Analysis of Data from the Israeli Emergency Department

Creating a Stochastic Model: $M_t/G_t/\infty$ (+over-dispersion)

IEOR 4615, Service Engineering, Professor Whitt,

with Xiaopei Zhang, IEOR PhD

Lecture 20, April 16, 2015

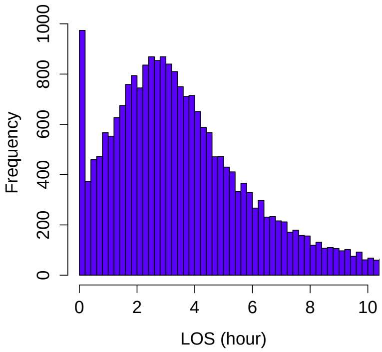
Data from an Israeli Emergency Department (ED)

- the Rambam Hospital Data from the Mandelbaum repository in the Technion SEELab
- Mor Armony, Shlomo Israelit, Avishai Mandelbaum, Yariv N. Marmor, Yulia Tseytlin, and Galit B. Yom-Tov. **Patient Flow in Hospitals: A Data-Based Queueing-Science Perspective**, Working paper, 2014. And longer longer 2011 draft. (Posted on Courseworks and in advanced survey reading on the course webpage.)
    - Like Chapter 3 in the PF paper above, we look only at the internal unit of the ED

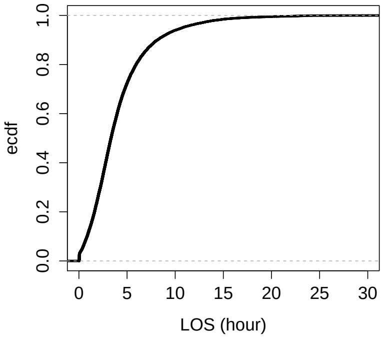
25 Weeks from December 5, 2004 to May 28, 2005

- Only internal Unit of Israeli ED
- $23,409$ patients
  - 25 weeks or 175 days
  - $23,409/175 = 133.8$ **arrivals per day**
  - Length of Stay (LoS) in hours
    - **mean =** 4.1 **hours**, standard deviation = 3.5
    - quantiles: $25\% = 1.90$, **median** $50\% = 3.31$, $75\% = 5.26$
  - Omit 12 patients with $LoS > 48$

The Length-of-Stay (LoS) Distribution

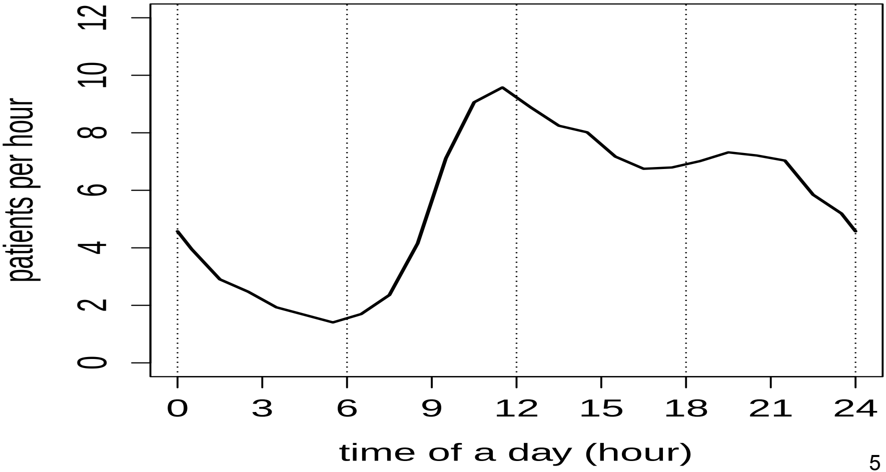


(a) LOS in limited interval



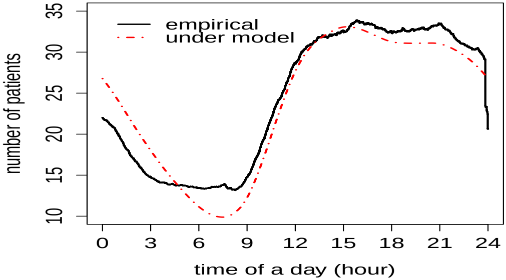
(b) ECDF of LOS

Average Arrival Rate in the ED by Hour of the Day



Quick Model: $M_t/GI/\infty$

- NHPP (M_t) arrival process with daily average arrival rate
- i.i.d. patient LoS with estimated distribution

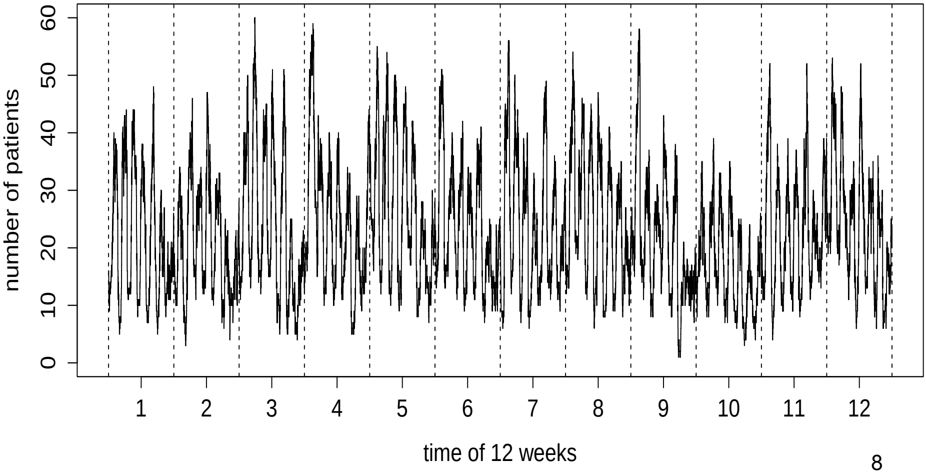


- Not too bad, but **not good enough**. Implies LoS distribution must be time varying (TV).

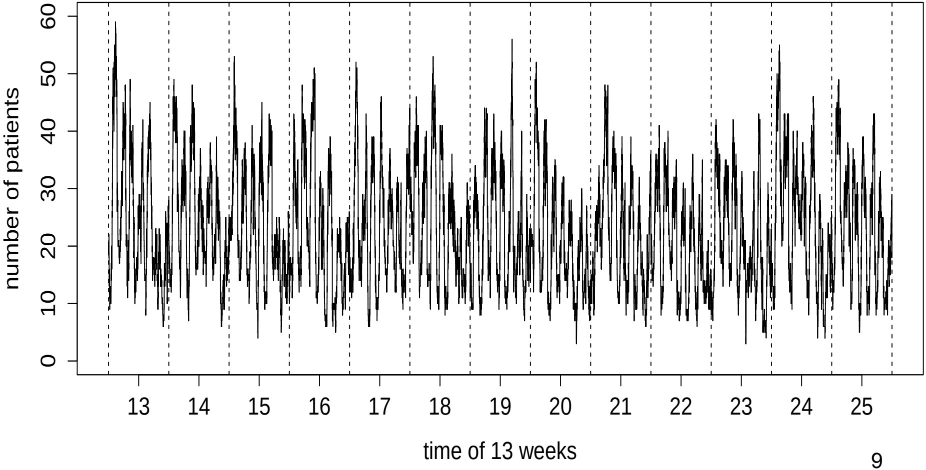
More Careful Analysis: OUTLINE

- Occupancy Levels
- Arrivals
  - Daily Totals
  - Arrival Rate Function
  - Statistical Test of a Nonhomogeneous Poisson Process (NHPP)
- Length-of-Stay (LOS) Distribution
- Odd Departures
  - LOS by Arrival Time
  - Examining the Odd Departures

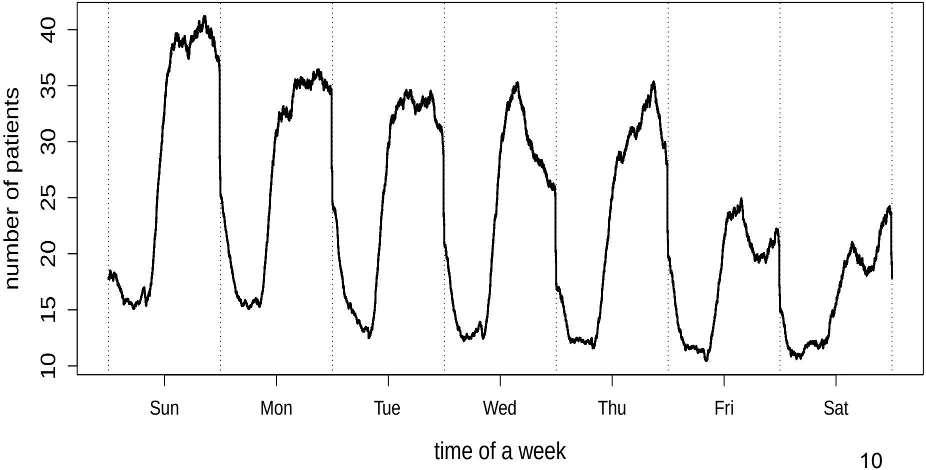
Occupancy Levels in the ED: First 12 of 25 Weeks



Occupancy Levels in the ED: Last 13 of 25 Weeks



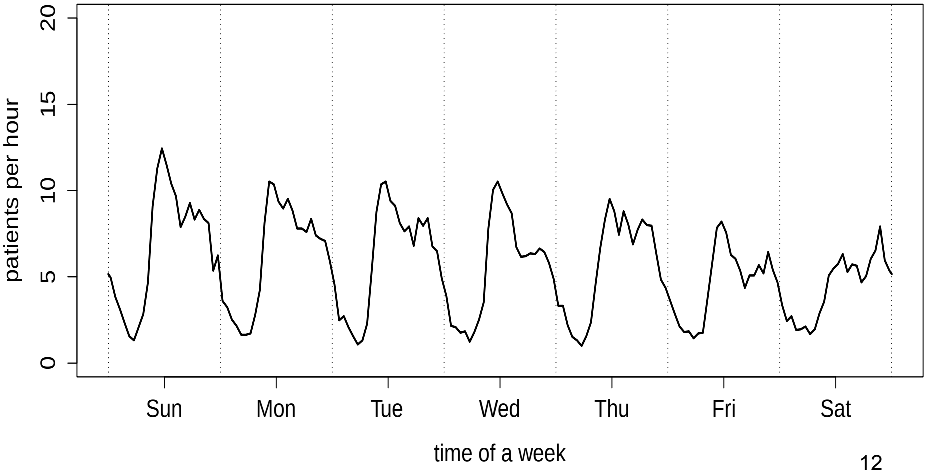
Average Occupancy Levels in the ED by Day of the Week



Arrivals: Daily Totals and Within Days

- Arrivals By Day of Week (DoW)
- Daily Totals
  - Linear Regression (ANOVA)
  - The Normalized Arrival Rate Function
- Statistical Test of NHPP Within Days

Average Arrival Rate in the ED by Day of the Week



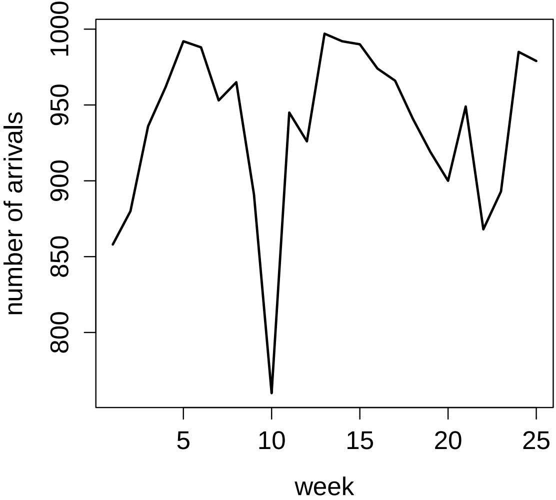
Daily Number of Arrivals to ED: First 13 of 25 Weeks

Week	Sun	Mon	Tue	Wed	Thu	Fri	Sat	sum	ave
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
14-25	next								
Mean	**162.00**	**148.40**	**145.04**	**132.40**	**133.24**	**109.88**	**105.40**	**936.36**	**133.8**
Var.	191.58	187.08	275.71	139.33	270.44	152.78	196.75	3110.99	

Daily Number of Arrivals to ED: Last 12 of 25 Weeks

Week	Sun	Mon	Tue	Wed	Thu	Fri	Sat	sum	ave
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
Mean	**162.00**	**148.40**	**145.04**	**132.40**	**133.24**	**109.88**	**105.40**	**936.36**	**133.8**
Var.	191.58	187.08	275.71	139.33	270.44	152.78	196.75	3110.99	

Average Number of Arrivals Per Week over 25 Weeks



A Model for the Daily Total Number of Arrivals

- Let $T(w, d)$ be total daily number of arrivals in **week** w on **DoW** d
- Model: $T(w, d) = G(m(w, d), \sigma^2)$, $\quad m(w, d) = A + B(w) + C(d)$
  - $G(m, \sigma^2)$ **Gaussian** with mean m and variance σ^2
  - $\sum_w B(w) = \sum_d C(d) = 0$, and $A = 133.8$, the overall daily average
  - with $\sigma^2 \approx m$ if arrivals approximately NHPP
- Fit model by doing **linear regression** (fit by least squares).
  - Construct Analysis of Variance (ANOVA) table
- Simplified Model: omit $B(w)$ (omit the week factor)

ANOVA Tables: With and Without Week 10

Factor	Sum of square	df	Mean sum of square	F statistics	P-value
Week	10666	24	444.4	2.75	$<$0.01
DoW	62893	6	10482.2	64.89	$<$0.01
Residuals	23262	144	161.5		

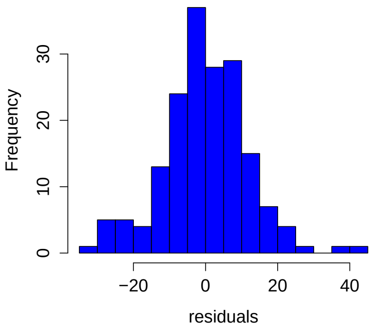
- $\hat{\sigma}^2 = 23262/144 = 161.5, \;\; V/M = 161.5/133.8 = 1.21$
- main model: $\hat{\sigma}^2_{DoW} = (23262 + 10666)/(144 + 24) = 202.0$

  $V/M = 202.0/133.8 = 1.51$, over-dispersion

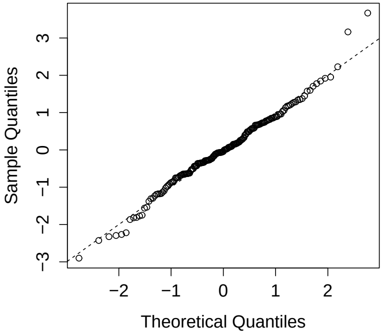
Factor	Sum of square	df	Mean sum of square	F statistics	P-value
Week	6038	23	262.5	1.58	0.058
DoW	60723	6	10120.5	60.77	$<$0.01
Residuals	22983	138	166.5		

- $\hat{\sigma}^2 = 22983/138 = 166.5, \;\; V/M = 166.5/134.8 = 1.23$
- $\hat{\sigma}^2_{DoW} = (22983 + 6038)/(138 + 23) = 180.2,$

The Residuals: Histogram and $Q - Q$ Plot (Studentized)

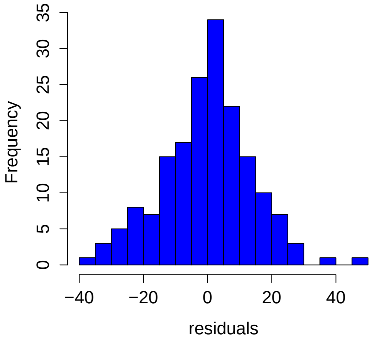


(c) Histogram of residuals.

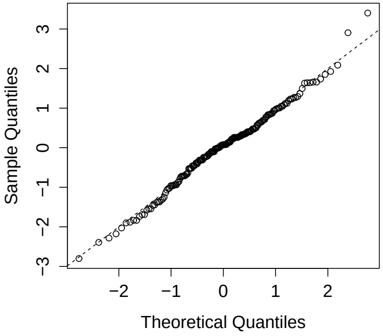


(d) Q-Q plot for studentized residuals.

The Residuals Without Week Factor: $m(w, d) = A + C(d)$



(e) Histogram of residuals.



(f) Q-Q plot for studentized residuals.

Regression Coefficients for $m(d, w) = A + B(w) + C(d)$

Coefficients	Estimate	SE	Coefficients	Estimate	SE
A	133.766	5.349			
B.Week1	-11.194	6.794	B.Week17	4.234	6.794
B.Week2	-8.051	6.794	B.Week18	0.663	6.794
B.Week3	-0.051	6.794	B.Week19	-2.480	6.794
B.Week4	3.663	6.794	B.Week20	-5.194	6.794
B.Week5	7.949	6.794	B.Week21	1.806	6.794
B.Week6	7.377	6.794	B.Week22	-9.766	6.794
B.Week7	2.377	6.794	B.Week23	-6.194	6.794
B.Week8	4.091	6.794	B.Week24	6.945	6.794
B.Week9	-6.480	6.794	B.Week25	6.091	6.794
B.Week10	-25.194	6.794	C.Sun	28.234	3.595
B.Week11	1.234	6.794	C.Mon	14.634	3.595
B.Week12	-1.480	6.794	C.Tue	11.274	3.595
B.Week13	8.663	6.794	C.Wed	-1.366	3.595
B.Week14	7.949	6.794	C.Thu	-0.526	3.595
B.Week15	7.663	6.794	C.Fri	-23.886	3.595
B.Week16	5.377	6.794	C.Sat	-28.366	3.595

Regression Coefficients for Model $m(d, w) = A + C(d)$

Coefficients	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

Summary for Daily Totals

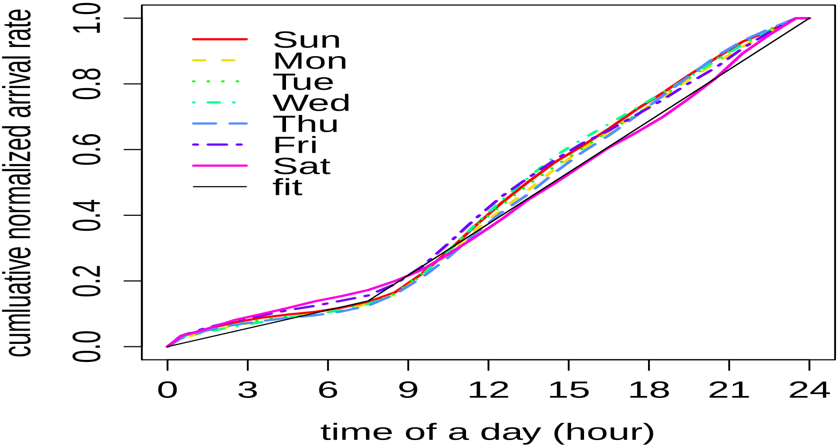
- Day of Week (DoW) has significant impact
  - mean daily total $m(d) = E[T(d)] = 133.8 + C(d)$
  - $C(Sun) = +28.2, \quad C(Fri) = -23.9 \quad$ and $\quad C(Sat) = -28.4$
  - the week can be important (week 10)
- Daily Totals Approximately Gaussian with some over-dispersion

  $V/M \approx 1.51$
  - Over-dispersion might be reduced by forecasting (predict week effect)
  - Might then be NHPP over week (roughly reasonable)

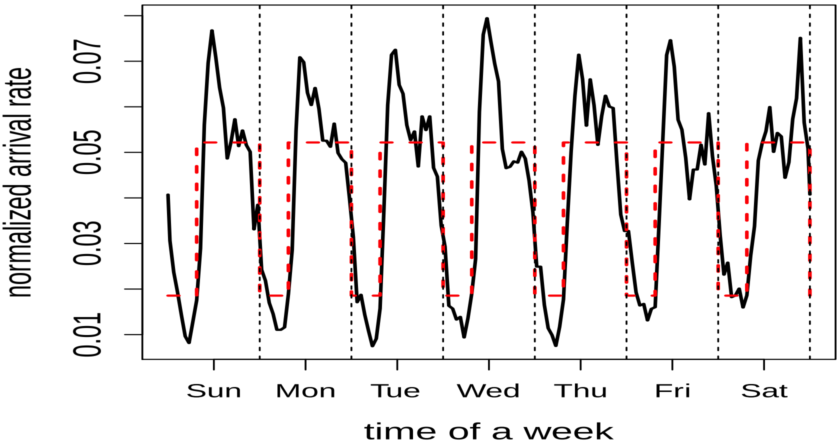
Arrivals Within Each Day

- Arrival Rate Function
- Statistical Test of a Nonhomogeneous Poisson Process (NHPP)

Cumulative Arrival Rate Divided By Daily Total



normalized arrival rate function



Statistical Tests of a NHPP Within Each Day

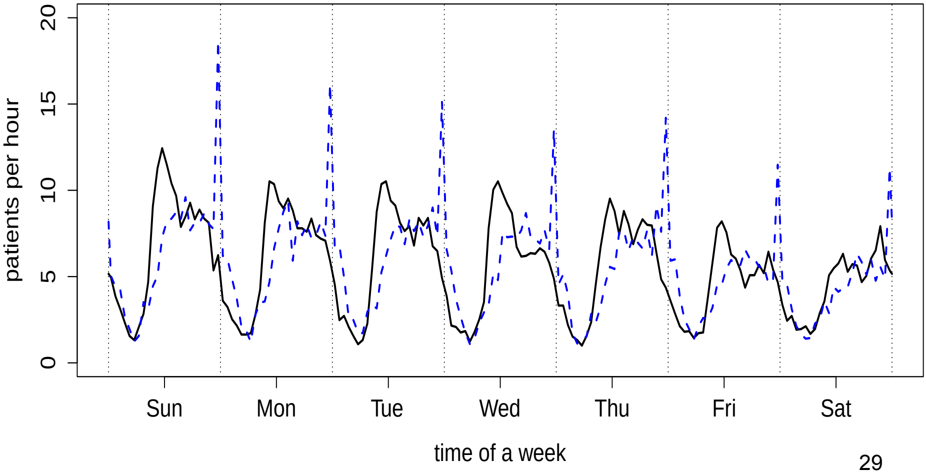
- Use Kolmogorov-Smirnov (KS) Statistical Tests (last week)
- Lewis and CU KS tests
- The arrival data mostly passes these tests using intervals $L = 1$ hour
  - divide day into subintervals of a few hours, so that 25 days produces about 400 arrivals
  - apply CU transformation within each hour and each day separately
  - effect of rounding to seconds

Statistical Tests of NHPP Within Days (Sample)

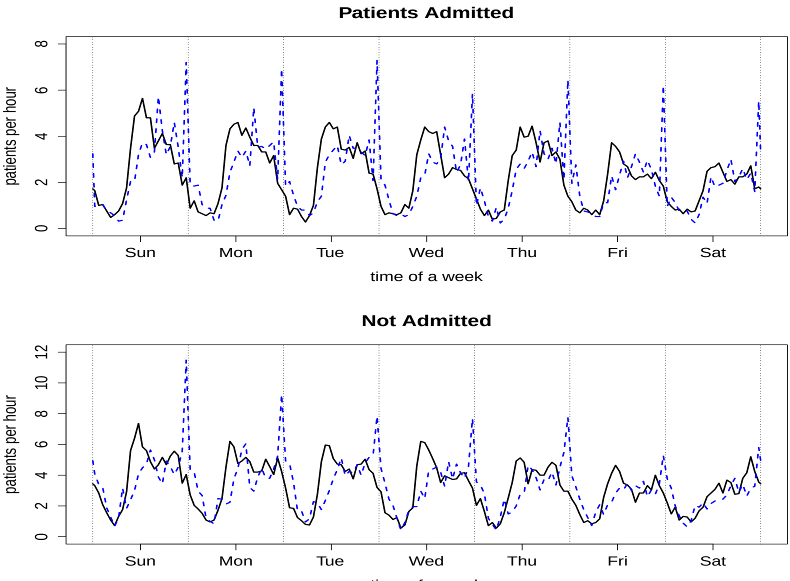
				$L = 1$, raw		$L = 1$, unrounded	
Int start	Int end	Day	N	CU	Lewis	CU	Lewis
7	9	Sun	188	0.81	0.69	0.81	0.69
7	9	Mon	176	0.03	0.26	0.03	0.27
7	9	Tues	192	0.00	0.49	0.00	0.49
7	9	Wed	150	0.00	0.24	0.00	0.24
7	9	Thurs	174	0.07	0.52	0.07	0.52
7	9	Fri	138	0.19	0.84	0.18	0.82
7	9	Sat	121	0.16	0.13	0.16	0.14
13	15	Sun	502	0.25	0.00	0.25	0.00
13	15	Mon	462	0.87	0.01	0.87	0.01
13	15	Tues	430	0.62	0.06	0.63	0.05
13	15	Wed	448	0.03	0.31	0.03	0.32
13	15	Thurs	408	0.88	0.13	0.88	0.13
13	15	Fri	308	0.17	0.75	0.17	0.74
13	15	Sat	290	0.84	0.26	0.84	0.26
		Average	371.6	0.40	0.33	0.40	0.33
		# Pass		58	50	58	50

Look at Departure Times

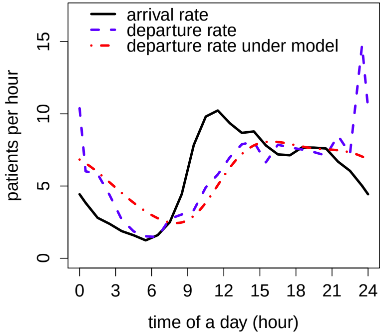
Average Arrival and Departure Rates by Day of the Week



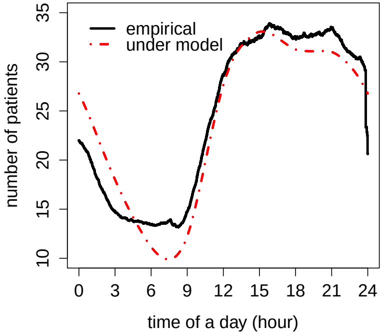
The Rates for Patients Admitted to Hospital and Not



The TVLL or $M_t/GI/\infty$ View (Assuming Fixed LOS CDF)

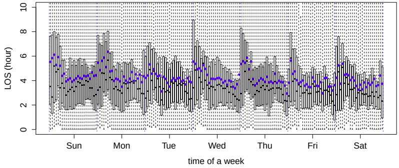


(g) Departure rate $\delta^e(t)$.

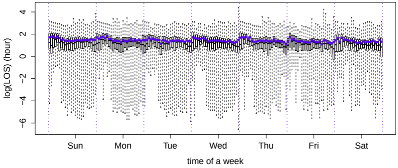


(h) Expected number in system $EQ^e(t)$.

Average LOS Distribution By Hour Over a Week

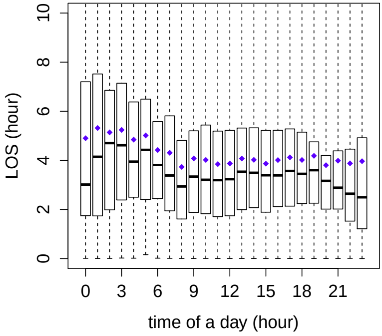


(i) LOS by time.

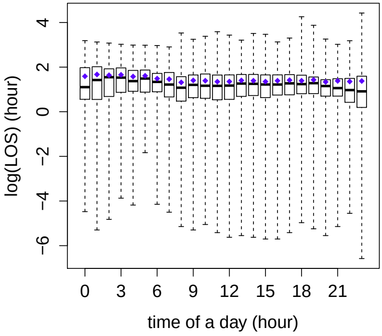


(j) log(LOS) by time.

Average LOS Distribution By Hour of the Day

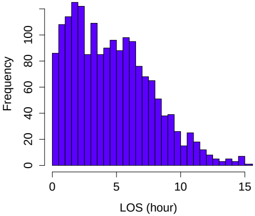


(k) LOS by time.

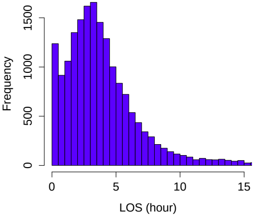


(l) log(LOS) by time.

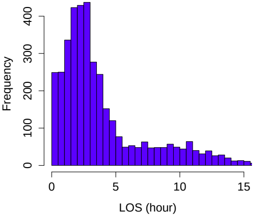
The average LOS distribution over three time intervals



(m) midnight to 6 a.m.



(n) 6 a.m. to 9 p.m.



(o) 9 p.m. to midnight

The CCDF of LOS by Hour of Arrival: Going Down

	Hour of Arrival							
LoS	1	2	3	4	5	6	7	8
0	0.9501	0.9450	0.9519	0.9676	0.9488	0.9634	0.9663	0.9367
1	0.8135	0.7898	0.8330	0.8466	0.8635	0.8415	0.8889	0.8321
2	0.5844	0.6228	0.7071	0.7493	0.7474	0.7317	0.7508	0.6764
3	0.4420	0.5285	0.6224	0.6608	0.5768	0.6179	0.5657	0.4988
4	0.3700	0.4833	0.5103	0.5192	0.4471	0.4878	0.4108	0.3552
5	0.3304	0.4381	0.3913	0.4041	0.3276	0.3577	0.2559	0.2822
6	0.2952	0.3222	0.2792	0.3127	0.2423	0.2358	0.1448	0.1995
7	0.2320	0.2515	0.2151	0.2360	0.1468	0.1707	0.0909	0.1484
8	0.1880	0.1768	0.1442	0.1475	0.1092	0.1301	0.0774	0.1265
9	0.1454	0.1218	0.1076	0.0855	0.0853	0.1098	0.0606	0.0827
10	0.0969	0.0923	0.0847	0.0678	0.0614	0.0854	0.0572	0.0608
11	0.0705	0.0668	0.0503	0.0472	0.0444	0.0488	0.0438	0.0535
12	0.0543	0.0609	0.0389	0.0413	0.0410	0.0407	0.0337	0.0389

Notes: Entry $X_{i,j} = P(LOS > i$ **hours**$|$**arrive in hour** $j)$ **for row** i **column** j. **Row max in red; row min in green.**

The CCDF of LOS by Hour of Arrival: Going Down

	Hour of Arrival							
LoS	1	2	3	4	5	6	7	8
13	0.0352	0.0550	0.0366	0.0295	0.0341	0.0407	0.0269	0.0316
14	0.0338	0.0491	0.0297	0.0265	0.0341	0.0407	0.0269	0.0292
15	0.0323	0.0432	0.0252	0.0236	0.0307	0.0325	0.0236	**0.0268**
16	0.0308	0.0413	0.0252	0.0236	0.0307	0.0325	**0.0236**	*0.0024*
17	0.0308	0.0413	0.0252	0.0206	0.0307	**0.0325**	*0.0034*	0.0024
18	0.0308	0.0413	0.0229	0.0206	**0.0273**	*0.0041*	0.0034	0.0000
19	0.0308	0.0413	0.0229	**0.0206**	*0.0000*	0.0041	0.0000	0.0000
20	0.0308	0.0413	**0.0229**	*0.0000*	0.0000	0.0000	0.0000	0.0000
21	0.0308	**0.0413**	*0.0000*	0.0000	0.0000	0.0000	0.0000	0.0000
22	**0.0308**	*0.0000*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
23	*0.0015*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
24	0.0015	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
25	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Notes: Entry $X_{i,j} = P(LOS > i$ hours$|$arrive in hour $j)$ for row i column j. after 11pm bold; after midnight italics; at midnight is the difference.

The CCDF of LOS by Hour of Arrival: Going Down

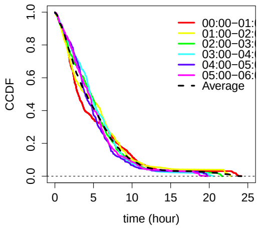
	Hour of Arrival							
LoS	9	10	11	12	13	14	15	16
0	0.8956	0.9261	0.9219	0.9176	0.9131	0.9357	0.9073	0.8923
1	0.7775	0.8112	0.8040	0.7855	0.7786	0.8140	0.8224	0.8029
2	0.6071	0.6434	0.6320	0.6243	0.6158	0.6674	0.6833	0.6608
3	0.4258	0.4779	0.4455	0.4618	0.4710	0.5048	0.4900	0.4773
4	0.2761	0.3317	0.3226	0.3411	0.3211	0.3527	0.3438	0.3264
5	0.1992	0.2378	0.2451	0.2097	0.2265	0.2303	0.2347	0.2275
6	0.1470	0.1799	0.1682	0.1410	0.1609	0.1584	0.1534	0.1508
7	0.1099	0.1237	0.1248	0.1027	0.1075	0.1072	0.1106	**0.1069**
8	0.0687	0.0892	0.0945	0.0759	0.0779	0.0816	**0.0842**	*0.0375*
9	0.0508	0.0707	0.0712	0.0544	0.0541	**0.0650**	*0.0285*	0.0239
10	0.0467	0.0554	0.0548	0.0442	**0.0463**	*0.0187*	0.0221	0.0200
11	0.0412	0.0442	0.0491	**0.0352**	*0.0161*	0.0145	0.0157	0.0168
12	0.0343	0.0402	**0.0441**	*0.0096*	0.0103	0.0131	0.0143	0.0136
13	0.0330	**0.0369**	*0.0038*	0.0090	0.0071	0.0118	0.0136	0.0128
14	**0.0330**	*0.0040*	0.0025	0.0066	0.0058	0.0118	0.0136	0.0120
15	*0.0027*	0.0032	0.0025	0.0048	0.0058	0.0118	0.0121	0.0112

The CCDF of LOS by Hour of Arrival: Going Down

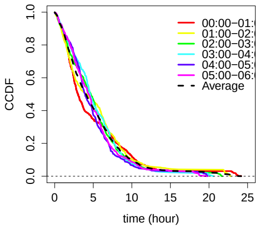
	Hour of Arrival							
LoS	17	18	19	20	21	22	23	24
0	0.9289	0.9505	0.9503	0.9602	0.9548	0.9480	**0.9609**	*0.8093*
1	0.8340	0.8405	0.8460	0.8673	0.8312	**0.8400**	*0.6960*	0.7089
2	0.6774	0.6961	0.7074	0.6948	**0.6704**	*0.5695*	0.5161	0.4961
3	0.4809	0.5189	0.4939	**0.5098**	*0.3502*	0.3834	0.3490	0.3385
4	0.3285	0.3291	**0.3472**	*0.2600*	0.2195	0.2380	0.2434	0.2679
5	0.2312	**0.2309**	*0.1573*	0.1702	0.1355	0.1755	0.1975	0.2249
6	**0.1643**	*0.1058*	0.1051	0.1132	0.1014	0.1511	0.1779	0.1996
7	*0.0728*	0.0781	0.0725	0.0890	0.0872	0.1348	0.1496	0.1797
8	0.0550	0.0579	0.0481	0.0765	0.0777	0.1162	0.1447	0.1411
9	0.0398	0.0462	0.0432	0.0679	0.0721	0.1040	0.1065	0.1080
10	0.0305	0.0395	0.0375	0.0640	0.0658	0.0877	0.0860	0.0838
11	0.0246	0.0378	0.0342	0.0585	0.0531	0.0674	0.0596	0.0595
12	0.0229	0.0361	0.0310	0.0484	0.0420	0.0479	0.0371	0.0408
13	0.0195	0.0327	0.0261	0.0328	0.0341	0.0325	0.0254	0.0254

Note: Entry $X_{i,j} = P(LOS > i \text{ hours} | \text{arrive in hour } j)$ for row i column j. row max red if not midnight.

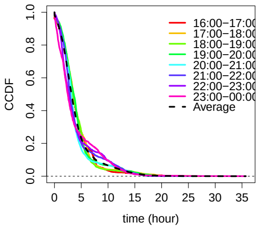
The average LOS CCDF over three time intervals



(p) midnight to 6 a.m.



(q) 6 a.m. to 4 p.m.



(r) 4 p.m. to midnight

References

- M. Armony et al. **On Patient Flow in Hospitals: A Data-Based Queueing Science Perspective.** forthcoming in *stochastic Systems*. Posted in Courseworks.
- S/ Maman **Uncertainty in the Demand for Service: The Case of Call Centers and Emergency Departments.** MS thesis, the Technion, 2009.
- Y. Tseytlin **Queueing Systems with Heterogeneous Servers: On Fair Routing of Patients in Emergency Departments.** MS thesis, The Technion, 2009.
- See http://iew3.technion.ac.il/serveng/References/references.html

Arrivals By Day of Week: Departure Same Day or Next Day

	Day	Sun.	Mon.	Tue.	Wed.	Thu.	Fri.	Sat.
Total	Mean	162.00	148.40	145.04	132.40	133.24	109.88	105.40
	Var.	191.58	187.08	275.71	139.33	270.44	152.78	196.75
Leave on the same day	Mean	136.84	123.80	124.12	115.16	113.60	95.00	87.72
	Var.	140.89	194.33	220.28	115.14	179.92	151.50	94.63
	Prop.	84.47%	83.42%	85.58%	86.98%	85.26%	86.46%	83.23%
Leave 1 day after	Mean	25.12	24.60	20.92	17.20	19.64	14.76	17.68
	Var.	53.78	42.75	49.91	26.92	43.07	41.11	39.31
	Prop.	15.51%	16.58%	14.42%	12.99%	14.74%	13.43%	16.77%
Leave 2+ days after	Num.	1	0	0	1	0	3	0
	Prop.	0.02 %	0%	0%	0.03%	0%	0.11%	0%