005. In particular, we use the 5th, 6th, 13th and 18th columns of the visit table in the database, which are the entry group, first department, entry time and ED duration. In the raw data, the time records are rounded to the nearest second.

A total of 58,332 patients visited the comprehensive ED, with 24,317 going to the EIMU (3955, 4360, 3530, 4324, 3965 and 4183 for each month). Table 1 provides the total number of arrivals to the ED and length-of-stay (LoS) statistics for each of the sample populations used in successive analyses. The LoS refers to the LoS within the ED up until the time that a decision is made to admit the patient to an IW or not. Thus, the LoS does not include the delay until transfer is completed after the admission decision, commonly called ED boarding.

From both the database and [7], we know that the ED patients can be divided into two groups according to the admission decision; we pay attention to whether or not patients are admitted. Even though the admission decision cannot be known in advance, we find that the proportion of admitted patients in successive hours is time-dependent and thus can be exploited in modeling and analysis.

There are several variables in the database that can be used to help classify the patients. In this paper we use the "exit_group", which we find to be consistent with the "exit_unit", "exit_department" and "num_dep" in the visits table. "exit_group=1" means the patient was released from the Emergency Department and was not admitted to any hospital department; "exit_group=2" means the patient was released from a hospital department, which means that he was admitted to at least one department from the ED. So when we focus on patients who entered ED first, this will tell us whether a patient was admitted to an IW or not. Among the $23,409$ patients that visited the ED (the EIMU) in the 25-week period, $9,669$ (about 40%) were admitted and $13,740$ were not.

No.	Data description	Sample Size	mean	stdv	1st qu.	median	3rd qu.
1	Larger ED	58,332 (182 days)	3.08	5.36	1.06	2.16	3.87
2	ED (EIMU)	24,317 (182 days)	4.10	3.49	1.90	3.31	5.26
3	Truncation of 2	23,409 (175 days)	4.10	3.49	1.90	3.31	5.26
4	Truncation of 2	23,421 (175 days)	4.09	3.41	1.90	3.31	5.26
5	Admitted from 3	9,669 (175 days)	4.01	3.26	1.68	3.45	5.53
6	Non-admitted from 3	13,740 (175 days)	4.17	3.65	2.01	3.24	5.05
7	Truncation of 2	23,407 (175 days)	4.10	3.50	1.90	3.31	5.27
8	Admitted from 7	9,668 (175 days)	4.01	3.27	1.68	3.45	5.53
9	Non-admitted from 7	13739 (175 days)	4.17	3.65	2.01	3.24	5.06

Table 1: Sample size (number of arrivals) and LoS statistics (in hours) in different views of the ED data.

Below is a detailed description of the datasets in Table 1. Throughout the paper, we will refer to Table 1 to specify the population.

- Dataset 2: the subset of dataset 1 containing those patients who went to the EIMU within the comprehensive ED during the 25 week period; we refer to the EIMU simply as the ED.

- Dataset 3: the subset of dataset 2 containing those patients that arrived within the 25 week period. (We use dataset 2 for occupancy statistics.)

- Dataset 4: the subset of the dataset 2 containing those patients who who both entered and departed the system between Dec. 5, 2004, and May 28, 2005, and have LOS less then 48 hours.

- Dataset 5: the subset of dataset 3 containing those patients who were admitted after visiting the ED ("exit_group=1").

- Dataset 6: the subset of dataset 3 containing those patients who were not admitted after visiting the ED ("exit_group=2").
- Dataset 7: the subset of dataset 2 containing those patients whose departure times are in the 25-week period (from Dec. 5, 2004, to May 28, 2005).
- Dataset 8: the subset of dataset 7 containing those patients who were admitted after visiting the ED ("exit_group=1").
- Dataset 9: the subset of dataset 7 containing those patients who were not admitted after visiting the ED ("exit_group=2").

3. The ED Arrival Process

In this section, we study the arrival process of patients at the ED (by which we always mean the EIMU). In §3.1 we look at the daily totals; we briefly discuss dependence among the daily totals in §3.2. In §3.3 we estimate the hourly arrival rates over a week. We evaluate the stochastic variability in the arrival process in §3.4, which leads to proposing the two-time-scale model involving a conditional nonhomogeneous Poisson process (NHPP). In §3.4.1 we estimate the index of dispersion for counts; in §3.4.2 we report results of statistical tests of the conditional NHPP property, drawing on [16, 17]. In §3.5 we examine the arrival processes of two separate groups of patients: those that are ultimately admitted to one of the IW's and those that are not. Finally, in §3.6, we summarize the two-time-scale model for the arrival process that we propose, based on that statistical analysis.

3.1. Daily Totals

Table 2 shows the number of patients that arrived at the ED on each day from Dec. 5, 2004, to May 28, 2005 (25 weeks). The $25 \times 7 = 175$ daily totals vary from 77 to 191, and have mean 133.8 and median 135.

Some of the fluctuation may be explained by Jewish holidays. In Israel, Dec. 8, 2004, to Dec. 12, 2004, in week 1 and 2 was Hanukkah, while Apr. 24, 2005, to Apr. 30, 2005, in week 21 was Passover. We see that the daily totals are somewhat low during weeks 1 and 2, but not very different in week 21. We also notice that another low period occurred between Feb. 4, 2005, and Feb. 11, 2005, in weeks 9 and 10, for which we have no explanation. It is possible that this was due to military hostilities, but we could not verify

Week	Sun.	Mon.	Tue.	Wed.	Thu.	Fri.	Sat.	Total	Mean
1	150	147	132	107	123	100	99	858	122.57
2	143	147	127	138	121	101	103	880	125.71
3	162	155	147	136	144	94	98	936	133.71
4	186	155	135	136	119	100	131	962	137.43
5	164	171	149	146	142	110	110	992	141.71
6	175	144	157	136	156	115	105	988	141.14
7	181	157	140	109	145	114	107	953	136.14
8	176	145	139	150	126	127	102	965	137.86
9	171	160	125	137	137	77	84	891	127.29
10	134	127	119	115	95	88	82	760	108.57
11	165	117	121	133	154	123	132	945	135.00
12	163	142	135	142	129	115	100	926	132.29
13	173	166	168	136	138	108	108	997	142.43
14	169	155	155	137	143	127	106	992	141.71
15	180	152	132	148	162	111	105	990	141.43
16	159	164	191	128	126	95	111	974	139.14
17	163	135	146	128	138	111	145	966	138.00
18	160	123	168	136	133	119	102	941	134.43
19	132	147	152	138	133	116	101	919	131.29
20	162	150	140	126	113	113	96	900	128.57
21	143	165	153	130	130	111	117	949	135.57
22	151	147	132	114	114	114	96	868	124.00
23	159	135	151	119	107	122	100	893	127.57
24	164	163	153	147	156	111	91	985	140.71
25	165	141	159	138	147	125	104	979	139.86
Total	4050	3710	3626	3310	3331	2747	2635	23409	
Mean	162.00	148.40	145.04	132.40	133.24	109.88	105.40	936.36	
Var.	191.58	187.08	275.71	139.33	270.44	152.78	196.75	3110.99	

Table 2: Number of arrivals at the ED on each day from Dec. 5, 2004, to May 28, 2005, (25 weeks, Dataset 2).

that. We do not omit these periods from our data because they represent unanticipated random events that do occur.

Within a week, Sunday has the largest number of arrivals, which is to be expected because it is the beginning of the work week in Israel. Then the average daily totals decrease over the week. Friday and Saturday have much fewer arrivals, which may be expected because that is the weekend. We also computed the variance of daily totals for each day of week. By looking at the dispersion (ratio of the variance to the mean), we see that there is a moderate level of overdispersion for the daily totals compared to a Poisson process (where the dispersion is 1).

Figure 1 is a plot of the weekly totals. It confirms our observation above about the low values in week 10. Table 2 shows that the mean weekly number of arrivals is 936. Hence, if the arrival process were an NHPP, then the standard deviation of the weekly total would be about 31. Figure 1 is roughly

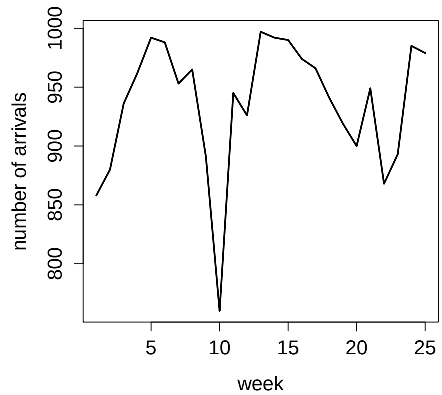


Figure 1: Weekly arrival totals over the 25-week study period. (Use dataset 3.)

consistent with that Poisson property, except for week 10, which is about 5 standard deviations below the mean.

We investigated models for the daily totals. We first considered a two-factor statistical regression model with Gaussian residuals for the daily total numbers of arrivals; see §§2.7, 3.7 and 6.5 of [23] for background. The daily total is represented as

$$T(w,d) \equiv A + Bw + Cd + G(0,\sigma^2), \tag{1}$$

where $\equiv$ denotes equality by definition, w represents the week and d is the day-of-week (DoW), while $G(0,\sigma^2)$ is a mean-0 Gaussian random variable with variance σ^2 (to be estimated) and A, B and C are constants. The week and the DoW are the two factors. Because there is redundancy in model (1), we set $\sum B \equiv 0$ and $\sum C \equiv 0$, so that A gives the average daily total number of arrivals for all days.

Table 3 is the usual Analysis of Variance (ANOVA) table for the regression. From the P-values in the last column of Table 3, we see that both factors

Factor	Sum of square	df	Mean sum of square	F statistics	P-value
Week	10,666	24	444.4	2.75	1.1×10^{-4}
DoW	62,893	6	10,482.2	64.89	$< 10^{-12}$
Residuals	23,262	144	161.5		

Table 3: ANOVA table for the two-factor model (1). (Use dataset 3.)

are statistically significant at the 1% level. From the residuals, the estimated variance is $\hat{\sigma}^2 = 161.5 = 12.71^2$. Under this model, the variance-to-mean ratio is $161.5/133.8 = 1.21$. The Gaussian two-factor model is supported by observing that the residuals are consistent with the Gaussian distribution, as can be seen from the histogram of the residuals and the QQ-plot of the studentized residuals in the appendix [22].

However, for applications, we would actually prefer the single-factor model with only the DoW as the single factor, because the DoW effect is known, whereas the the week effect is not, but the results above show the consequence if we can assume that it can be known or, more generally, if better estimates of the daily totals can be generated from forecasting. Hence, instead of (1), we propose the single-factor model

$$T(d) \equiv A + Cd + G(0, \sigma^2), \tag{2}$$

where again d represents the DoW factor and $G(0, \sigma^2)$ is the Gaussian random variable, while A and C are constants. Again, we set $\sum C = 0$ to avoid redundancy.

Table 4 shows the estimated coefficients for model (2), while Figure 2 shows the histogram and QQ-plot for the residuals. The coefficients C_j quantify the decreasing trend of the daily total arrivals within a week. Figure 2 shows that the normality of the residuals remains good. The ANOVA table can be computed from Table 3. The estimated variance and dispersion (variance-to-mean ratio) are

$$\hat{\sigma}^2 = \frac{10666 + 23262}{24 + 144} = 202.0 \quad \text{and} \quad D \equiv \frac{\hat{\sigma}^2}{\hat{m}} = \frac{202.0}{133.8} = 1.51,$$

where $\hat{m}$ is the estimated mean, which again represents a moderate level of is overdispersion relative to an NHPP.

3.2. Dependence Among Daily Totals and Residuals

We also examined the dependence among the residuals in the single-factor model. we first directly estimated the autocorrelation function and found the the first seven coefficients were all positive. We then fit and compared autoregressive AR(p) models, and found that the fitting was not very good, but positive coefficients again indicate some positive dependence among the residuals. Nevertheless, when we performed four different statistical tests of the residuals, we found that none could reject the independence hypothesis. Finally, we also fit ARMA(p,q) models for the actual daily totals for various

Coeffitients	Estimate	SE
A	133.766	2.842
C.Sun	28.234	4.019
C.Mon	14.634	4.019
C.Tue	11.274	4.019
C.Wed	-1.366	4.019
C.Thu	-0.526	4.019
C.Fri	-23.886	4.019
C.Sat	-28.366	4.019

Table 4: Estimated regression coefficients for the single-factor model in (2). (Use dataset 3.)

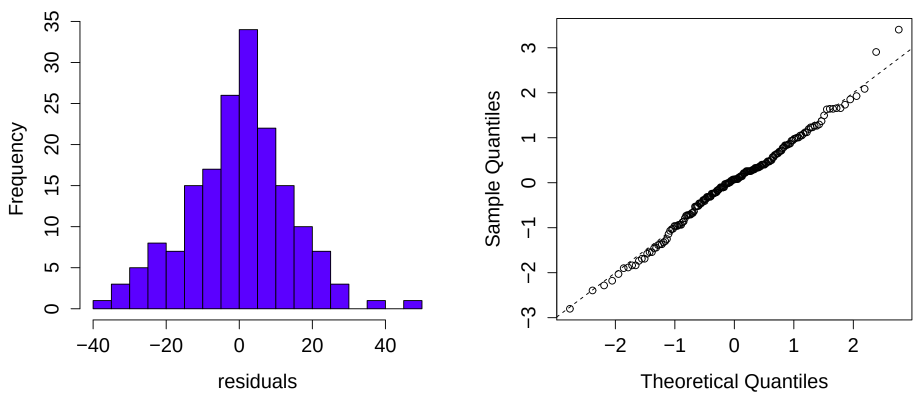


Figure 2: Supprting detail for model (2): histogram of the residuals (left) and QQ plot of studentized residuals (right). (Use dataset 3.)

p and q, with $p = 7$ being a natural choice because of the observed DoW effect. Overall, we did not find a better model to suggest. Thus the details are left to the appendix [22].

3.3. Arrival Pattern Within Days

We now estimate the time-varying arrival rate by computing hourly averages and using a piecewise-linear plot. Unlike most service systems, we find that it is important to take a week view as opposed to a day view. Thus, we combine all the 25 weeks and estimate the hourly arrival rate over a week, as is shown in Figure 3. The vertical dashed lines are at midnight between successive days. Figure 3 shows that the arrival rate is lowest in the early morning, about 6am, and increases rapidly to a peak just before noon, after which it declines irregulary, with a steep decline around midnight. As expected, the arrival rate is lower at night than during the day. We can also see that the arrival rate is lower on weekends and has a somewhat different pattern.

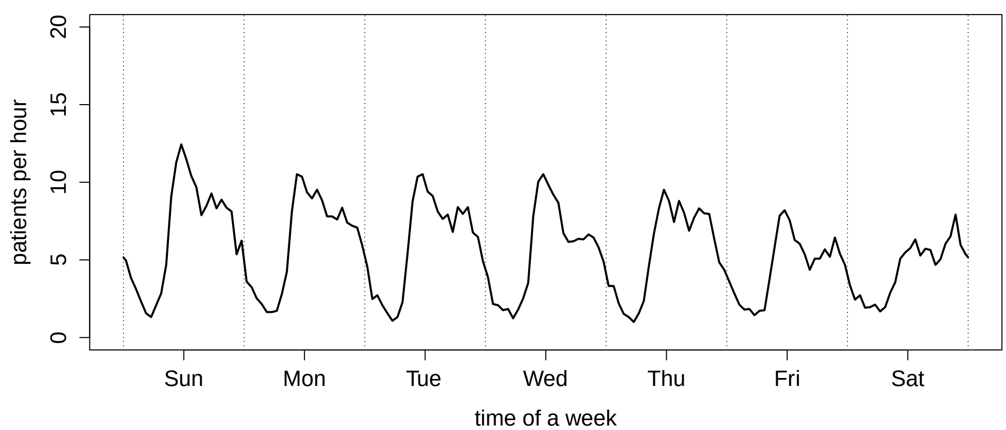
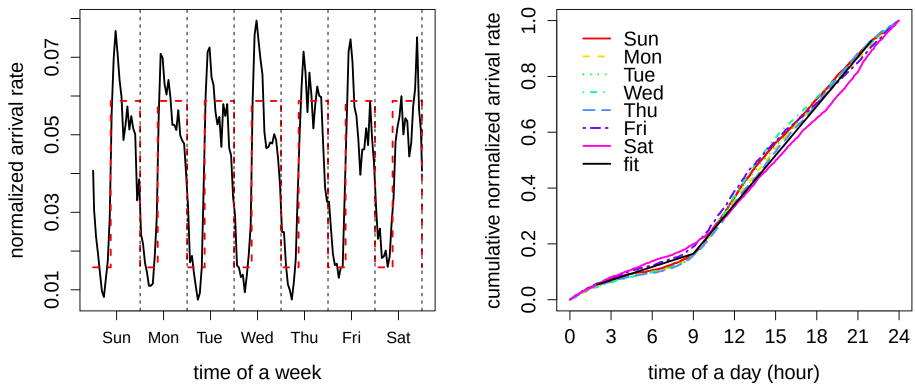


Figure 3: Estimated arrival rate at the ED over a week. (Use dataset 3.)

Since we have demonstrated a strong DoW effect on the daily totals, it is natural to examine the daily pattern without the DoW effect. To do so, we can normalize the arrival rate by the daily totals; i.e., we divide the arrival rate in Figure 3 by the average daily total arrivals of each day of week. Figure 4 (left) shows the arrival rate after normalizing, while Figure 4 (right) shows the corresponding estimated cumulative arrival rate function. Figure 4 shows that the normalized arrival rates still look different for different days, but we see more regular behavior with the cumulative view. Figure 4 suggests that it should not be unreasonable to approximate the arrival rate by a lower constant rate from midnight to 9am and a higher constant rate from 9am to midnight. This relatively simple arrival rate model is appealing, but we found that it did not perform as well in simulation comparisons.

Figure 4: Estimated normalized arrival rate function over the week (left) and the corresponding estimate cumulative arrival rate function (right). These are both compared to a piecewise-constant approximation with two pieces divided at 9 am and midnight.



3.4. Stochastic Variability in the Time-Varying Arrival Process

It is commonly accepted that the arrival process to an ED can be modeled by a nonhomogeneous Poisson process (NHPP), because the arrivals typically come from the independent medical incidents of many different people, each of whom uses the ED infrequently. Mathematical support is provided by the Poisson superposition theorem; e.g., §11.2 of [24], but that should be verified, as in [16, 17].

Indeed, we have already seen strong stochastic variation in the daily totals that suggests overdispersion relative to a Poisson process. To illustrate unsuspected bunching of arrival that can occur, anecdotally from New York, ED employees report surges of arrivals at public transportation arrival times at the hospital.

Accordingly, we investigated the stochastic variability in the arrival process by (i) estimating the index of dispersion for counts, as in [25, 26], and by performing statistical tests of the NHPP property as in [16, 17]. We briefly summarize the results of our investigations and refer to the appendix for more details.

3.4.1. The Index of Dispersion for Counts

The index of dispersion for counts (IDC) is the ratio of the variance to the mean of the arrival counting process, as a function of time. Let $A(t)$ be the number of arrivals in interval $[0,t]$, so that $\{A(t),: t \geq 0\}$ is the arrival counting process. Let $\Lambda(t) \equiv \mathbb{E}[A(t)]$ and $V(t) \equiv Var(A(t))$ be the mean and variance functions. Then the IDC is $I(t) \equiv V(t)/\Lambda(t)$, $t \geq 0$.

It is instructive to consider three different views: (i) the week view, (ii) the day view and (iii) the DoW view. In the week view we take $T = 7 * 24 = 168$ hours, and estimate $\Lambda(t)$ and $V(t)$ hourly by taking the 25 weeks as samples, then compute the ratio to estimate $I(t)$. In the day view we take $T = 24$ hours, and take the $25 * 7 = 175$ days as samples. In DoW view we take $T = 24$ hours, and take each specific day of week in the 25 weeks, so that the sample size is 25 for each day of week. Notice that it is natural to regard successive Tuesdays as i.i.d., but not successive days, so that the DoW view is likely to have less dependence.

Figure 5 shows estimates of the IDC in all three views. In both the week and day views IDC is steadily increasing, which reveals dependence over multiple days. In contrast, in the DoW view the IDC is much more flat, at a level that is not much greater than 1 for Poisson. The DoW view in Figure 5 shows that the average IDC is about 1.5, which coincides with the regression result for the daily total arrivals in §3.1. Figure 5 provides strong evidence that the overall arrival process is not too well modeled as an NHPP, but is quite well modelled as a conditional NHPP, where the arrival process conditional on the daily total is regarded as an NHPP. As explained in §3.2 of [27], that means that, after we condition on the daily total, those arrival times can be regarded as i.i.d. random variables over the day, each having a pdf proportional to the time-varying arrival-rate function. Our analysis is consistent with the conclusions in [11] and in §3.2.1 of [7], but very different in detail.

3.4.2. Statistical Tests of the NHPP Property

To statistically test the deviations from the conditional NHPP assumption, we used the statistical tests in [16, 17], in particular, the conditional uniform KolmogorovSmirnov test (CU KS test) and the Lewis KS test. The test results are shown in the appendix. The results indicate that most intervals passed these KS tests, indicating that it is reasonable to regard the arrival processes as NHPP within each day. As emphasized in [16], that does not imply that the arrival-rate function should be regarded as deterministic. Instead, it supports the conditional NHPP property, because these statistical cannot distinguish between the conditional NHPP and the direct NHPP when the separate days are analyzed separately, as in the DoW view in the

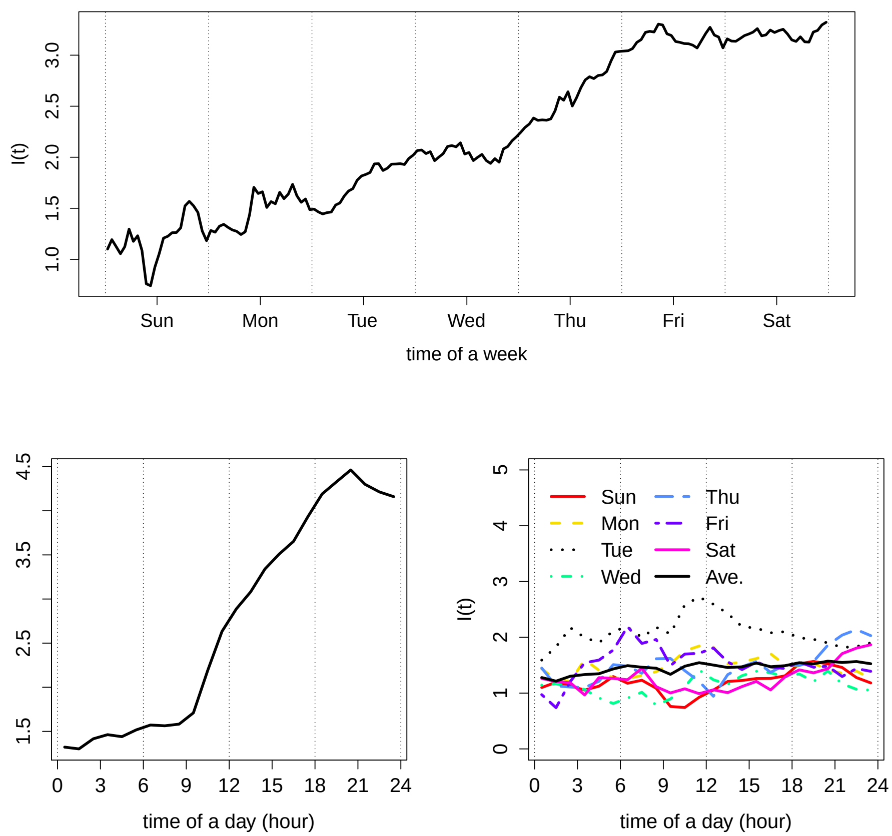


Figure 5: The estimated IDC in a week view (top), day view (bottom left) and DoW view (bottom right). (Use dataset 3.)

previous subsection.

In summary, we propose a two-time-scale model that has random daily totals and, conditional on those totals, assumes that the arrival process within each day is an NHPP. The conditioning feature means that, conditional on the daily totals, that number of arrivals is modeled as i.i.d. random variables over the entire day, each having a pdf proportional to the arrival rate function. We use M_t^T to denote this two-time-scale conditonal NHPP arrival process, where T denotes conditioning on the daily totals. A variant of this M_t^T arrival process model was proposed for appointment-generated arrival processes in [15]. For appointment-generated arrival processes, the arrival process tended to be under-dispersed compared to a Poisson process.

3.5. Arrival Processes of the Two Groups: Admitted and Non-Admitted

In §2 we mentioned that the patients in ED can be divided into two groups according to the admission decision. The non-admitted patients are released after being treated in the ED while the admitted ones are transferred to the IW's in the main hospital. A priori, we judge that these two arrival processes can be regarded as an independent thinning from the whole arrival process. for managing ED's, we wanted to investigate if this thinning might be time-dependent. Figure 6 shows the estimated arrival rates of the admitted and non-admitted patients for a week.

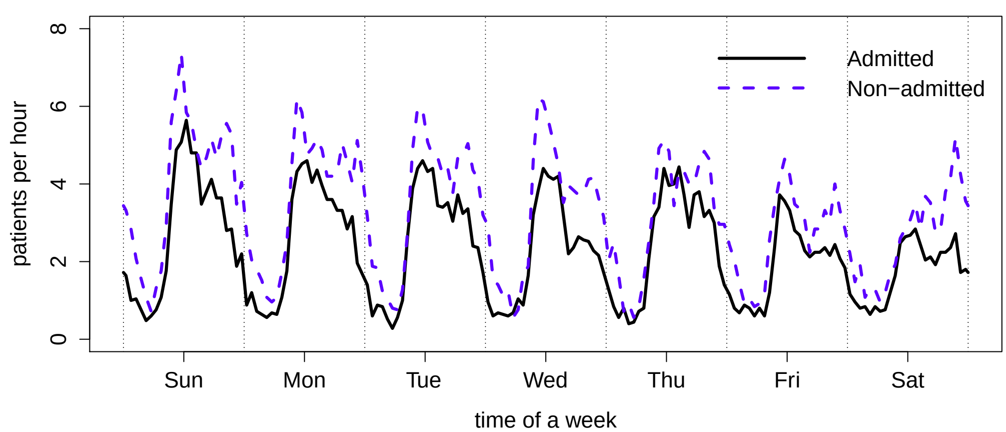
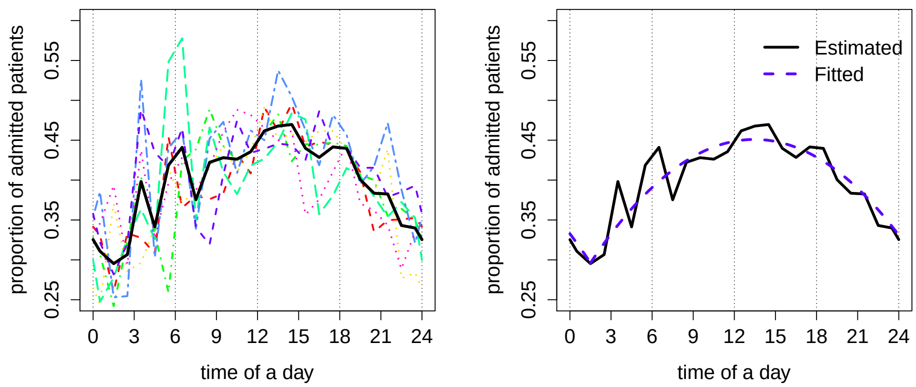


Figure 6: Estimated arrival rates for the admitted and non-admitted patients. (Use datasets 5 and 6.)

We also looked at the proportion of admitted patients as a function of time. Figure 7 shows estimates of the proportion of admitted patients by time of day over a single day, using all 175 days. Figure 7 presents strong evidence that the probability of admission is indeed time-varying. From a modelling perspective, it is significant that time-dependent, but stochastically independent, thinning also preserves the NHPP property; i.e., if A is an NHPP, then the two separate arrival processes will be NHPP's as well; see Proposition 2.3.2 of [28].

Figure 7: Estimated proportion of admitted patients as a function of the arrival time within each DoW plus the overall average shown by the black solid line (left) and compared to the fitted quadratic function (right). (Use datasets 5 and 6.)



Furthermore, we use least squares to fit a quadratic function to $p(t)$ with a maximum at 2:30 pm. Figure 7 (right) shows fitted function, which is $\hat{p}(t) = -0.001082(x - 13.5)^2 + 0.451996$, where $x = ((t - 1.5) \texttt{ mod } 24) + 1.5$ and $t \in [0, 24]$. The modulus function is used to treat the data as periodic with a daily cycle.

3.6. Summary: Full Model of the ED Arrival Process

We combine the analysis in the previous subsections to develop a full arrival process model that can be used in simulation studies. First, the daily totals for the number of arrivals are modeled as independent random variables with a Gaussian distribution, as determined by the single factor Gaussian model in (2). Then, given the daily totals, the arrival process is modelled as an NHPP, which means that the given random daily number of arrivals are treated as i.i.d. random variables over the entire day with a pdf proportional to the estimated arrival rate function for that day. We refer to that arrival process model as M_t^T. Finally, a patient that arrives at time t is admitted with probability $p(t)$, estimated by the quadratic function above. We conduct simulation experiments using the model in §6.

4. Length of Stay

In this section, we investigate the patient LoS distribution. We first find that the LoS distribution should be regarded as time-varying. Then we introduce a discrete-time analysis to expose the structure in more detail.

4.1. Failure of the $G_t/GI/\infty$ Aggregate Model

It is common to directly examine the LoS distribution, as if that should be a natural primitive. For modeling, that means that the LoS of successive patients would be modeled as i.i.d. random variables with that distribution. Given that perspective, we started by estimating the overall LoS distribution; we refer to the appendix for the details. That was accomplished by looking at the difference between the exit time and entrance time of each patient. A more elaborate model of what happens in between arrival and departure was not possible, because such extra information was not included in the data.

It also turns out to be highly significant that the departure or exit time was defined as the time that the ED doctor made the decision whether or not to admit the patient. Thus, for admitted patients, the additional time until the transfer to the Internal Ward (IW) was not included in the data. Thus, we were unable to directly study the important problem of ED boarding (the extra delay between the admission decision and the patient being transferred to the IW).

Given that perspective, a natural aggregate model for the ED would be an $M_t/GI/\infty$ or $G_t/GI/\infty$ infinite-server queue, combining an arrival process with a time-varying arrival-rate function with the patient LoS modeled by a sequence of independent and identically distributed (i.i.d.) service times with a general cumulative distribution function (cdf) G.

To see if these models with GI LoS times are approximately appropriate, we calculated the time-varying departure rate $\delta(t)$ and the mean occupancy level $m(t) \equiv E[Q(t)]$ in the $G_t/GI/\infty$ model using Theorem 1 of [19] together with the estimated arrival-rate function $\lambda(t)$ and LoS cdf G. (As emphasized by §5 of [29], these formulas apply to G_t as well as M_t arrivals, and so also apply to M_t^T arrivals, as assumed in §3.6.) Figure 8 compares the directly estimated departure rate and mean occupancy to the indirect estimator exploiting the model. Figure 8 shows that the model with GI LoS does not nearly approximate the actual departure rate and mean occupancy levels. Especially striking is the surge in departures at the end of the day, around midnight, which is totally missed by the $M_t/GI/\infty$ model. In closing, we remark that Figure 8 parallels Figure 3 in [7]. There it is emphasized that the peak occupancy lags after the peak arrival rate, which can be seen from Figure 8 as well.

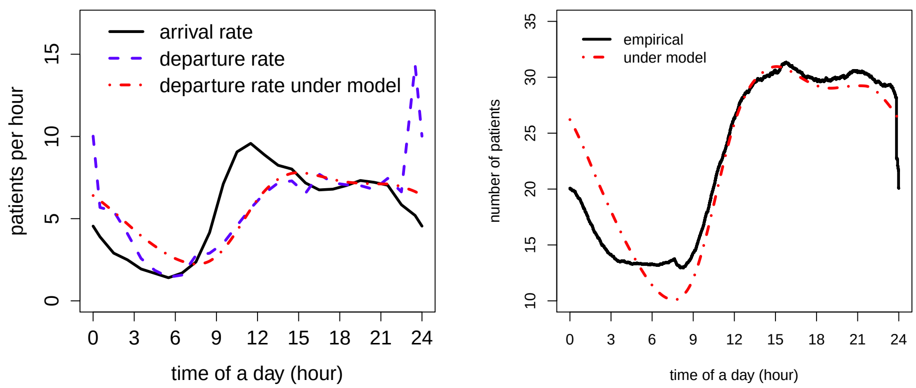


Figure 8: A comparison of direct estimates of the time-varying departure rate $\delta(t)$ (left) and the mean occupancy level $m(t)$ (right) at the ED to indirect estimates based on the $M_t/GI/\infty$ model using the estimated arrival rate and LoS ecdf (Use dataset 3.)

4.2. The Time-Varying LoS Distribution

To directly see the time-varying structure of the LoS distribution, we looked at a box-plot of the LoS for each hour; see [30] for backgound. The time-varying behavior of the LoS can be seen from a week view (see the appendix), but is especially clear in a day view, as shown in Figure 9. The boxes show the 25% and 75% percentiles, while the blue diamonds are the means and the black bars are the medians. Consistent with intuition, the LoS is longer for patients arriving after midnight, when there tends to be fewer staff. The LoS also tends to be somewhat less for arrivals in the evening. This may be explained by extra effort to release non-admitted patients by midnight, which we will discuss soon.

Given the time-dependence in the LoS distribution, we decided to do a careful analysis in discrete time. For that purpose, let $X_{k,j}$ be the number of arrivals in discrete time period k that have a LoS of j time periods, i.e., that depart in discrete time period $k + j$, $j \geq 0$. We let a discrete time period be one hour. We still focus on the time period from Dec. 5, 2004, to May

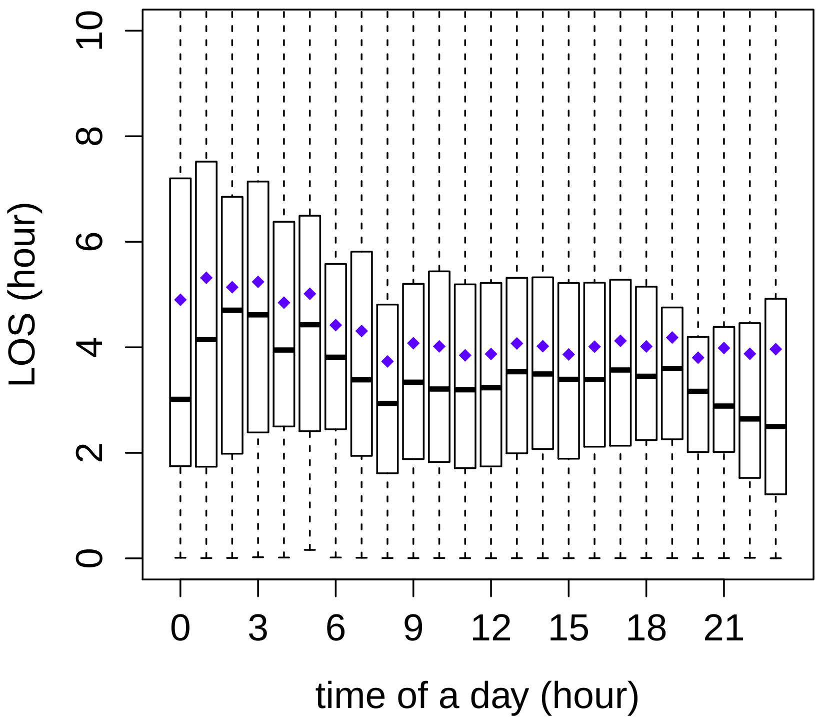


Figure 9: A box plot of the LoS distribution by hour of the day. The blue diamonds are the means, while the black bars are the medians (Use dataset 3.)

28, 2005 (25 weeks, 175 days), so we make 00:00-01:00 on Dec. 5 2004 to be discrete time period $k = 1$. In order to have the correct number of patients in the system, we only count patients who entered or left the system in that time period and have a reasonable LoS; i.e. we use dataset 4 through this part. We have to include 1 extra day (Dec. 4, 2004) at the beginning to let X include all the patients we focused on. The longest LoS is less than 37 hours after cleaning the data. So our X matrix has dimension $(24*176)\times 37$, i.e. for $X_{k,j}$, $-23 \le k \le 24*175$, $0 \le j \le 36$. The full X matrix is displayed in the appendix.

Let A_k and D_k be the number of arrivals and departures in the time period k. Then we have

$$A_k = \sum_{j=0}^{36} X_{k,j} \quad \text{and} \quad D_k = \sum_{j=0}^{36} X_{k-j,j},$$

where we assume $X_{k,j} = 0$ for all k, j except $-23 \leq k \leq 24 * 175$, $0 \leq j \leq 36$.

Now we assume a periodic structure over successive periods of d discrete times. We assume that we have sufficient data to estimate averages over n periods, containing nd discrete time periods. Specifically, if we consider a period to be 1 week, then we have $n = 25$ and $d = 7 * 24$; if we consider a period to be 1 day, then we have $n = 175 = 7 * 25$ and $d = 24$.

In this periodic setting, we construct averages. In particular, let

$$\bar{A}_k = n^{-1} \sum_{m=1}^{n} A_{(m-1)d+k}, \;\; \bar{D}_k = n^{-1} \sum_{m=1}^{n} D_{(m-1)d+k}$$

and

$$\bar{X}_{k,j} = n^{-1} \sum_{m=1}^{n} X_{(m-1)d+k,j},$$

for $1 \leq k \leq d$ and $0 \leq j \leq 36$. The $\bar{X}$ matrix for $d = 24$ is shown in the appendix. Table 5 shows part of the the transpose of the $\bar{X}$ matrix; i.e., the entry in row j and column k is the proportion of all arrivals in hour k who had a LoS equal to j hours, so that the bold values correspond to the surge just before midnight

To make the structure more evident, We show some of the cells shadowed and bold. Those diagonally arranged cells correspond to the proportion of patients that arrived in the column hour whose row value of LoS made them depart from the ED in the hour before midnight. Table 5 shows that many patients depart from the ED just before midnight. For example, consider the arrival in hour (column) 10. The discrete LoS probability mass function increases from $j = 1$ to $j = 2$, but then decreases to the low value 0.023 at $j = 13$ before jumping up to 0.234 at $j = 14$, a value 10 times higher, before devlining rapidly toward 0.

Again, we emphasize that the data we used only provides the entry time and exit time for each patient, where the exit time is when the ED doctor made the admission decision. Evidently there is a change in medical staff at midnight that increases the number of admission decisions just before midnight.

	7	8	9	10	11	12	13	14	15
1	0.131	0.246	0.491	0.817	1.069	1.263	1.194	1.006	0.680
2	0.234	0.366	0.709	1.194	1.560	1.543	1.446	1.211	1.114
3	0.314	0.417	0.754	1.177	1.691	1.554	1.286	1.343	1.549
4	0.263	0.337	0.623	1.040	1.114	1.154	1.331	1.257	1.171
5	0.263	0.171	0.320	0.669	0.703	1.257	0.840	1.011	0.874
6	0.189	0.194	0.217	0.411	0.697	0.657	0.583	0.594	0.651
7	0.091	0.120	0.154	0.400	0.394	0.366	0.474	0.423	0.343
8	0.023	0.051	0.171	0.246	0.274	0.257	0.263	0.211	0.211
9	0.029	0.103	0.074	0.131	0.211	0.206	0.211	0.137	**0.446**
10	0.006	0.051	0.017	0.109	0.149	0.097	0.069	**0.383**	0.051
11	0.023	0.017	0.023	0.080	0.051	0.086	**0.269**	0.034	0.051
12	0.017	0.034	0.029	0.029	0.046	**0.246**	0.051	0.011	0.011
13	0.011	0.017	0.006	0.023	**0.366**	0.006	0.029	0.011	0.006
14	0.000	0.006	0.000	**0.234**	0.011	0.023	0.011	0.000	0.000
15	0.006	0.006	**0.126**	0.006	0.000	0.017	0.000	0.000	0.011
16	0.000	**0.057**	0.000	0.006	0.006	0.006	0.011	0.017	0.011
17	**0.034**	0.000	0.000	0.006	0.000	0.006	0.006	0.017	0.011
18	0.000	0.006	0.000	0.000	0.000	0.000	0.000	0.017	0.011

Table 5: Part of the transpose of the $\bar{X}$ matrix; i.e., the entry in row j and column k is the proportion of all arrivals in hour k who had a LoS equal to j hours, so that the bold values correspond to the surge just before midnight (Use dataset 4).

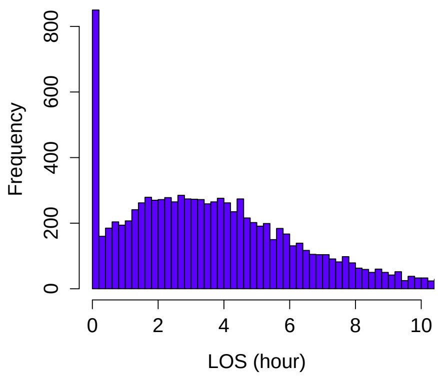
4.3. The LoS of the Two Groups

Just as for the arrival process, we want to study differences in the LoS distribution for the admitted and non-admitted patients. Figure 10 shows the empirical LoS distribution for the two groups without time structure. The admitted patients have a smaller mean LoS but a longer median, because about 7% of the admitted patients have an extremely low LoS. Evidently, these patients were transfered immediately to the IW's. If we omit the admitted patients whose LoS is less than 2 minutes (657 patients), then the mean LoS of the admitted group increases to 4.30 hours, which is larger than the non-admitted group.

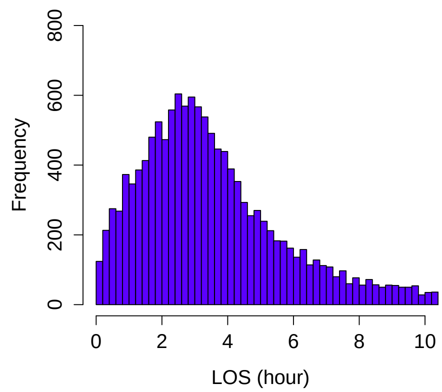
Then we look at the time-varying feature of the LoS for the two groups, again using box plots. Figure 11 shows that the time-varying LoS distribution is more regular for the non-admitted patients. We see quite striking differences for admitted patients before and after midnight.

4.4. The LoS Model and Occupancy

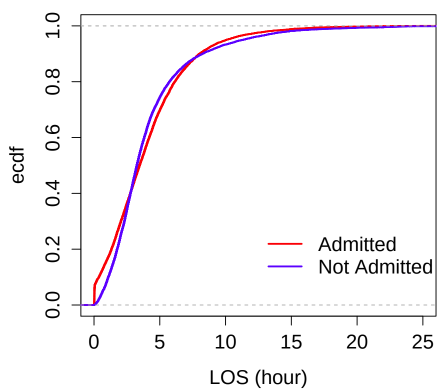
Our analysis of the LoS data, leads us to model the LoS distribution as (i) time-dependent and (ii) depending on whether the patient is admitted or not. If we use the M_t^T two-time-scale arrival process model in §3.6 and ignore the distinction between the admitted and non-admitted patients, this



(a) LoS of admitted patients. (x-axis limited to 10 hours)



(b) LoS of non-admitted patients. (x-axis limited to 10 hours)



(c) empirical cdf's for the 2 groups.

Figure 10: Estimated LoS distributions of the admitted and non-admitted patients truncated to $[0, 10]$. (Use datasets 5 and 6.)

produces an $M_t^T/GI_t/\infty$ infinite-server aggregate model. Extending it to the two types of patients, the model becomes two independent $M_t^T/GI_t/\infty$ models, again using the arrival process model from §3.6, one for the admitted patients and another for the non-admitted patients. We would use the separate time-varying LoS distribution for each group. In conclusion, we note that this pair of independent infinite-server models does not capture

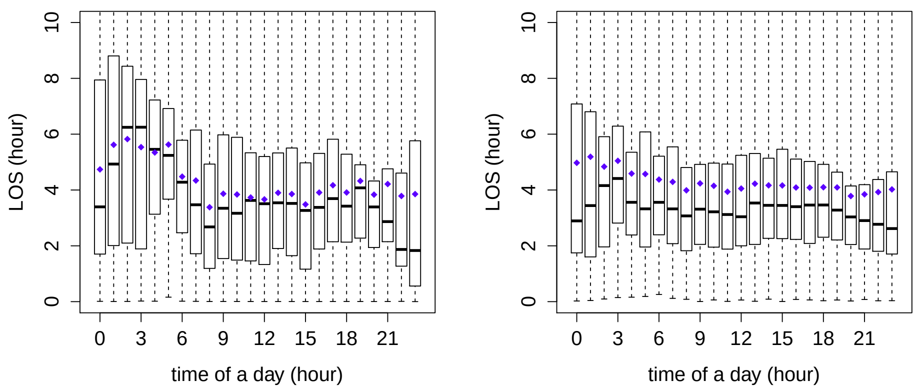


Figure 11: Box plots of the LoS distribution as a function of the arrival time for admitted (left) and non-admitted (right) patients. (Use datasets 5 and 6.)

the congestion one group causes for the other. Nevertheless, even though an infinite-server model was not suggested in [7], this model is consistent with several observations in [7]. First, in §3.1 [7] the authors emphasize that the bed capacity of the ED is highly flexible, so that there is effectively unbounded. Second, in Figures 4 and 5 in §3.2.2 of [7] the authors observe that a time-varying Gaussian distribution fits the occpancy data well, but that is consistent with the theoretical time-varying Poisson distribution in the time-varying $M_t/GI/\infty$ model and the heavy-traffic Gaussian approximations for infinite-server models in [31].

5. The Departure process

In this section, we investigate the departure process from the ED. As a theoretical reference point, for an the $M_t/GI_t/\infty$ model, the departure process is also an NHPP. We find it useful to look at the departure process and the entire ED in reverse time, so that we can think of the departure process as an arrival process and use the same methods we have used in previous sections. That reverse-time perspective is especially revealing to look at the time-varying proportion of admitted patients and the time-varying LoS, where the time refeers to the departure time instead of the arrival time.

5.1. Daily Totals

Paralleling §3.1, we first look at the daily totals of departures, but we provide only a brief overview; see the appendix for the tables and figures.

The reverse time perspective forces us to change the data a little. Now we consider the patients that left the from Dec. 5, 2004, to May 28, 2005, which is 23,407 patients in total (see Table 1). The mean values for each week and each DoW are almost the same as for the arrivals, but there is a significant difference in the variances. The variance of the total numbers of departures by DoW is higher than for the arrivals. Evidently, there is less regularity in departures than in arrivals.

Again, we fit the Gaussian regression models in (1) and (2) in §3.1 for the departures. The parameters have the same meaning as before. Table 6 shows the ANOVA results. As before, both the Week factor and the DoW

Factor	Sum of square	df	Mean sum of square	F statistics	P-values
Week	10661	24	444.2	2.19	<0.01
DoW	56146	6	9357.6	46.22	<0.01
Residuals	29156	144	202.5		

Table 6: ANOVA table for the two-factor model (1) for the departures. (Use dataset 7.)

factor are statistically significant, but the DoW factor explains most of the variance. For the two-factor model, the mean sum of square for the residuals is $\hat{\sigma}^2 = 202.5 = 14.23^2$, which is higher than that of the arrival process. The variance-to-mean ratio is $202.5/133.8 = 1.51$. If we omit the Week factor and consider the single factor model. Then the mean sum of square for the residuals is $(10661 + 29156)/(24 + 144) = 237.0$ and the variance-to-mean ratio is $237.0/133.8 = 1.77$.

5.2. Departure Pattern Within Each Day

Now we turn to the time structure of departure rate within days. Figure 12 shows the reverse-time view. Paralleling and amplifying Figure 8, Figure 12 shows clearly that the departure rate has midnight surges and that the peaks are increasing over the week.

As before, we divided the patients into two groups according to the admission decision. (See Table 1 for basic statistics.) Figure 13 shows the time-varying proportion of admitted patients as a function the departure time. We see that the proportion of admitted patients is extremely low at 7-8 a.m. of each day. Evidently, admission decisions at that time are postponed until new doctors arrive after morning staff changes.

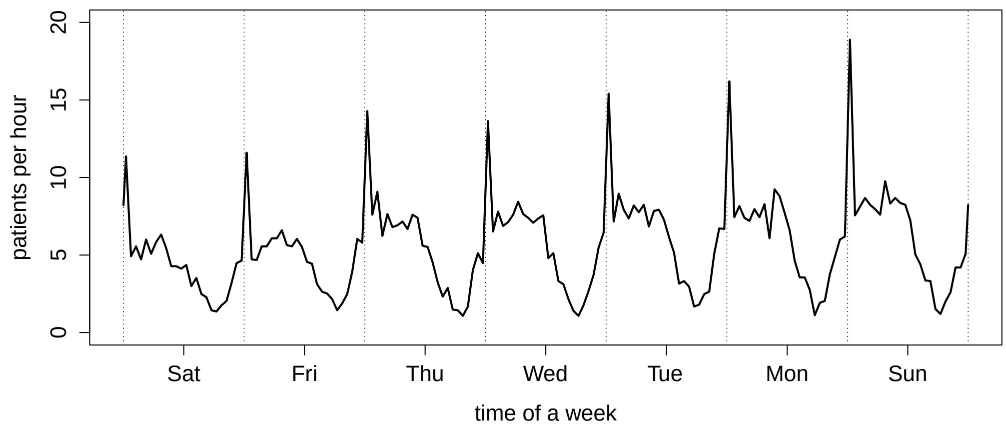


Figure 12: Estimated departure rate at the ED in reversed time. (Use dataset 7.)

Figure 14 presents box plots of the LoS distribution as a function of the departure time (in reverse time) for admitted (left) and non-admitted (right) patients. We see that the midnight surge is caused by the non-admitted patients, and that the LoS of non-admitted patients are more influenced by time of the day.

6. Comparison with Simulation

In this section we conduct simulations to substantiate our model.

6.1. Comparing Alternative LoS Models

We conduct simulation experiments with our model to see how it represents the data. First, we focus on the LoS model. To do so, we use the original arrival data. We repeat the 25 weeks 40 times, so that the sample size is 1000 weeks. To examine alternative LoS models, we treat them in three different ways: (A) The first option is GI, i.e., we assume that the LoS distribution is not time-varying; we use the overall estimated cdf; (B) The second option is GI_t but with a day view; i.e., we assume that the LoS distribution is time-varying over each day; we use the estimated time-varying cdf depending on the arrival time within a day; (C) The third option is also

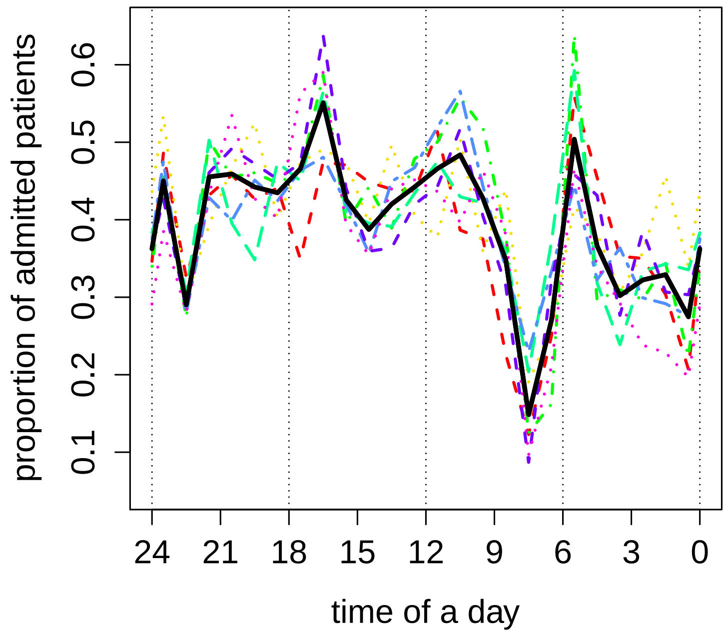


Figure 13: Estimated time-varying proportion of admitted patients as a function of the departure time from the ED over a day, for each DoW and overall, combining all days together (black solid line). (Use datasets 8 and 9.)

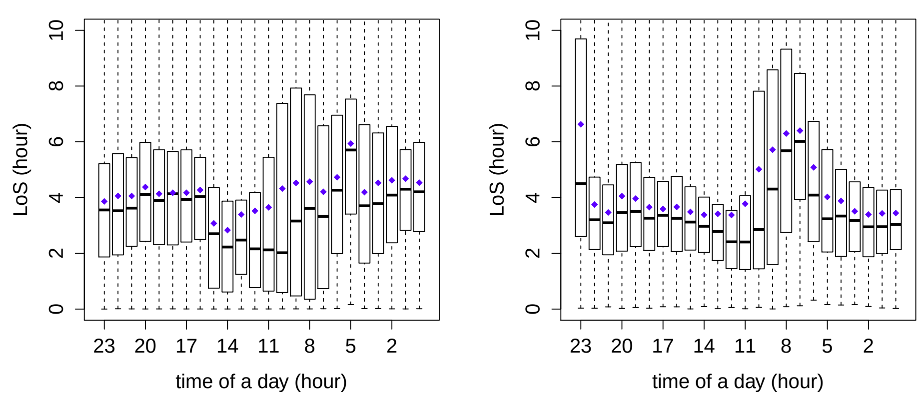


Figure 14: Box plots of the LoS distribution as a function of the departure time (in reverse time) for admitted (left) and non-admitted (right) patients. (Use datasets 5 and 6.)

GI_t but with a week view; i.e., we assume that the LoS distribution is time-varying over each week; we use the estimated time-varying cdf depending on the arrival time within a week.

Figures 15 and 16 compare the indirect model estimates to direct simulation estimates of the time-varying expected occupancy $m(t)$ and the time-varying departure rate $\delta(t)$, respectively, based on each of these three LoS models.

The top plots of Figures 15 and 16 show the consequence of ignoring the time-varying LoS distribution. Consistent with Figure 8, Figure 15 shows that the GI LoS model significantly underestimates the occupancy at the end of the day, before midnight, and overestimates it at the beginning of the day, after midnight, while Figure 16 shows that the GI LoS model completely misses the midnight surge of departures.

The middle plots (B) of Figures 15 and 16 show that the GI_t LoS model with a day view does much better than the GI model, capturing the midnight surge in departures. Nevertheless, there is a clear gap between the mean occupancy curves. Remarkably, the bottom plots (C) of Figures 15 and 16 show that the GI_t LoS model with a week view show near-perfect agreement.

6.2. Evaluating the Full Model

We obtain a full ED model when we (i) incorporate the M_t^T arrival model summarized in §3.6, (ii) divide the arrivals into the two groups, admitted and non-admitted, using independent thinning according to the time-varyiing p[robability $p(t)$ estimated in §3.5, and (iii) when we use a separate LoS model for each group.

We repeated the three experiments Figures 15 and 16 using the full model. We applied the three LoS models to each group separately. The new simulation results look virtually identical to Figures 15 and 16, and so they are only shown in the appendix.

6.3. The Time-Varying Little's Law

The spectacular agreement between the simulation estimates for case (C) were initially puzzling. However, we find that this can be explained in large part by the time-varying Little's law (TVLL), as in [20, 21]. The TVLL Little's law applies to a $G_t/G_t/\infty$ model and thus to our $M_t^T, GI_t/\infty$ model. The discrete-time study in this paper motivated us to also consider a discrete-time version of the TVLL. We intend to discuss the discrete-time TVLL and the implications of the TVLL in [32].

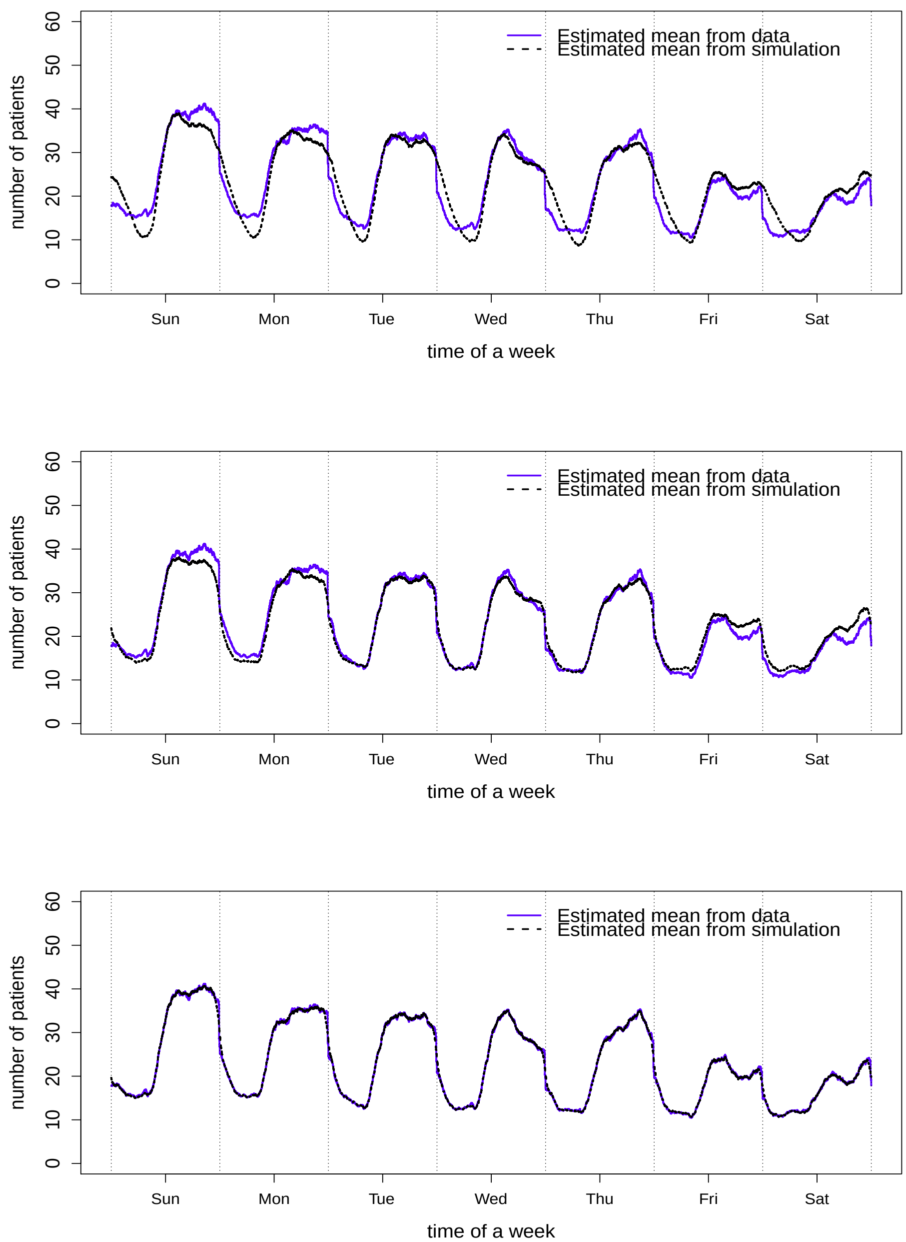


Figure 15: Simulation estimates of the time-varying expected occupancy $m(t)$, based on the arrival data plus three LoS models: (A) GI (top), (B) GI_t with day view (middle) and (C) GI_t with week view (bottom)

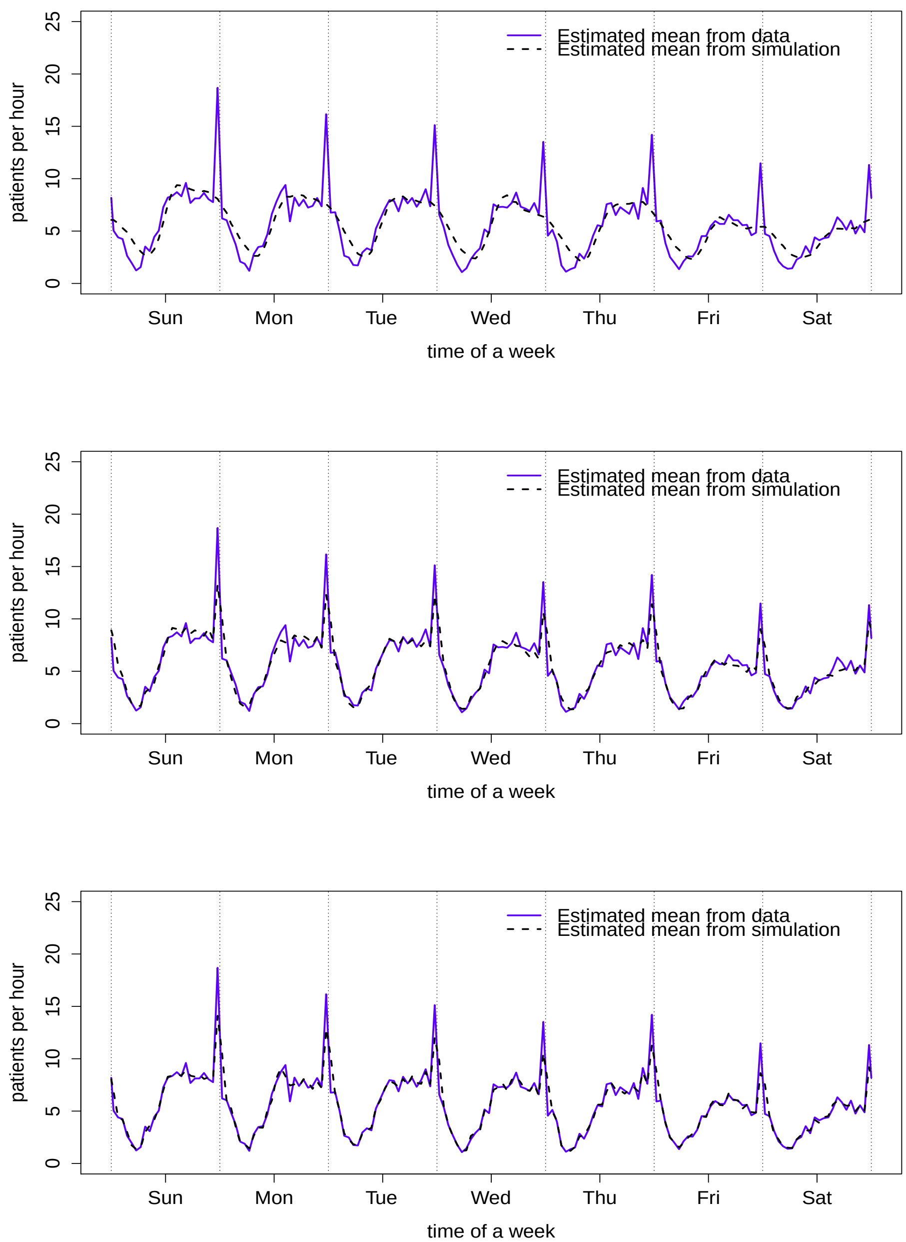


Figure 16: Simulation estimates of the time-varying departure rate $\delta(t)$ based on the arrival data plus three LoS models: (A) GI (top), (B) GI_t with day view (middle) and (C) GI_t with week view (bottom)

7. Conclusions

We studied a 25-week portion of the ED data used in the patient flow study by Armony et al. [7]. We carefully studied the arrival process to the ED and the patient LoS distribution, reaching several important conclusions.

First, for the arrival process, we think that it is helpful to use the two-time-scale approach, in which we first look at daily totals and then look at the arrival process within each day conditional on the daily totals, which leads to the arrival process model summarized in §3.6. In §3.1 we examined factor regression models for the daily totals, and adopted the single-factor model in (2), which expresses the daily totals as an expected value depending on the day of the week (DoW) plus a mean-0 Gaussian distribution with a variance that is determined by the regression. This directly leads to a model of independent daily totals with a Gaussian distribution depending on the DoW. The two-time-scale model is useful, because it provides a useful framework for future research. It is natural to next look for improvements to the model of daily totals by exploiting (i) time-series models, (ii) forecasting methods and (iii) more context knowledge to capture the dependence in successive daily totals. Preliminary investigation revealed positive dependence among the residuals, as indicated in §3.2 and expanded upon in the appendix [22]. With further work, it may be possible to capture the dependence over multiple days shown in Figure 5 (first two plots).

We studied the time-varying arrival rate in §3.3. We concluded that it is important to take a week view, as shown in Figure 3, rather than the common day view. An important new finding is the dependence of the admission decision on the time of arrival, discussed in §3.5. Even though the admission decision cannot be known in advance for individual patients, we can exploit the time-dependence in the observed admission decisions to model these two groups of patients differently. Finally, we examined the stochastic variability in the arrival process in §3.4 and found support for the two-time-scale model, where conditional on the daily totals, the arrival within the day can be modeled as an NHPP. We denote this arrival process as M_t^T.

Second, we analyzed the patient length-of-stay (LoS) distribution in §4. We concluded that this too should depend on the arrival time. Figures 8, 15 and 16 dramatically show the consequence of ignoring this time-varying feature. Of course, it is desirable to do a more detailed modeling of the flow within the ED, presumably with a queueing network model, so that the overall LoS distribution can be analyzed through its component parts, but the available data did not permit that. Even after that is done, an aggregate model should be helpful for comparison.

Combining the arrival process model in §3 and the LoS model in §4, we obtain the proposed $M_t^T/GI_t/\infty$ time-varying infinite-server aggregate model of the ED. This model becomes expanded to two independent such infinite-server models if we separately model the admitted and non-admitted patients, with independence following from the independent thinning of an NHPP. This model can be used for capacity planning and for comparison in more detailed queueing network models of the ED.

We think it is also important to analyze the departure process from the ED, which we do in reverse time in §5. The departure rate function in Figure 12 clearly shows the midnight surge, which can be missed from other views. Figures 13 and 14 show that the admission decision and the LoS both depend on the departure time as well as on the arrival time.

Finally, we compared our model to simulation in §6. We found remarkable agreement in the average occupancy level and the departure rate, but discovered that these high-quality approximations can largely be explained by the time-varying Little's law in [20, 21], as we plan to discuss in [32].

Acknowledgement. We thank Avishai Mandelbaum, Galit Yom-Tov and their colleagues at the Technion IE&M Laboratory for Service Enterprise Engineering (SEELab) for providing access to the Israeli Rambam hospital data. We thank NSF for research support through CMMI 1265070.

References

[1] R. B. Fetter, J. D. Thompson, The simulation of hospital systems, Operations Research 13 (5) (1965) 689–711.

[2] M. L. Brandeau, F. Sainfort, W. P. Pierskalla, Operations research and health care: a handbook of methods and applications, Springer Science & Business Media, 2004.

[3] B. T. Denton, Handbook of Healthcare Operations Management, 2nd Edition, Springer, 2013.

[4] R. W. Hall, Handbook of Healthcare System Scheduling, Springer, 2012.

[5] R. W. Hall, Patient flow: reducing delay in healthcare delivery, 2nd Edition, Springer Science & Business Media, 2013.

[6] R. S. Kaplan, M. E. Porter, How to solve the cost crisis in healthcare, Harvard Business Review 89 (2011) 46–64.

[7] M. Armony, S. Israelit, A. Mandelbaum, Y. Marmor, Y. Tseytlin, G. Yom-Tov, Patient flow in hospitals: a data-based queueing-science perspective, Stochastic Systems 5 (1) (2015) 146–194.

[8] M. A. Ahmed, T. M. Alkhamis, Simulation optimization for an emergency department healthcare unit in kuwait, European Journal of Operations Research 198 (2009) 936–942.

[9] A. M. D. Bruin, A. C. van Ossum, M. C. Visser, G. M. Koole, Modeling the emergency cardiac in-patient flow: an application of queueing theory, Health Care Management Science 10 (2007) 125–137.

[10] L. V. Green, J. Soares, J. F. Giglio, R. A. Green, Using queueing theory to increase the effectiveness of emergency department provider staffing, Academic Emergency Medicine 13 (2006) 61–68.

[11] S. Maman, Uncertainty in the demand for service: The case of call centers and emergency departments, Master's thesis, Technion-Israel Institute of Technology (2009).

[12] D. Sinreich, Y. Marmor, Emergency department operations: the basis for developing a simulation tool, IIE Transactions 37 (2005) 233–245.

[13] L. V. Green, P. J. Kolesar, W. Whitt, Coping with time-varying demand when setting staffing requirements for a service system, Production Oper. Management 16 (2007) 13–29.

[14] P. Shi, M. C. Chou, J. G. Dai, D. Ding, J. Sim, Models and insights for hospital inpatient operations: time-dependent ed boarding time, Management Science 62 (1) (2016) 1–28.

[15] S.-H. Kim, P. Vel, W. Whitt, W. C. Cha, Poisson and non-Poisson properties in appointment-generated arrival processes: the case of an endrocrinology clinic, Operations Research Letters 43 (2015) 247–253.

[16] S. Kim, W. Whitt, Are call center and hospital arrivals well modeled by nonhomogeneous Poisson processes?, Manufacturing and Service Oper. Management 16 (3) (2014) 464–480.

[17] S. Kim, W. Whitt, Choosing arrival process models for service systems: Tests of a nonhomogeneous Poisson process, Naval Research Logistics 17 (2014) 307–318.

[18] M. Ramakrishnan, D. Sier, P. G. Taylor, A two-time-scale model for hospital patient flow, IMA Journal of Management Mathematics 16 (2005) 197–215.

[19] S. G. Eick, W. A. Massey, W. Whitt, The physics of the $M_t/G/\infty$ queue, Oper. Res. 41 (1993) 731–742.

[20] D. Bertsimas, G. Mourtzinou, Transient laws of nonstationary queueing systems and their applications, Queueing Systems 25 (1997) 315–359.

[21] B. H. Fralix, G. Riano, A new look at transient versions of Little's law, Journal of Applied Probability 47 (2010) 459–473.

[22] W. Whitt, X. Zhang, A data-driven model of an emergency department, appendix: Additional details, Columbia University, http://www.columbia.edu/∼ww2040/allpapers.html (2016).

[23] M. H. Kutner, C. Nachtsheim, J. Neter, Applied linear regression models, 4th Edition, McGraw-Hill/Irwin, 2004.

[24] D. J. Daley, D. Vere-Jones, An Introduction to the Theory of Point Processes, 2nd Edition, Vol. II, Springer, Oxford, U. K., 2008.

[25] D. R. Cox, P. A. W. Lewis, The Statistical Analysis of Series of Events, Methuen, London, 1966.

[26] K. W. Fendick, W. Whitt, Measurements and approximations to describe the offered traffic and predict the average workload in a single-server queue, Proceedings of the IEEE 71 (1) (1989) 171–194.

[27] S. Kim, W. Whitt, The power of alternative Kolmogorov-Smirnov tests based on transformations of the data, columbia University, http://www.columbia.edu/∼ww2040/allpapers.html (2013).

[28] S. M. Ross, Stochastic Processes, 2nd Edition, Wiley, New York, 1996.

[29] O. B. Jennings, A. Mandelbaum, W. A. Massey, W. Whitt, Server staffing to meet time-varying demand, Management Sci. 42 (1996) 1383–1394.

[30] W. Härdle, L. Simar, Applied multivariate statistical analysis, 3rd Edition, Vol. 22007, Springer, 2012.

[31] G. Pang, W. Whitt, Two-parameter heavy-traffic limits for infinite-server queues, Queueing Systems 65 (2010) 325–364.

[32] W. Whitt, X. Zhang, Data analysis and the time-varying little's law, in preparation, Columbia University, http://www.columbia.edu/∼ww2040/allpapers.html (2016).