Queueing Models of Patient Flow in Hospitals: What Does the Data Tell Us?

Mor Armony

Joint work with: Avi Mandelbaum, Yariv Marmor, Yulia Tseytlin and Galit Yom-Tov

NYU, Technion, Mayo, IBM, Columbia

November 2010

Patient Flow in Hospitals as a Queueing Network

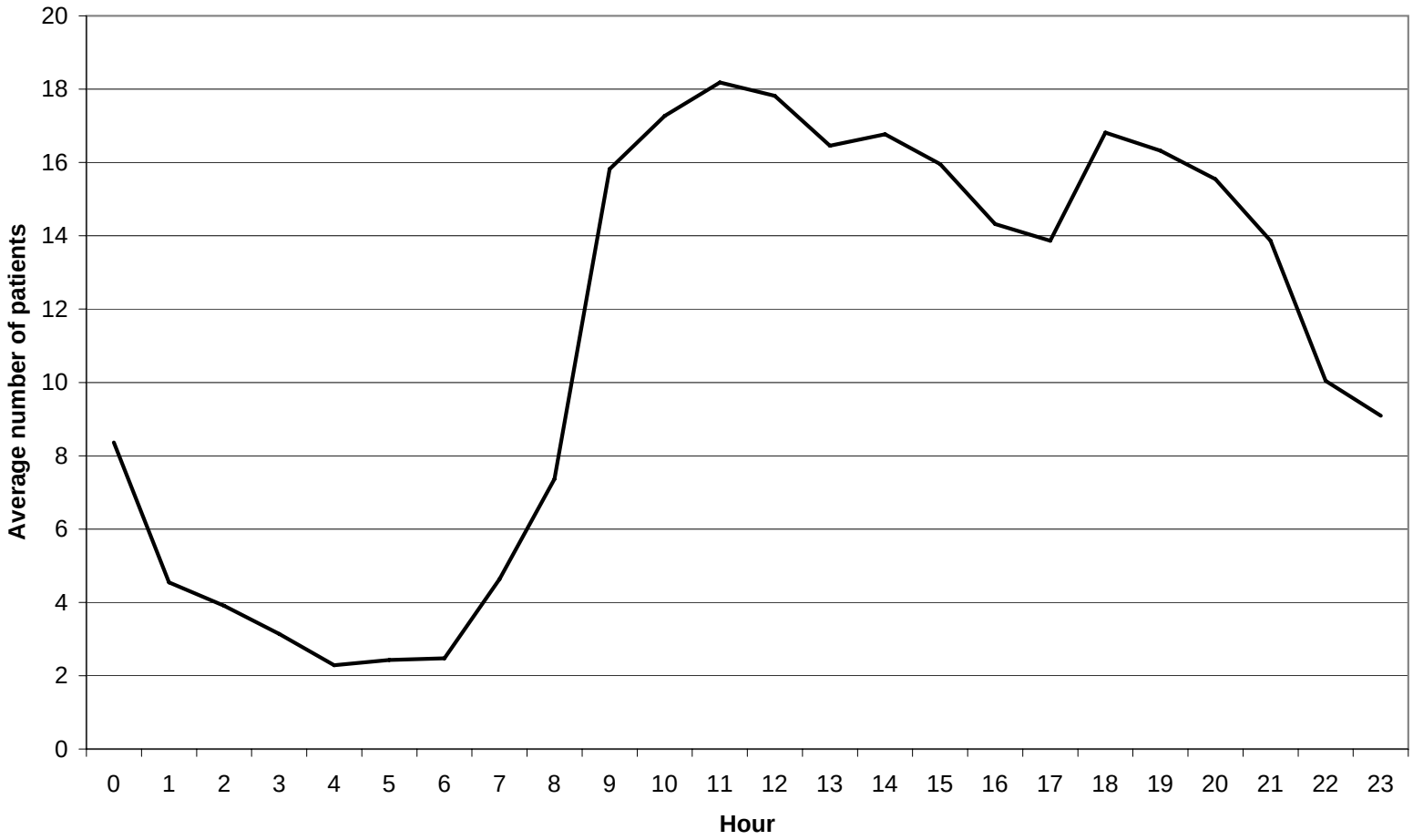
Questions:

- How to model arrivals, departures and transitions?
- Who are the servers?
- What are the service and arrival rates?
- What are the relevant performance measures?

Our data:

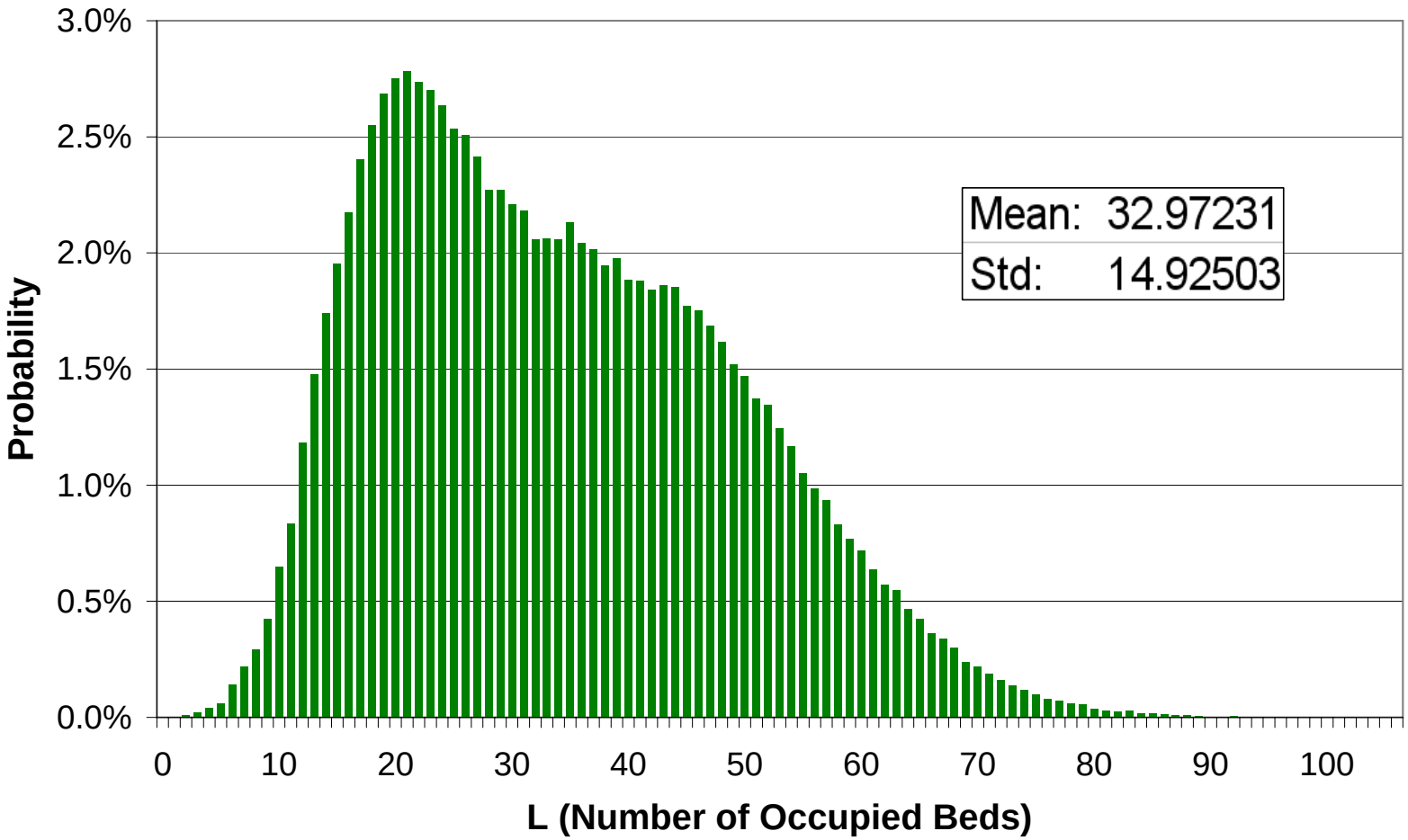
- Anonymous Israeli hospital with 1000 beds and 45 medical units
- 75,000 patients are admitted annually
- Years data collected: 2004 - 2008
- Individual patient level data, time stamps (admission, transfers and discharge)
- Our focus: ED, IW and transfers ED –> IW

Arrival rates into ED

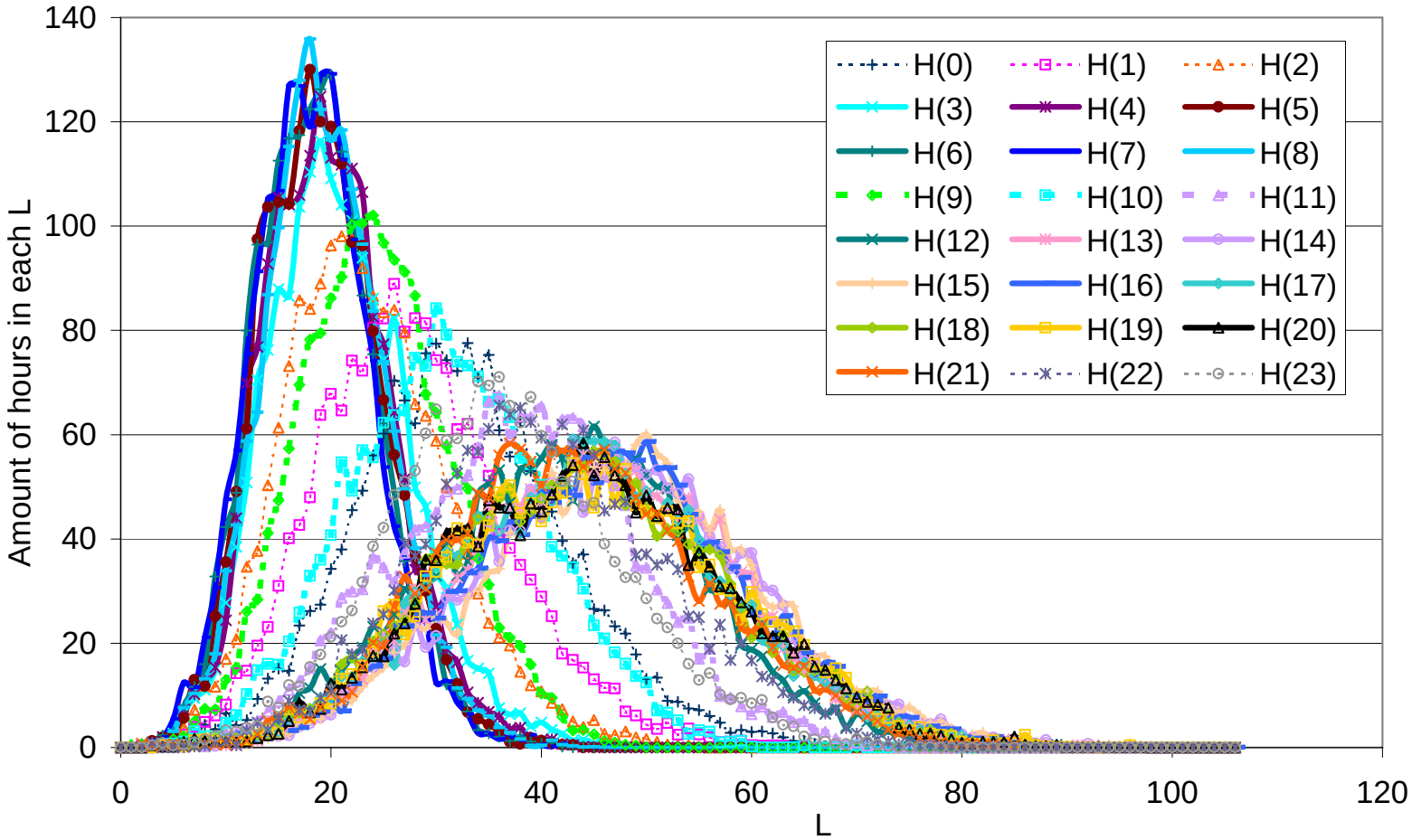


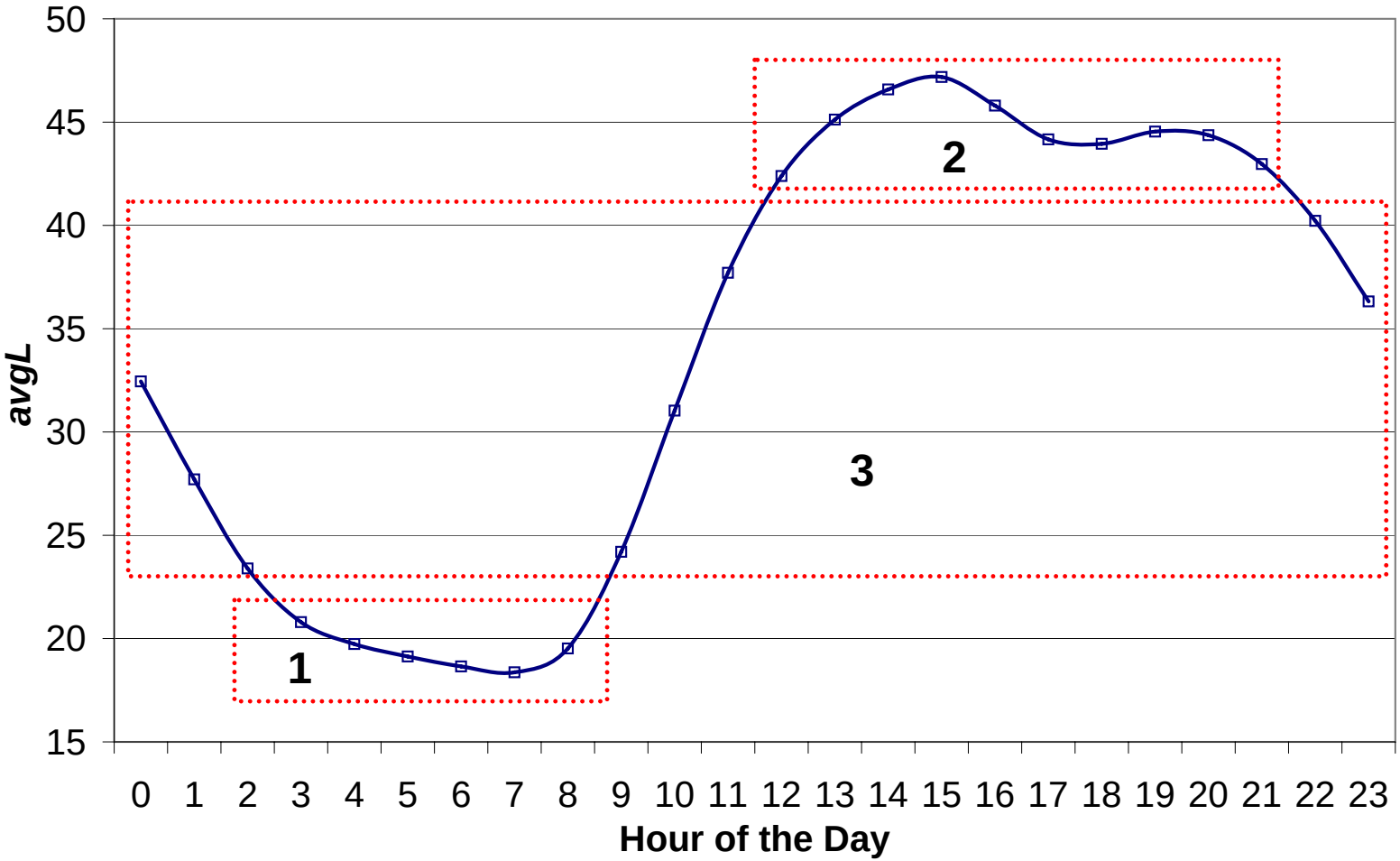
Average number of patients per hour (Jan 2005, weekdays)

ED: Distribution of the number of occupied beds



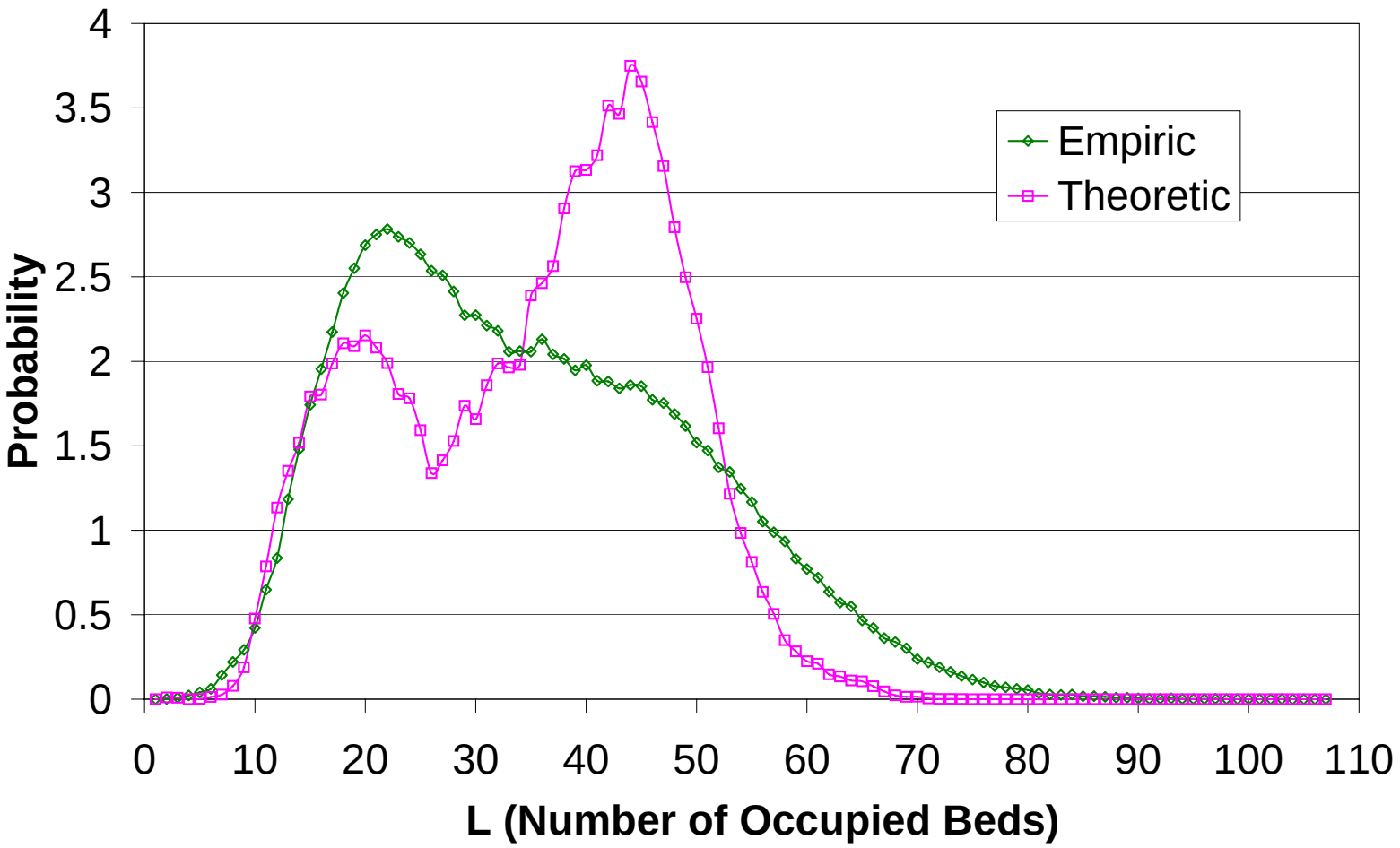
ED: Hourly distribution of the number of occupied beds



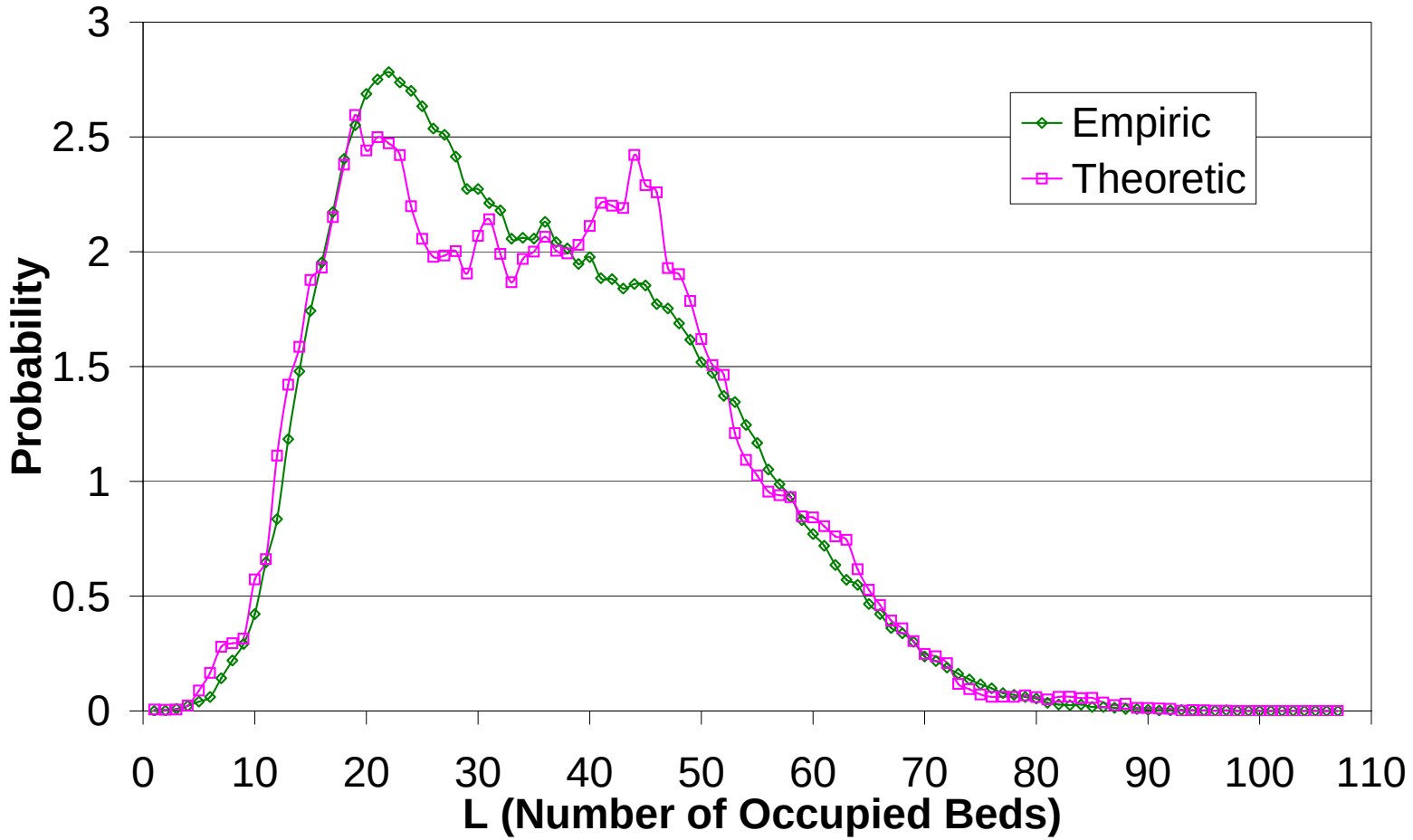


Average number of beds per hour of the day

ED: Number of occupied beds as an $M_t/M/\infty$ queueing model

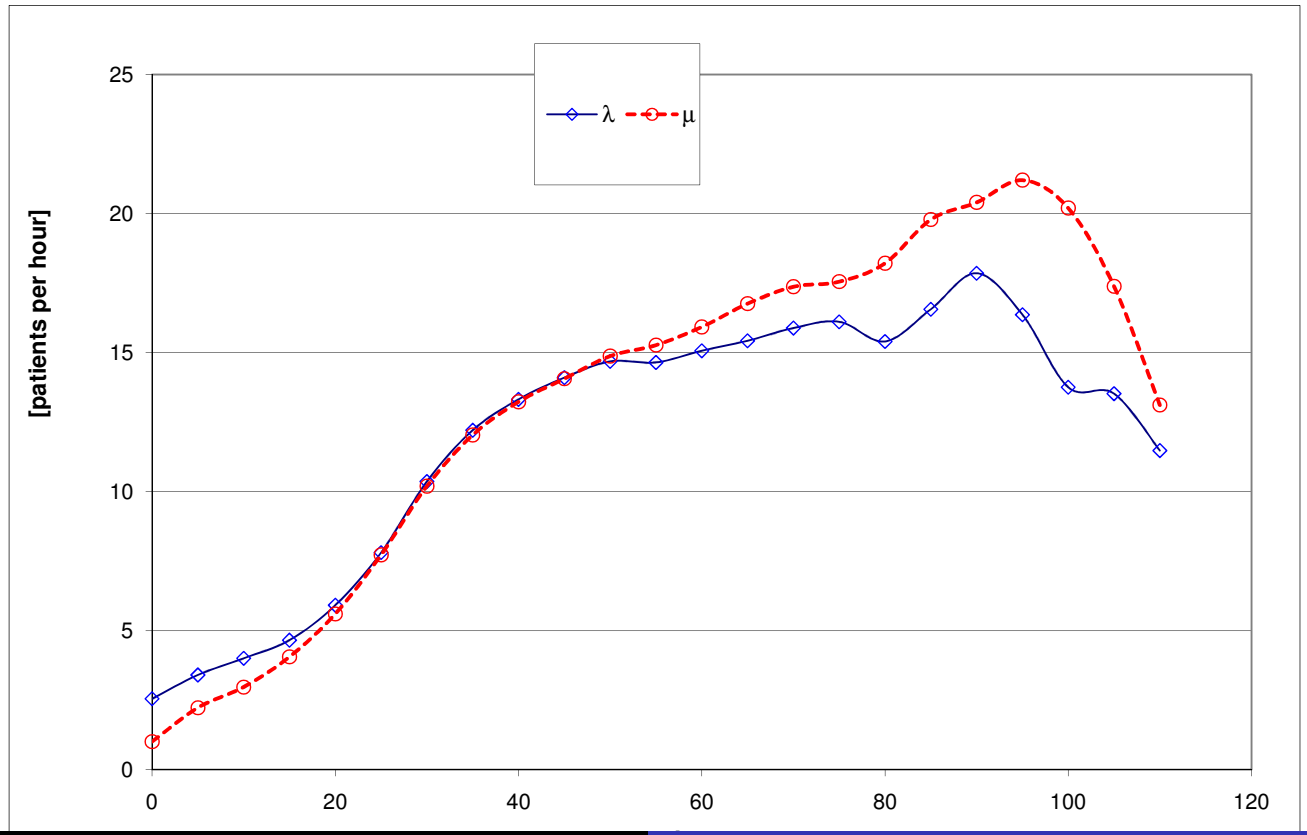


ED: Number of occupied beds as Birth & Death model



ED: Arrival and departure rates

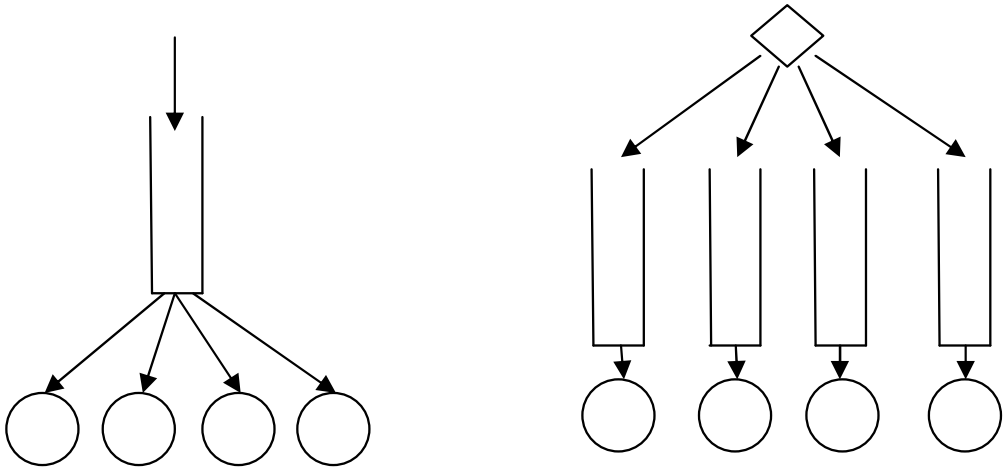
Observation: Both arrival and departure rates are state dependent.
Consistent with Diwas and Terwiesch (2008)



Internal Wards: Capacities and ALOS

	Ward A	Ward B	Ward C	Ward D	Ward E
Standard capacity (# beds)	45	30	44	42	24
Maximal capacity (# beds)	52	35	46	44	27
ALOS (days)	6.368	4.474	5.358	5.562	4.11

Routing: ED to IW



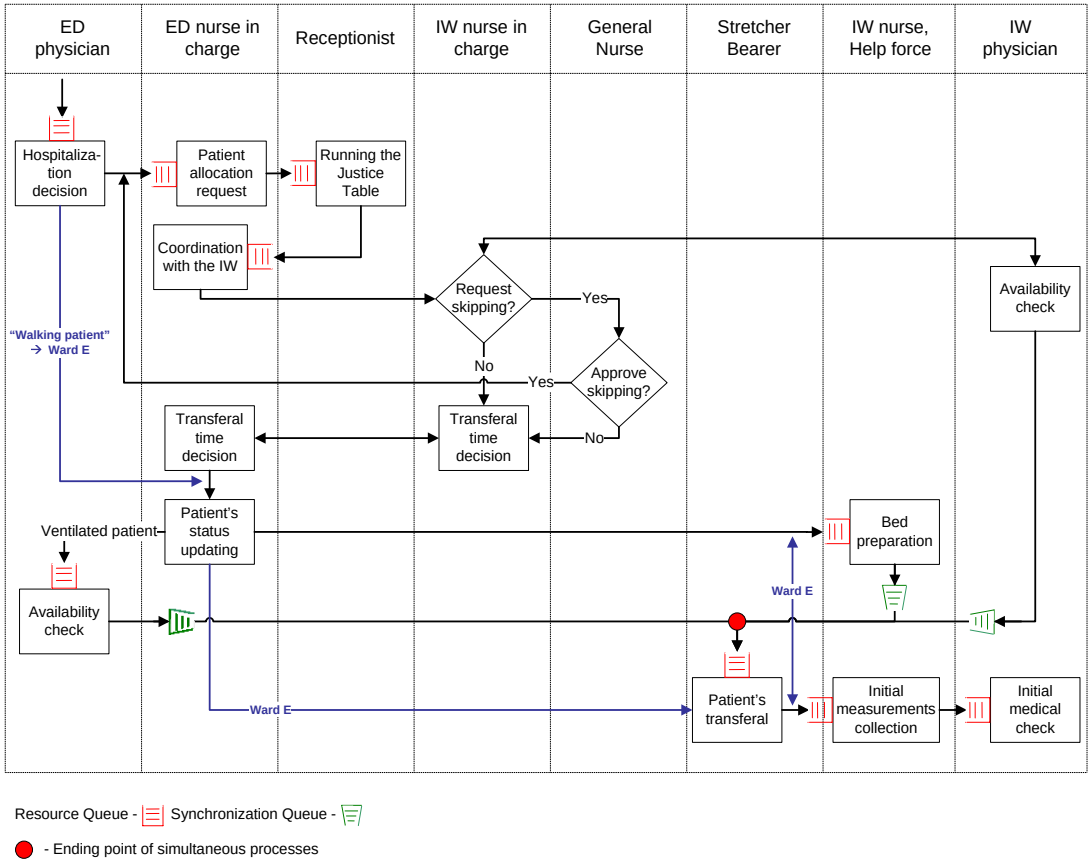
- Single line system is more efficient
- Reality requires multiple lines
- Patients require care even when in queue
- Push versus Pull
- Justice table is meant to ensure fairness

Fairness and Dis-economies of Scale

	Ward A	Ward B	Ward C	Ward D
ALOS (days)	6.368	4.474	5.358	5.562
Mean Occupancy Rate	97.8%	94.4%	86.8%	91.1%
Mean # Patients per Month	205.5	187.6	210.0	209.6
Standard capacity	45	30	44	42
Mean # Patients per Bed per Month	4.57	6.25	4.77	4.77
Return Rate (within 3 months)	16.4%	17.4%	19.2%	17.6%

* Data refer to period 1/05/06 - 30/10/08 (excluding the months 1-3/07)

Fork-Join networks revisited



- Zviran (2008): Diffusion limits and control
- Zaied (2010): Offered load

Operational Regime of IW

- Beds: QED regime.
  - Prediction: Erlang-B in QED: $N \simeq R + \beta\sqrt{R} \Rightarrow P(block) \simeq \frac{1}{\sqrt{N}} \frac{\phi(\beta)}{\Phi(\beta)}$ and $\rho \simeq 1 - \frac{\beta + \phi(\beta)/\Phi(\beta)}{\sqrt{N}}$.
  - For our data, the QED regime predicts: $P(block) \simeq 2.9\%$ and $\rho \simeq 91.7\%$.
  - Actual numbers: $P(block) = 3.54\%$ and $\rho = 93.1\%$.
- Doctors: ED regime.
  - Average handling time for patient admission: 30 minutes.
  - Average wait for admission (once a bed is ready): 2.5 hours.

Operational performance measures

Focus: Quality of Care

- Patient surveys
- Rate of return
- Mortality rates

Conclusions

- Hospital as a queueing network
- Arrival and departure rates are state dependent
- Push versus Pull in routing
- Fairness: Occupancy + Flux
- Economies and dis-economies of scale
- Fork-Join networks
- QED and ED regime co-exist in a single system.
- Operational measures should be in line with quality of care measures