Pre-Scheduled Random Arrivals for modelling air traffic operations

Analytical and applied results

Carlo Lancia

Mathematical Institute Leiden University

TU Delft, October 16, 2017
Seminar Series in Probability and Statistics

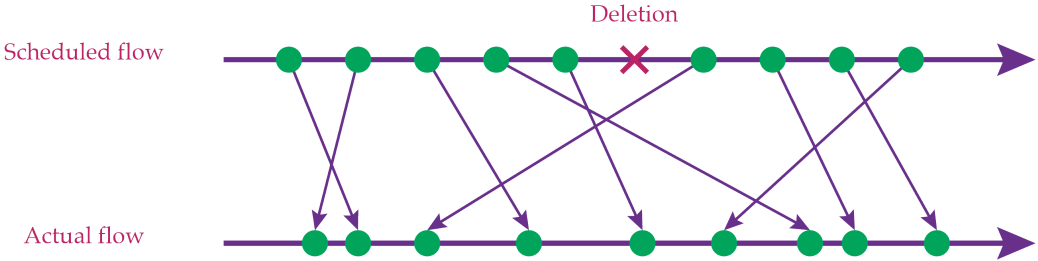
Outline

Pre-Scheduled Random Arrivals (PSRA)

- Continuous-time point process of the form

$$t_i = {}^{i}/_{\lambda} + \xi_i$$

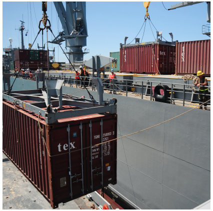
- λ is a fixed constant
- $\{\xi_i\}_i$ is a sequence of i.i.d. random variables
- Optional *thinning* of intensity ρ (later)



Range of applications

- Public transportation systems
- Vessels marine traffic
- Crane handling in loading/unloading operations
- Outpatient scheduling



All systems where scheduled arrivals are intrinsically subject to random variations.

Properties of PSRA

$$t_i = {}^{i}\!/\!_{\lambda} + \xi_i$$

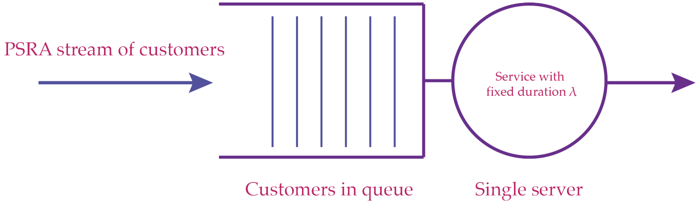
- We usually require $\mathbb{E}(\xi_i) = 0$ and $\sigma(\xi_i) < \infty$
- Interpretation as a random translation of the integers $z \in \mathbb{Z}$ (1D crystal)
- Weak convergence to the Poisson process for $\sigma \to \infty$ (1D gas)
- If N_1 and N_2 are the number of arrivals in $(t, t+T]$ and $(t+T, t+2T]$ then $\mathrm{Cov}(N_1, N_2) < 0$
- A congested period is likely to be followed by one with less arrivals than expected

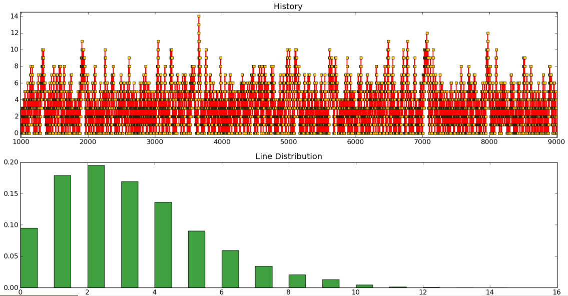
London Heathrow Airport

- Second busiest airport in the world by international passenger traffic
- Busiest airport in Europe by passenger traffic (both transit and terminal)
- Seventh busiest airport in the world by total passenger traffic (enplaned, deplaned, and direct-transit)
- 75.7 million passengers in 2016 (+1.0% increase from 2015)

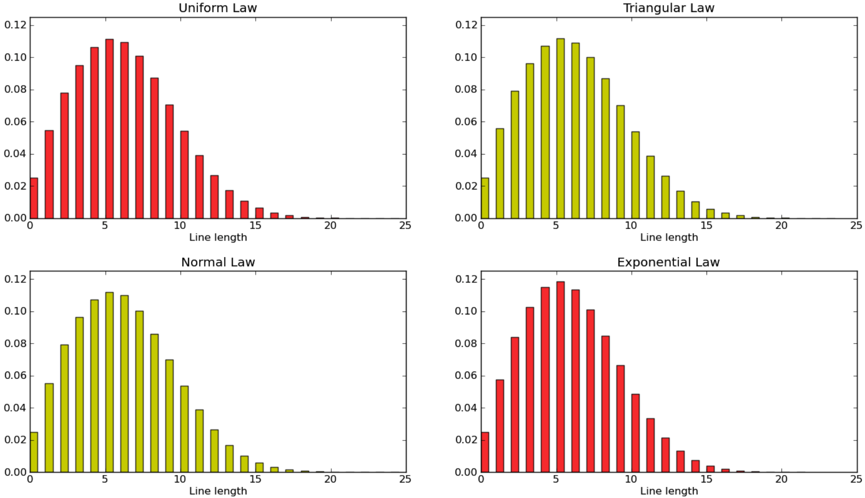
`http://news.images.itv.com/image/file/704067/stream_img.jpg`

PSRA/D/1


PSRA stream of customers
Service with
fixed duration λ
Customers in queue
Single server

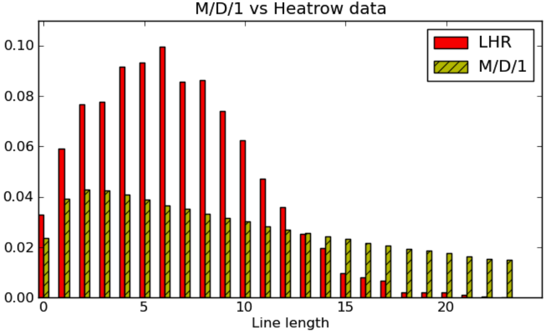

History
Line Distribution

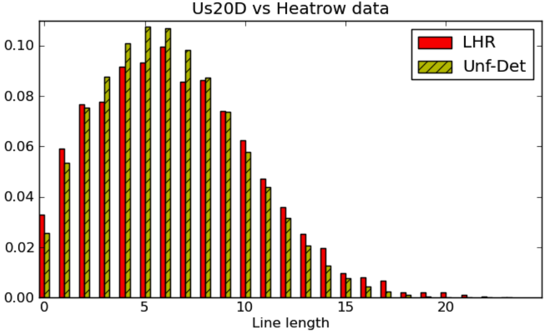
Insensitivity to the delays law with fixed σ



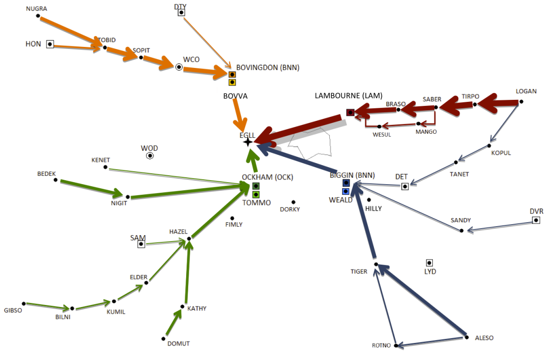
PSRA/D/1 vs M/D/1

- Poisson arrivals (M) are the *golden stardard* in the Air Traffic Management domain
- PSRA thinned with intensity $\rho = \frac{40}{41}$
- Uniform delays with $\sigma = {}^{20}/_{\lambda} = 30$ min
- Service parameter for both models is ρ





Limitations



From Caccavale et al. (2014) *J Air Transp Manag*, 34, 116–122

- Indirect comparison through a $\cdot$/D/1 queue model
- Analysis overlooks action of Traffic Control
- *This leads to overestimating PSRA parameter* σ
- Deterministic schedule of PSRA is equally spaced in time (OK for Heathrow, though)

Airport selection and Arrival areas

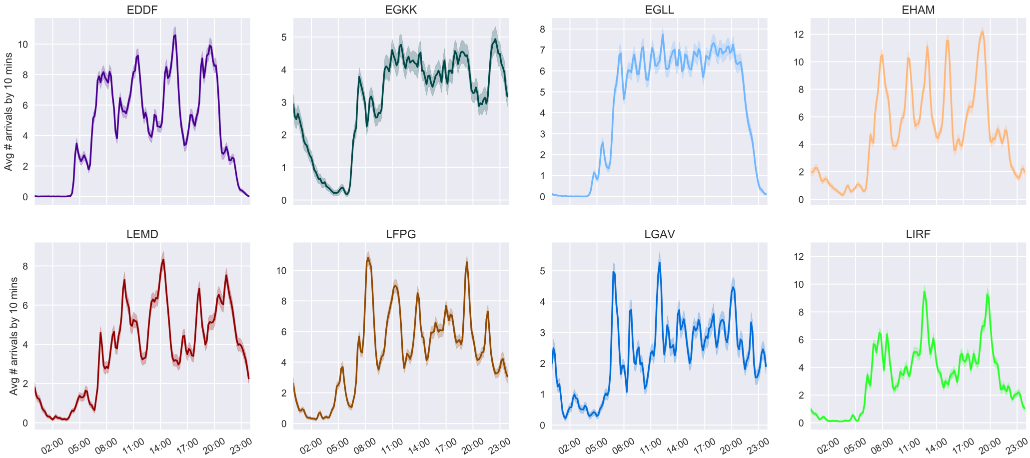
Dataset overview

airport name	ICAO code	sample size
Frankfurt am Main International Airport	EDDF	58167
London Gatwick Airport	EGKK	39746
London Heathrow Airport	EGLL	56716
Amsterdam Airport Schiphol	EHAM	63279
Madrid Barajas International Airport	LEMD	48162
Charles de Gaulle International Airport	LFPG	60122
Athens International Airport	LGAV	29503
Rome Fiumicino International Airport	LIRF	43333

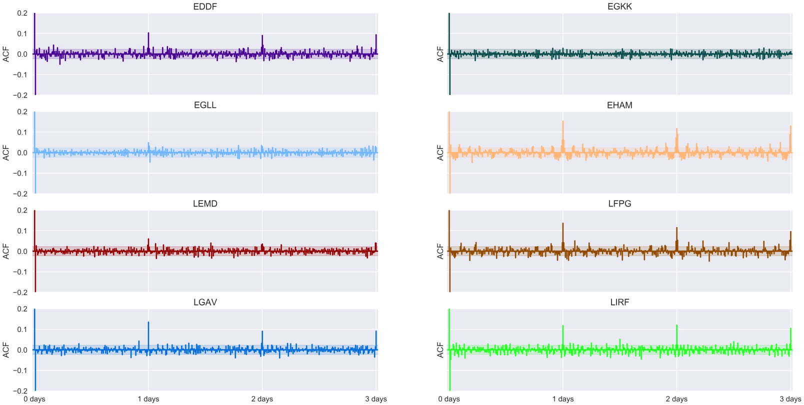
Study period goes from June 15 to September 15, 2016

We model the demand using first time within 40 NM from airport

Average daily demand



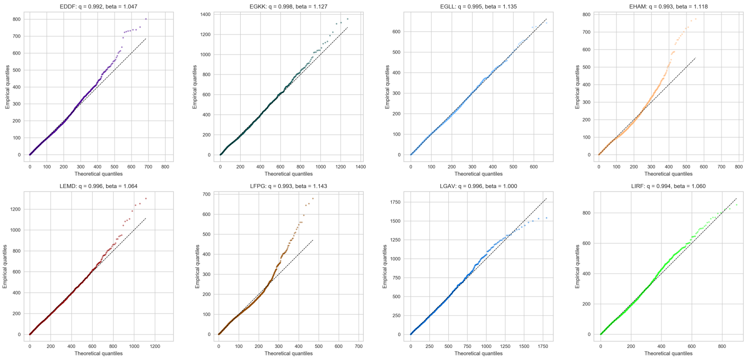
Significant autocorrelations at lag 10 mins and 1-2 days



We expect batch arrivals *(typically Weibull interarrivals)* and a daily-periodic demand

Interarrival times are (nearly) exponential

Fitted Weibull: $P_W(X = x; q, \beta) = q^{x^\beta} - q^{(x+1)^\beta}$

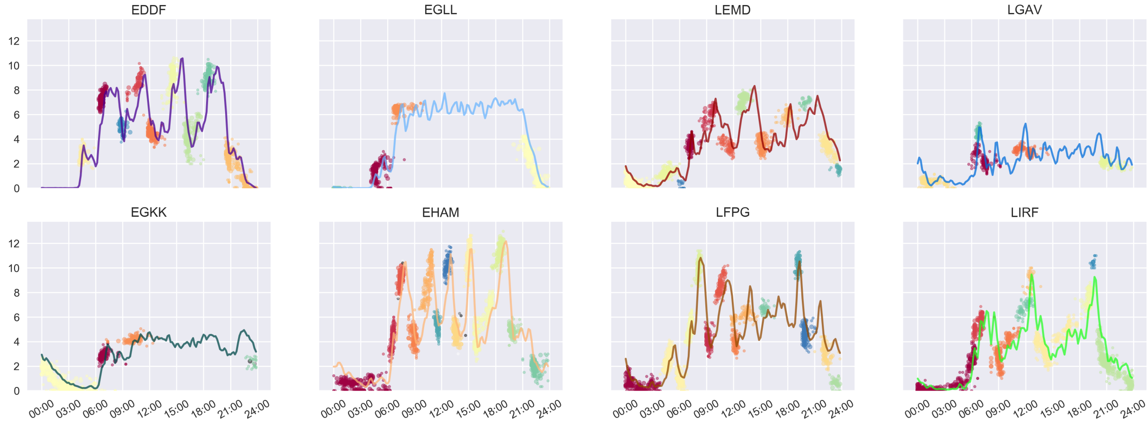


Poisson process explains shape $\simeq 1.0$ *but clashes with negative lag-1 autocorrelation*

Data-driven Poisson process

1. Aggregate arrivals TS, e.g. by intervals of 10 minutes
2. Run a changepoint-detection algorithm, e.g. PELT, under the null hypothesis of Poissonian arrivals to obtain
    - changepoint time $\hat{t}_k$
    - $\hat{\lambda}_k$, estimated intensity in $[\hat{t}_k, \hat{t}_{k+1})$
3. Cluster couples $\{(\hat{t}_k, \hat{\lambda}_k)\}_k$, e.g. via DBSCAN
4. Compute the centroid $(\bar{t}_i, \bar{\lambda}_i)$ of each cluster
5. Define a step-wise, periodic intensity function that takes on $\bar{\lambda}_i$ for $t \in [\bar{t}_i, \bar{t}_{i+1})$

Implementation of data-driven Poisson process



Data-driven PSRA

Recall

Fixed, deterministic schedule perturbed by i.i.d. random variables

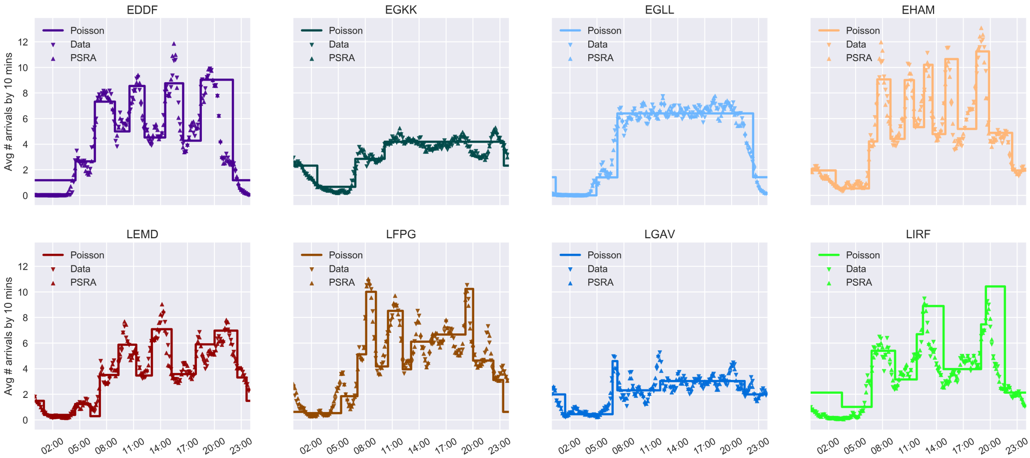
$$t_i = \frac{i}{\lambda} + \xi_i$$

1. Let t_i^{M3} the *actual* arrival time at 40 NM
2. Let t_i^{M1} the *anticipated* arrival time at 40 NM (from the last flight plan agreed with Eurocontrol)
3. Compute the *delays* $\delta_i = t_i^{\mathrm{M3}} - t_i^{\mathrm{M1}}$
4. Define

$$t_i = t_i^{\mathrm{M1}} + \xi_i$$

where $\{\xi_i\}_i$ are IID rvs drawn from the empirical distribution of the delays $\{\delta_i\}_i$

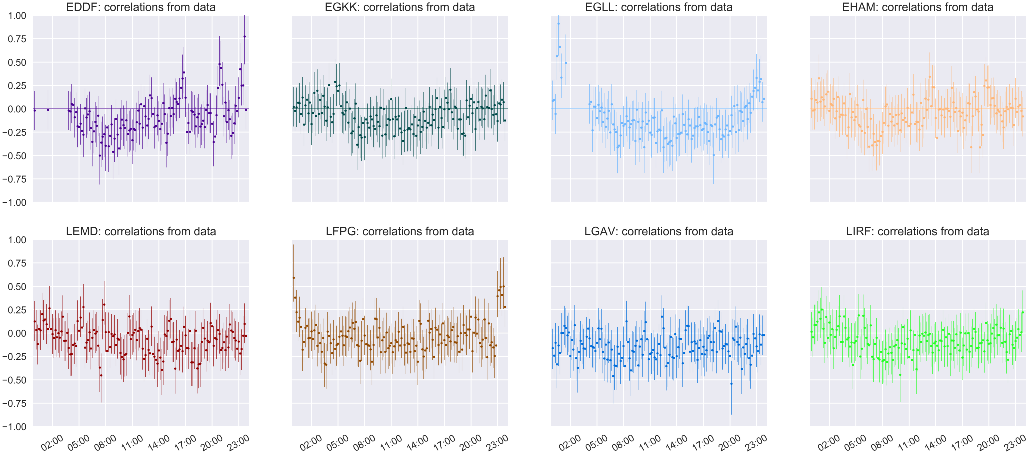
Data-driven Poisson vs PSRA



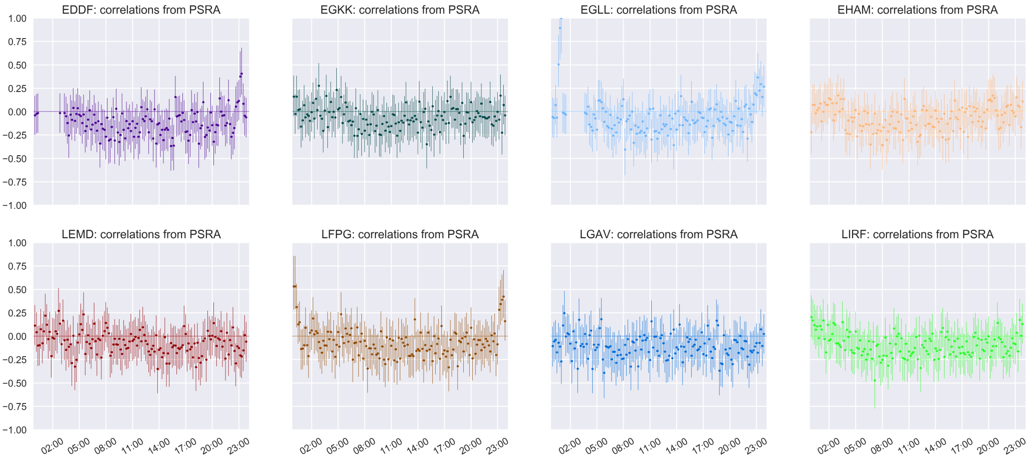
Data-driven Poisson vs PSRA

- PSRA outperform Poisson arrivals in reproducing the average daily demand
- Average demand reproduced with Poisson model by forcing intensity to vary over a finer time scale, e.g. every 10 minutes
  - CAVEAT number of parameters required increases in a sensible manner (144)
  - CAVEAT system mostly out of equilibrium
  - CAVEAT mathematical tractability is greatly reduced
- PSRA has 1 parameter at most
- PSRA inherits arrivals correlation-structure from the M1 schedule
- In contrast, Poissonian arrivals have independent increments by definition

Arrivals correlation



Correlations from simulation of PSRA



The EDA/D/1 queue model

Exponentially Delayed Arrivals (EDA)

Particular case of PSRA ($\lambda = 1$ for convenience)

$$t_i = i + \xi_i \,,$$

where now ξ_i are independent exponential random variables, i.e.

$$f_\xi(t) = \begin{cases} \beta\, e^{-\beta\, t} & \text{if } t > 0\,, \\ 0 & \text{otherwise.} \end{cases}$$

Thinning

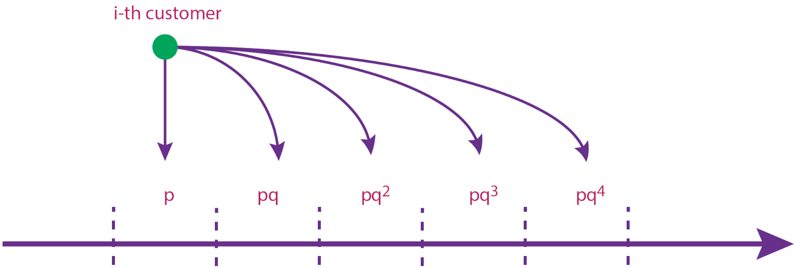
Each customer is deleted independently with probability $1 - \rho$, whereas it is kept with probability ρ.

The EDA/D/1 queue model

Probability of a customer arriving/not arriving in each slot

$$p = \int_0^1 f_\xi(t)\,dt = 1 - e^{-\beta} \quad , \quad q = 1 - p = e^{-\beta}$$

Conditioned on accepting the customer (probability ρ)



The EDA/D/1 queue model

Previous results about EDA

- *Winstein, 1959.* Solution for general delays $\xi_i \in [0, 2]$ and exponential service times;
- *Mercer, 1960.* Continuation of Winsten's studies;
- *Lewis, 1961.* Statistical properties of inter-arrivals;
- *Kendall, 1964.* Arrival process seen as output of stationary $D/M/\infty$ system;
- *Nelsen and Williams, 1970.* Exact characterization for the distribution of the inter-arrival time intervals;
- After that, only numerical studies applied to trasportation systems, outpatient scheduling and crane handling optimization.

System recursion

Queue length at time t

$$n_{t+1} = n_t + a_{(t,t+1]} - (1 - \delta_{n_t,0})$$

where $a_{(t,t+1]}$ is the number of arrivals in the interval $(t, t+1]$.

This quantity does depend on the whole trajectory up to time t!

How many arrivals in the unit interval?

In the interval $(t, t+1]$ there may arrive

- the t-th customer, provided it is not deleted;
- any of the l_t customers that have not been deleted and have not yet arrived at time t.

System recursion

Queue length at time t

$$n_{t+1} = n_t + a_{(t,t+1]} - (1 - \delta_{n_t,0})$$

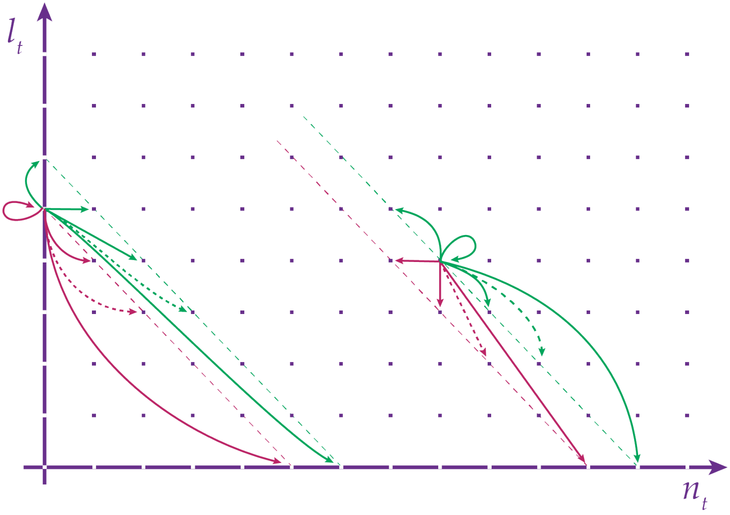
where $a_{(t,t+1]}$ is the number of arrivals in the interval $(t, t+1]$.

This quantity does depend on the whole trajectory up to time t!

Number of arrivals in a slot

$$\mathbb{P}\left(a_{(t,t+1]} = j \,\middle|\, l_t = k\right) = \begin{cases} \binom{k}{j} p^j\, q^{k-j} & \text{deletion at time } t \\ \binom{k+1}{j} p^j\, q^{k+1-j} & \text{otherwise} \end{cases}$$

EDA/D/1 as a random walk in the quarter plane



- l_t, number of late customers at time t
- n_t depends only on n_{t-1} and l_{t-1}
- (n_t, l_t) is a bivariate Markov chain

Existence of a unique stationary state

- (n_t, l_t) is *positive recurrent*, i.e. finite mean first return time to any state
- This is a consequence of the *stochastic stability* of the system for $q, \rho < 1$
- Use Foster's criterion with the Lyapunov function

$$V(n_t, l_t) = M\alpha_t + l_t + 1\,,$$

where

- $M = 2/1-\rho$
- $\alpha_t = (n_t + l_t)$

The bivariate generating function $P(z,y)$

Let us define $P(z,y) = \sum_{n,l} P_{n,l}\, z^n\, y^l$

Theorem (Lancia et al, 2017)

The bivariate generating function $P(z,y)$ satisfies

$$P(z,y) = \frac{1-\rho+\rho(z+q(y-z))}{z}\Big[(z-1)P(0,z+q(y-z)) + P(z,z+q(y-z))\Big],$$

where

- $q = e^{-\beta}$
- β *is the parameter of the exponential delays*
- ρ *is the thinning intensity*

$P(z, y)$ as a q-series

$$P(z,y) = \frac{1-\rho+\rho(z+q(y-z))}{z}\Big[(z-1)P(0, z+q(y-z)) + P(z, z+q(y-z))\Big],$$

Key idea

$$P(z,y) = \sum_{k\geqslant 0} q^k P^{(k)}(z,y)\ , \quad P(z,y) = \sum_{j\geqslant 0} \frac{(y-z)^j}{j!}\frac{\partial^j}{\partial y^j}P(z,y)\Big|_{y=z}$$

- The j-th derivative of $P(z, y)$ gives a factor q^j
- after rearranging, we get a recursive relation for $P^{(k)}(z,y) = [q^k]P(z,y)$

$P(z, y)$ as a q-series

Theorem (Lancia, 2013)

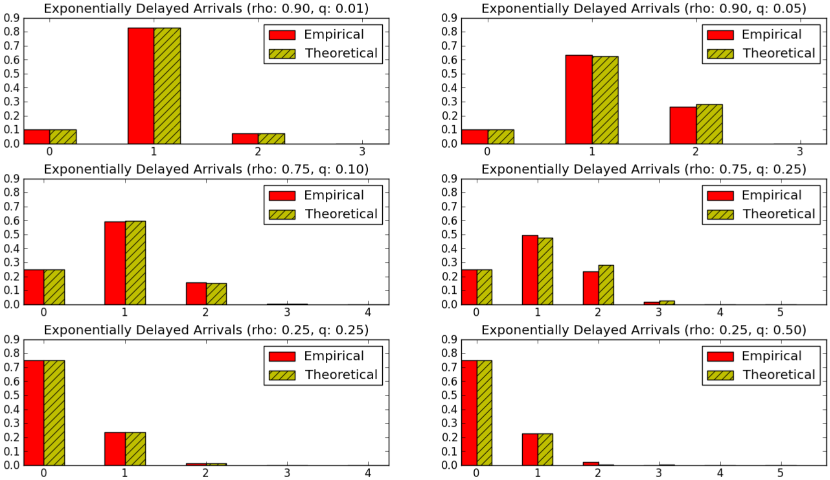
The coefficients $P^{(k)}(z, y)$ *satisfy* $P^{(0)}(z, y) = 1 + \rho(z - 1)$ *and*

$$P^{(k)}(z, y) = \sum_{j=1}^{k} \left[\frac{(y - z)^j}{j!} A_j^k(z) + \frac{1 + \rho(z - 1)}{1 - \rho} \, \frac{z^j - 1}{j!} A_j^k(0) \right]$$

where

$$\begin{aligned} A_j^k(z) &= j\rho \frac{\partial^{j-1}}{\partial y^{j-1}} \left(P^{(k-j)}(0, y) + \frac{P^{(k-j)}(z, y) - P^{(k-j)}(0, y)}{z} \right) \Bigg|_{y=z} \\ &+ [1 + \rho(z - 1)] \frac{\partial^j}{\partial y^j} \left(P^{(k-j)}(0, y) + \frac{P^{(k-j)}(z, y) - P^{(k-j)}(0, y)}{z} \right) \Bigg|_{y=z} \end{aligned}$$

Simulation vs Truncation to $k = 3$



The marginal distribution of late customers

If we evaluate in $z = 1$ the expression

$$P(z, y) = \frac{1 - \rho + \rho(z + q(y - z))}{z} \Big[(z - 1)P(0, z + q(y - z)) + P(z, z + q(y - z)) \Big],$$

we get

$$P(1, y) = [1 - \rho + \rho q(1 - y)]\, P(1, 1 + q(1 - y))$$

Iterating yields

$$P(1, y) = \sum_{l} y^{l} \sum_{n} P_{n,l} = \prod_{k \geqslant 1} (1 + \rho q^{k}(y - 1)).$$

Remarks

- $P(1, y) = \frac{(\rho(1-y); q)_\infty}{1+\rho(y-1)}$
- $(a; q)_\infty$ is the infinite q-*Pochhammer symbol*, aka q-ascending factorial in a

The marginal distribution of late customers

Theorem (Lancia et al, 2017)

The marginal distribution of the number of late customers satisfies

$$\sum_n P_{n,l} = \sum_{k \geqslant l} (-1)^{k-l} \rho^k \, q^{\binom{k+1}{2}} \binom{k}{l} \left[1 + \prod_{i=1}^{k} \frac{1}{1-q^i} \right].$$

- The marginal distribution decays super-exponentially fast in l.
- The factor $1 + \prod_{i=1}^{k} \frac{1}{1-q^i}$ is the generating function of the number of partitions into at most k parts;
- The problem shows a rich combinatorial structure.

Asymptotic result

- $\sum_n P_{n,l} = O(\rho^l \, q^{\binom{l+1}{2}})$
- Let $\alpha_t = n_t + l_t$
- The generating function of the equilibrium distribution of α_t is

$$\sum_a \sum_{n+l=a} P_{n,l} z^a = \sum_{n+l} P_{n,l} z^{n+l} = P(z,z)$$

Since

$$\sum_{n+l=a} P_{n,l} = \frac{1}{a!} \frac{d^a}{dz^a} P(z,z) = P_{0,a} + \frac{\rho}{1-\rho} P_{0,a-1} \, ,$$

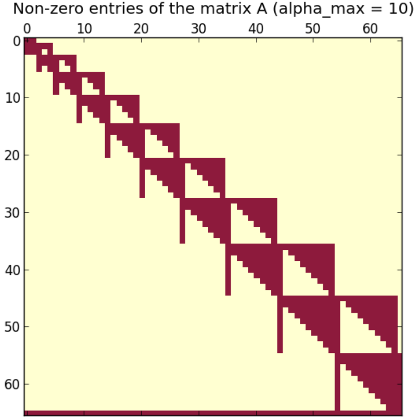
Theorem (Lancia et al, 2017)

For $a = n + l$,

$$P_{n,l} = O(\rho^a \, q^{\binom{a}{2}})$$

Approximation scheme

- We truncate the infinite system of balance equations
- Fix an integer α_{max}
- Impose $P_{n,l} = 0$ for $n + l > \alpha_{\mathrm{max}}$

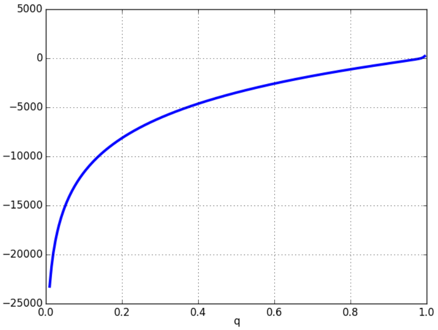


A priori bound

- From perturbation theory

$$\sum_{n,l<\alpha_{\max}} |P_{n,l}\tilde{P}_{n,l}| \leqslant \kappa(A)\epsilon_{n,l}$$

- $\epsilon_{n,l}$ can be uniformly bounded by $\frac{(-q;q)_\infty}{(q;q)_\infty} q^{\binom{\alpha_{\max}}{2}}$
- $\kappa(A) = \|A\|_1 \|A^{-1}\|_1$ is the condition number of the truncated matrix A
- Plot of the uniform bound for $\alpha_{\max} = 100$

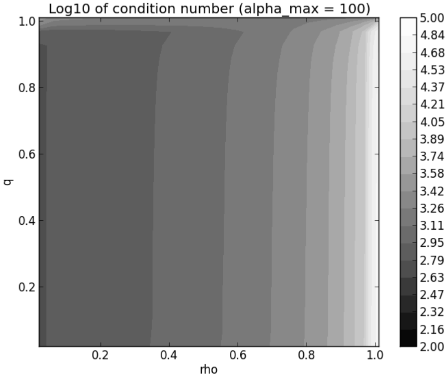


A priori bound

- From perturbation theory

$$\sum_{n,l<\alpha_{\max}} |P_{n,l}\tilde{P}_{n,l}| \leqslant \kappa(A)\epsilon_{n,l}$$

- $\epsilon_{n,l}$ can be uniformly bounded by $\frac{(-q;q)_\infty}{(q;q)_\infty} q^{\binom{\alpha_{\max}}{2}}$
- $\kappa(A) = \|A\|_1 \|A^{-1}\|_1$ is the condition number of the truncated matrix A
- Plot of the condition number for $\alpha_{\max} = 100$



Conclusions

- PSRA describes better the inbound flow than Poisson arrivals
- Poisson arrivals lead to overestimation of the queue
- Only advantage of Poisson is mathematical tractability
- If time scale is small, only the transient matters, *i.e. no benefit*
- PSRA are difficult to tackle, even in the *simple* case of EDA
- Efficient approximation schemes are possible, see Guadagni et al 2011

Acknowlegments

Lorenzo Capanna and Luigi De Giovanni for querying the DDR database

Preprints

- Lancia et al (2017) `https://arxiv.org/abs/1302.1999`
- Lancia, Lulli (2017) `https://arxiv.org/abs/1708.02486`