The $M/G/\infty$ estimation problem revisited

Alexander Goldenshluger

Department of Statistics

University of Haifa

Israel

Mathematical Statistics and Inverse Problems

Luminy, February 8–12, 2016

Talk outline

- The $M/G/\infty$ estimation problem
  - background and problem formulation
- Estimation from the input–output data
  - some results on the bivariate arrival–departure point process
  - estimator and its accuracy
- Estimation from the queue–length data
  - results on the queue–length process
  - estimator and its accuracy
- Conclusion

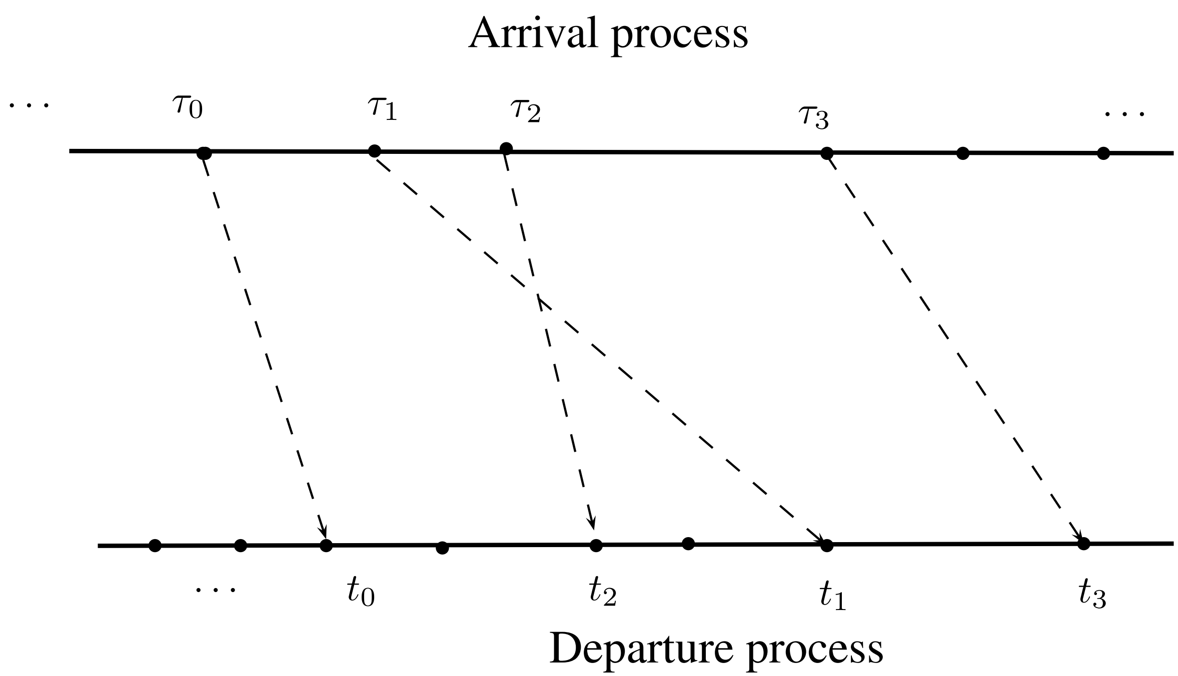
I. Background and problem formulation

1. The $M/G/\infty$ estimation problem

- Arrival process: customers come to a system according to a Poisson process of intensity λ.
- Service times: upon arrival, every customer obtains service and leaves the system after the service is completed. The service times are i.i.d. random variables, independent of the arrival process, with common distribution G.
- Observations: during some observation period arrival and departure time instances are recorded without matchings.
- Goal: estimate (make inference on) the service time distribution G.

2. The $M/G/\infty$ estimation problem

- The departure point process is obtained by translating the input points by i.i.d. random variables with distribution G. The correspondences (arrows) are not observed.



- The departure process is also Poisson of intensity λ.

Examples

- The $M/G/\infty$ model was used in many applications:
  - A telephone exchange model

    Beneš (1957),...
  - Mobility of particles

    dates back to Smoluchowski (1906)

    Bingham & Dunham (1997),...
  - A traffic density model

    Renyi (1964), Brown (1970)

“Sequence of differences” estimator of G

- Brown’s (1970) estimator:
  - Associate each output point t_j in $[t_0, t_n]$ with the nearest input point τ_k to the left of t_j. Call the corresponding distances between these points z_j, $j = 1, \ldots, n$.
  - The sequence $\{z_j\}$ is stationary and ergodic, z_j has distribution D:

$$D(x) = 1 - (1 - G(x))e^{-\lambda x} \quad \Leftrightarrow \quad G(x) = 1 - (1 - D(x))e^{\lambda x}.$$

  - Estimate D empirically and invert for G.
  - Consistency of the estimator is proved.

References and research questions

- Extensions:
  - *Blanghaps, Nov & Weiss (2013):* distances to rth nearest input point; consistency of the estimator...
  - *Schweer & Wichelhaus (2015):* discrete model, central limit theorem...

- Related literature:

  other observation schemes, inference for point processes

  Pickands & Stine (1997), Bingham & Pitts (1999), Hall & Park (2004),

  Brillinger (1972, 74, 75), Cox & Lewis (1972).

- Reserach questions:

  - estimation accuracy in the original $M/G/\infty$ problem?

  - how to construct estimators?

II. Estimation from the input–output data

The random translation model

- Input: M is a stationary Poisson process of intensity λ with the representation

$$M := \sum_{j\in\mathbb{Z}} \epsilon_{\tau_j}, \quad \epsilon_x(A) := \begin{cases} 1, & x \in A, \\ 0, & x \notin A, \end{cases} \quad x \in \mathbb{R},\ A \in \mathscr{B}.$$

- Output: $N = \mathscr{L}[M]$, where $\mathscr{L}$ is random translation, independent of M,

$$N := \sum_{j\in\mathbb{Z}} \epsilon_{t_j}, \quad t_j = \tau_j + \sigma_j, \quad \sigma_j \overset{iid}{\sim} G,$$

  $(\sigma_j)_{j\in\mathbb{Z}}$ are not necessarily non–negative random variables.

- Observation: a realization of the bivariate point process $(M, N)|_{\mathcal{T}}$, restricted to a time "window" $\mathcal{T} = \mathcal{T}_M \times \mathcal{T}_N$.

- The goal is to estimate (to make inference on) G.

Some properties of (M, N)

- Proposition 1: Let $\{A_i\}_{i=1,\dots,m}$ and $\{B_l\}_{l=1,\dots,n}$ be two families of disjoint intervals of the real line; then

$$\log \mathrm{E}_G \exp\left\{\sum_{i=1}^{m} \eta_i M(A_i) + \sum_{l=1}^{n} \xi_l N(B_l)\right\} = \lambda \sum_{i=1}^{m} (e^{\eta_i} - 1)|A_i|$$
$$+ \lambda \sum_{l=1}^{n} (e^{\xi_l} - 1)|B_l| + \lambda \sum_{i=1}^{m} \sum_{l=1}^{n} (e^{\eta_i} - 1)(e^{\xi_l} - 1) Q(A_i, B_l),$$

where $|\cdot|$ is the Lebesgue measure, and

$$Q(A, B) := \int_A G(B - x)\mathrm{d}x.$$

- Notation: $G(I) := G(b) - G(a)$, for $I = (a, b]$, $a < b$.

Remarks

- The case of $m = 2$ and $n = 2$ is proved in Milne (1970).
- (M, N) is closely related to Gauss–Poisson processes whose distribution is completely determined by the first two moment measures: the probability generating functional of (M, N) is

$$\begin{aligned}\mathscr{G}_{(M,N)}[\eta, \xi] &:= \mathrm{E}_G \exp\left\{\int \log \eta(\tau)\mathrm{d}M(\tau) + \int \log \xi(t)\mathrm{d}N(t)\right\} \\ &= \exp\left\{\lambda \int [\eta(\tau) - 1]\mathrm{d}\tau + \lambda \int [\xi(t) - 1]\mathrm{d}t\right. \\ &\qquad \left. + \lambda \iint [\eta(\tau) - 1)][\xi(t) - 1]Q(\mathrm{d}\tau, \mathrm{d}t)\right\}\end{aligned}$$

$\forall\, 0 \le \eta \le 1$, $0 \le \xi \le 1$ s.t. $1 - \eta$ and $1 - \xi$ vanish outside a compact interval, and $Q(\mathrm{d}\tau, \mathrm{d}t) = \mathrm{d}G(t - \tau)\mathrm{d}\tau$.

Proof outline

- Step 1: conditioning on (τ_j):

$$\mathrm{E}_G\left[e^{\sum_{i=1}^n \eta_i M(A_i)+\sum_{l=1}^m \xi_l N(B_l)}\middle|(\tau_j)\right] = \exp\left\{\sum_{j\in\mathbb{Z}} f(\tau_j)\right\},$$

where

$$f(x) := \sum_{i=1}^n \eta_i \mathbf{1}_{A_i}(x) + \log\left[\sum_{l=1}^m (e^{\xi_l}-1)G(B_l-x)+1\right].$$

- Step 2: the use of Campbell's formula

$$\mathrm{E}_G \exp\left\{\sum_j f(\tau_j)\right\} = \exp\left\{\lambda\int_0^\infty [e^{f(x)}-1]\mathrm{d}x\right\}.$$

Moment measures

- Corollary 1: For any two intervals A and B one has

$$\begin{aligned} \mathrm{E}_G\big[M(A)N(B)\big] &= \lambda^2|A|\cdot|B| + \lambda Q(A,B), \\ Q(A,B) &:= \int_A G(B-x)\mathrm{d}x. \end{aligned}$$

- Corollary 2: For two pairs of disjoint intervals A_1, A_2 and B_1, B_2

$$\begin{aligned} &\mathrm{E}_G\big[M(A_1)M(A_2)N(B_1)N(B_2)\big] \\ &\quad - \mathrm{E}_G\big[M(A_1)N(B_1)\big]\cdot\mathrm{E}_G\big[M(A_2)N(B_2\big] \\ &\quad = \lambda^3\Big[Q(A_1,B_2)|A_1|\cdot|B_1| + Q(A_2,B_1)|A_1|\cdot|B_2|\Big] \\ &\qquad\qquad\qquad\qquad + \lambda^2 Q(A_1,B_2)Q(A_2,B_1). \end{aligned}$$

- The proof is by differentiation of the formula in Proposition 1.

Important formula

- Corollary 3:

  For any function v satisfying

$$\iint |v(\tau,t)|\mathrm{d}\tau\mathrm{d}t < \infty, \quad \iint |v(\tau,t)|\mathrm{d}G(t-\tau)\mathrm{d}\tau < \infty,$$

$$\begin{aligned} \mathrm{E}_G\Big[\iint v(\tau,t)\mathrm{d}M(\tau)\mathrm{d}N(t)\Big] &= \mathrm{E}_G\Big[\sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} v(\tau_j,t_k)\Big] \\ &= \lambda^2\iint v(\tau,t)\mathrm{d}\tau\mathrm{d}t + \lambda\iint v(\tau,t)\mathrm{d}G(t-\tau)\mathrm{d}\tau. \qquad (*) \end{aligned}$$

- Immediate consequence of Corollary 1.

- Corollary 1 as well as $(*)$ are known results; see e.g., Cox & Lewis (1972), Mori (1975).

Back to statistical problem...

- Input: Poisson point process of intensity λ on $\mathbb{R}$,

  $M = \sum_{j\in\mathbb{Z}} \epsilon_{\tau_j}$.

- Output: $N = \sum_{j\in\mathbb{Z}} \epsilon_{t_j}$, $t_j = \tau_j + \sigma_j$, $\sigma_j \overset{iid}{\sim} G$.

- Observations: $\mathscr{D}_{\mathcal{T}} := (M, N)\big|_{\mathcal{T}}$, $\quad \mathcal{T} = \mathcal{T}_M \times \mathcal{T}_N \subset \mathbb{R} \times \mathbb{R}$.

- The goal is to estimate G; in fact, for $I = (a, b]$ we consider the problem of estimating

$$\theta_I = G(I) := G(b) - G(a).$$

- Risk: for any estimator $\hat{\theta}_I = \hat{\theta}_I(\mathscr{D}_{\mathcal{T}})$

$$\mathrm{Risk}[\hat{\theta}_I; \theta_I] := \mathrm{E}_G|\hat{\theta}_I - \theta_I|^2.$$

Estimator

- Data: realization of $(M,N)|_{\mathcal{T}}$ restricted to

$$\mathcal{T} = [\tau_{\min}, \tau_{\max}] \times [\tau_{\min} + a, \tau_{\max} + b], \quad T := \tau_{\max} - \tau_{\min},$$

so that

$$\mathscr{D}_{\mathcal{T}} = \Big\{ (\tau_j : \tau_{\min} \le \tau_j \le \tau_{\max}),\ (t_k : \tau_{\min} + a \le t_k \le \tau_{\max} + b) \Big\}.$$

- Estimator: Let $I := (a,b]$, $v_0(\tau,t) := \mathbf{1}_{[\tau_{\min},\tau_{\max}]}(\tau)\mathbf{1}_I(t-\tau)$; and

$$\begin{aligned} \hat{\theta}_I &:= \tfrac{1}{\lambda T} \iint v_0(\tau,t)\mathrm{d}M(\tau)\mathrm{d}N(t) - \lambda|I| \\ &= \tfrac{1}{\lambda T} \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} \mathbf{1}_{[\tau_{\min},\tau_{\max}]}(\tau_j)\mathbf{1}_I(t_k - \tau_j) - \lambda|I|. \end{aligned}$$

Accuracy of $\hat{\theta}_I$

- Theorem 1:

  For any G one has

$$
\begin{aligned}
\mathrm{E}_G\left[\hat{\theta}_I\right] &= \theta_I = G(I) = G(b) - G(a), \\
\mathrm{var}_G\{\hat{\theta}_I\} &= \tfrac{2\lambda|I|}{T}\int_{-T}^{T} G(I+u)\Big(1-\tfrac{|u|}{T}\Big)\mathrm{d}u \\
&\quad + \tfrac{1}{T}\int_{-T}^{T} G(I+u)G(I-u)\Big(1-\tfrac{|u|}{T}\Big)\mathrm{d}u. \\
&\quad + \tfrac{\lambda|I|^2}{T} + \tfrac{2|I|}{T}G(I) + \tfrac{1}{\lambda T}\big(\theta_I + \lambda|I|\big).
\end{aligned}
$$

For the $M/G/\infty$ setting...

- Estimator: in the $M/G/\infty$ setting $G(0) = 0$, $[\tau_{\min}, \tau_{\max}] = [0, T]$, $I = (0, x_0]$, so that the estimator is given by

$$\hat{G}(x_0) \quad = \quad \tfrac{1}{\lambda T} \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \mathbf{1}_{[0,T]}(\tau_j) \mathbf{1}_{[0,x_0]}(t_k - \tau_j) - \lambda x_0.$$

- Notation
  - *service rate* μ: $\frac{1}{\mu} := E_G[\sigma] = \int_0^\infty [1 - G(u)]\mathrm{d}u$;
  - *traffic intensity:* $\rho = \lambda / \mu$;
  - *normalized integrated tail of* G:

$$H(t) := \mu \int_t^\infty [1 - G(u)]\mathrm{d}u = \frac{\int_t^\infty [1 - G(u)]\mathrm{d}u}{\int_0^\infty [1 - G(u)]\mathrm{d}u}.$$

The result for the $M/G/\infty$ estimation problem

- Corollary 4: Let $G(0) = 0$, $x_0 \in (0, T)$; then $\hat{G}(x_0)$ is unbiased,

$$\mathrm{var}_G\{\hat{G}(x_0)\} = \tfrac{2\lambda x_0}{T} \int_{-T}^{T} [G(x_0 + u) - G(u)]\Big(1 - \tfrac{|u|}{T}\Big)\mathrm{d}u$$

$$+ \tfrac{1}{T} \int_{-T}^{T} [G(x_0 + u) - G(u)][G(x_0 - u) - G(-u)]\Big(1 - \tfrac{|u|}{T}\Big)\mathrm{d}u.$$

$$+ \tfrac{\lambda x_0^2}{T} + \tfrac{2x_0}{T} G(x_0) + \tfrac{1}{\lambda T}[G(x_0) + \lambda x_0].$$

Moreover, if $\frac{1}{\mu} := E_G[\sigma] = \int_0^\infty [1 - G(u)]\mathrm{d}u < \infty$ then

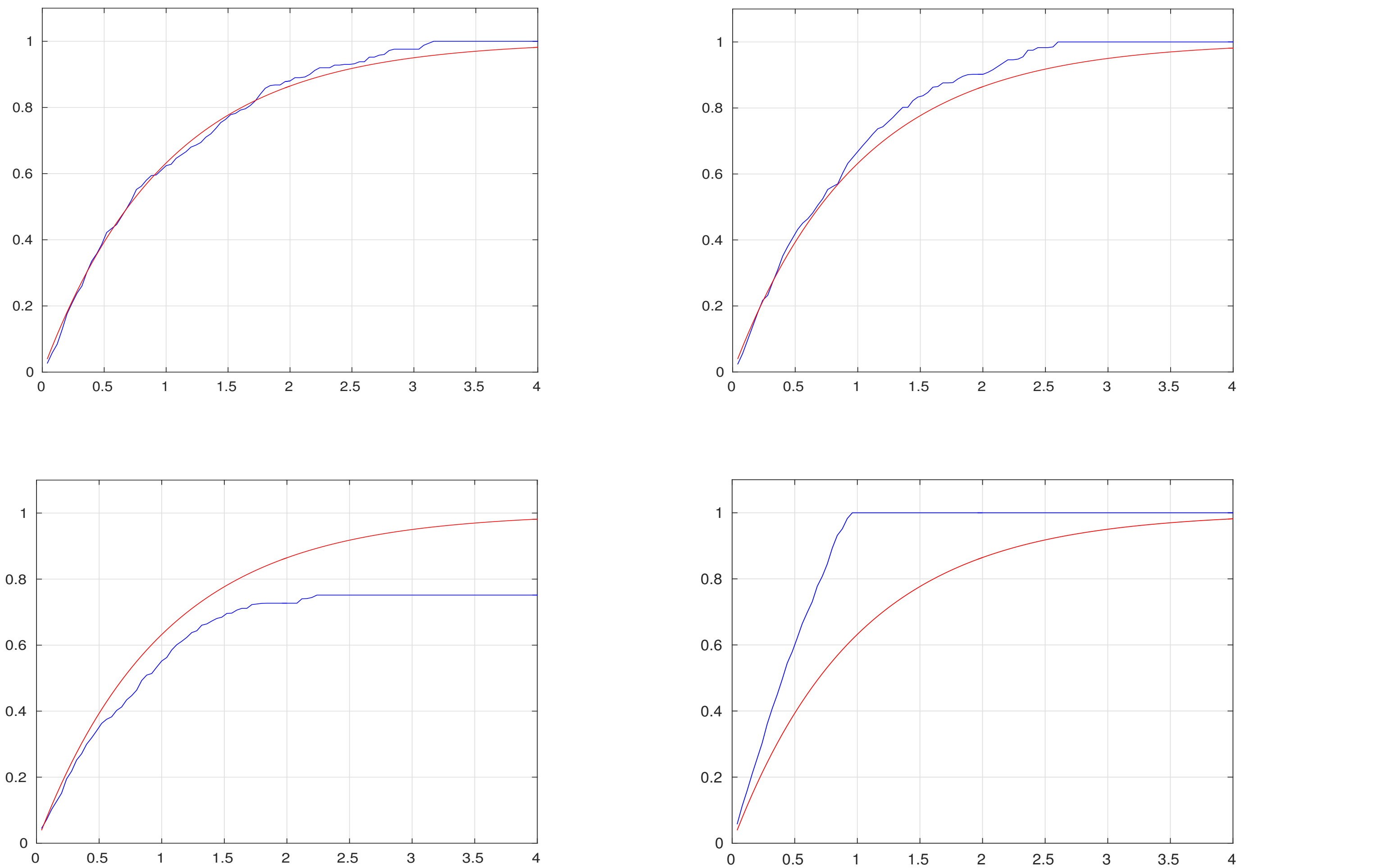
$$\mathrm{var}_G\{\hat{G}(x_0)\} \ \le \ \tfrac{C}{T}\Big[\lambda x_0^2 + \lambda x_0 \min\{\tfrac{1}{\mu}[1 - H(x_0)], x_0\} + x_0 + \tfrac{1}{\lambda} G(x_0)\Big].$$

Remarks

- Proof outline:
  - the estimator is unbiased in view of formula $(*)$;
  - the variance is calculated using formulas for the moment measures, established in Corollary 2.
- The $M/G/\infty$ setting:
  - $G(x_0)$ can be estimated with parametric rate $O(\frac{1}{T})$;
  - the estimator is not accurate for "large" λ and x_0.
- Can we do better?

Numerical illustration

- Setting: $\sigma \sim \exp(1),\ T = 1000,\ \lambda \in \{0.5, 1, 5, 15\}$

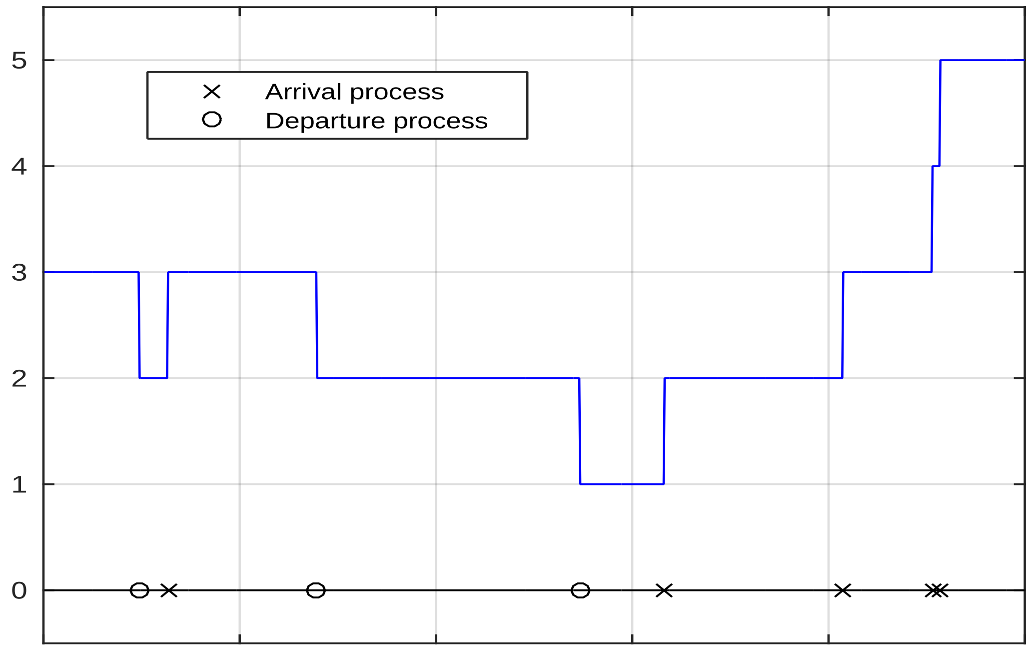


III. Estimation from the queue–length data

Another look at the $M/G/\infty$ problem

- Queue–length (number of busy servers) process $X(t)$ encodes input–output streams up to initial conditions:

$$X(t) = \sum_{j \in \mathbb{Z}} \mathbf{1}\{\tau_j \le t,\ \sigma_j > t - \tau_j\}, \quad t \in \mathbb{R}.$$



- Assume that $\{X(k\delta),\ k = 1, \ldots, n,\ T = n\delta\}$ is observed...

1. Properties of the queue–length process

Proposition 2:

* $X(t) \sim \mathrm{Poisson}(\rho)$, $\forall t \in \mathbb{R}$, where $\rho = \lambda/\mu$.

* $\{X(t), t \in \mathbb{R}\}$ is stationary, and

$$\mathrm{cov}_G\{X(t), X(s)\} = \rho H(t-s), \quad \forall t, s \in \mathbb{R}.$$

* Let $X = (X(t_1), \ldots, X(t_n)) = (X_1, \ldots, X_n)$; then for any $\theta \in \mathbb{R}^n$

$$\log \mathrm{E}_G\big[\exp\{\theta^T X\}\big] = \rho S_n(\theta),$$

$$S_n(\theta) := \sum_{k=1}^{n} (e^{\theta_k} - 1) + \sum_{k=1}^{n-1} H_k \sum_{m=k}^{n-1} \big(e^{\theta_{m-k+1}} - 1\big) e^{\sum_{i=m-k+2}^{m} \theta_i} \big(e^{\theta_{m+1}} - 1\big),$$

where $H_k := H(t_k)$, $k = 1, \ldots, n$.

2. Properties of the queue–length process

- Covariance function of $\{X(t)\}$:

$$\begin{aligned} R(t) &:= \operatorname{cov}_G\{X(s), X(s+t)\} = \rho H(t) \\ &= \rho \cdot \frac{\int_t^\infty [1-G(u)]\mathrm{d}u}{\int_0^\infty [1-G(u)]\mathrm{d}u} = \lambda \int_t^\infty [1-G(u)]\mathrm{d}u. \end{aligned}$$

Hence,

$$1 - G(t) = -\frac{1}{\lambda} R'(t), \quad t \in \mathbb{R}_+. \qquad (**)$$

- The idea is to estimate the first derivative of the covariance function of $X(t)$ at point x_0, and then recover $G(x_0)$ from $(**)$.

1. Estimator construction

- Estimators of $R_k := R(k\delta)$: $\hat{\rho} = \frac{1}{n}\sum_{t=1}^{n} X_t$

$$\hat{R}_k := \frac{1}{n-k}\sum_{t=1}^{n-k}(X_t - \hat{\rho})(X_{t+k} - \hat{\rho}), \qquad k = 0, 1, \ldots, n-1.$$

- Local window: Let $h > 0$, $D_x := [x-h, x+h]$, $\forall x \in [h, T-h]$, and $M_{D_x} = \{k : k\delta \in D_x\}$, $N_{D_x} = \#\{M_{D_x}\}$.
- Differentiating filter: Fix positive integer ℓ, real $h \geq \frac{1}{2}(\ell+2)\delta$, and let $\{a_k(x), k \in M_{D_x}\}$ be the solution to

$$\begin{aligned} \min \quad & \sum_{k \in M_{D_x}} a_k^2(x) \\ \text{s.t.} \quad & \sum_{k \in M_{D_x}} a_k(x) = 0, \\ & \sum_{k \in M_{D_x}} a_k(x)(k\delta)^j = jx^{j-1}, \quad j = 1, \ldots, \ell. \end{aligned}$$

2. Estimator construction

- Remarks
  - $h \geq \frac{1}{2}(\ell+2)\delta$ ensures at least $\ell+1$ grid points in M_{D_x}.
  - The filter reproduces the first derivative of any polynomial of degree $\leq \ell$:

$$\sum_{k \in M_{D_x}} a_k(x) p(k\delta) = p'(x), \quad \forall p : \deg(p) \leq \ell.$$

- Estimator of $G(x_0)$:

$$\tilde{G}_h(x_0) = 1 + \tfrac{1}{\lambda} \sum_{k \in M_{D_{x_0}}} a_k(x_0) \hat{R}_k.$$

- Two design parameters to be chosen:

degree of the fitted polynomial ℓ and window width h.

Functional class

- Local smoothness: let $\beta > 0$, $L > 0$ and $I \subset (0, \infty)$ be a closed interval containing x_0. We say that $G \in \mathcal{H}_\beta(L, I)$ if
$$|G^{(\lfloor\beta\rfloor)}(x) - G^{(\lfloor\beta\rfloor)}(y)| \le L|x-y|^{\beta-\lfloor\beta\rfloor}, \;\; \forall x, y \in I,$$
where $\lfloor\beta\rfloor := \max\big\{k \in \{0,1,2,\dots\big\} : k < \beta\}$.

- Tail (moment) conditions: we say that $G \in \mathcal{M}_p(K)$ with $p \ge 1$, $K > 0$ if
$$\mathrm{E}_G[\sigma^p] = \int_0^\infty p x^{p-1}[1 - G(x)]\mathrm{d}x \le K < \infty.$$
If $G \in \mathcal{M}_2(K)$ then $\{H_i\}$ is summable $\Rightarrow$ *short–range dependence*.

- Functional class: we consider
$$\Sigma_\beta = \Sigma_\beta(L, I, K) := \mathcal{H}_\beta(L, I) \cap \mathcal{M}_2(K).$$

Upper bound

- Theorem 2:

  Let $I = [x_0 - d, x_0 + d] \subset [0, (1-\varkappa)T]$ for some $\varkappa \in (0,1)$. Let $\tilde{G}_*(x_0)$ be the estimator $\tilde{G}_{h_*}(x_0)$ associated with

  $$\ell \geq \lfloor \beta \rfloor + 1, \quad h_* = \left[\frac{K}{L^2 \varkappa T}\left(1 + \tfrac{1}{\lambda}\right)\right]^{1/(2\beta+2)}.$$

  If

  $$\frac{K}{L^2 \varkappa}\left(1 + \tfrac{1}{\lambda}\right) d^{-2\beta-2} \;\leq\; T \;\leq\; \frac{K}{L^2 \varkappa}\left(1 + \tfrac{1}{\lambda}\right)\left[\frac{2}{(\ell+2)\delta}\right]^{2\beta+2}$$

  then

  $$\sup_{G \in \Sigma_\beta}\left[\mathrm{E}_G |\tilde{G}_*(x_0) - G(x_0)|^2\right]^{1/2} \;\leq\; C(\ell) L^{1/(\beta+1)}\left[\frac{K}{\varkappa T}\left(1 + \tfrac{1}{\lambda}\right)\right]^{\beta/(2\beta+2)}.$$

Remarks

- Under local smoothness and second moment conditions:

$$\mathrm{Risk}_{x_0}[\tilde{G}_*;\Sigma_\beta] := \sup_{G\in\Sigma_\beta}\left[\mathrm{E}_G|\tilde{G}_*(x_0)-G(x_0)|^2\right]^{1/2}$$

$$\asymp C\left[\tfrac{1}{T-x_0}(1+\tfrac{1}{\lambda})\right]^{\beta/(2\beta+2)},\ \ T\to\infty.$$

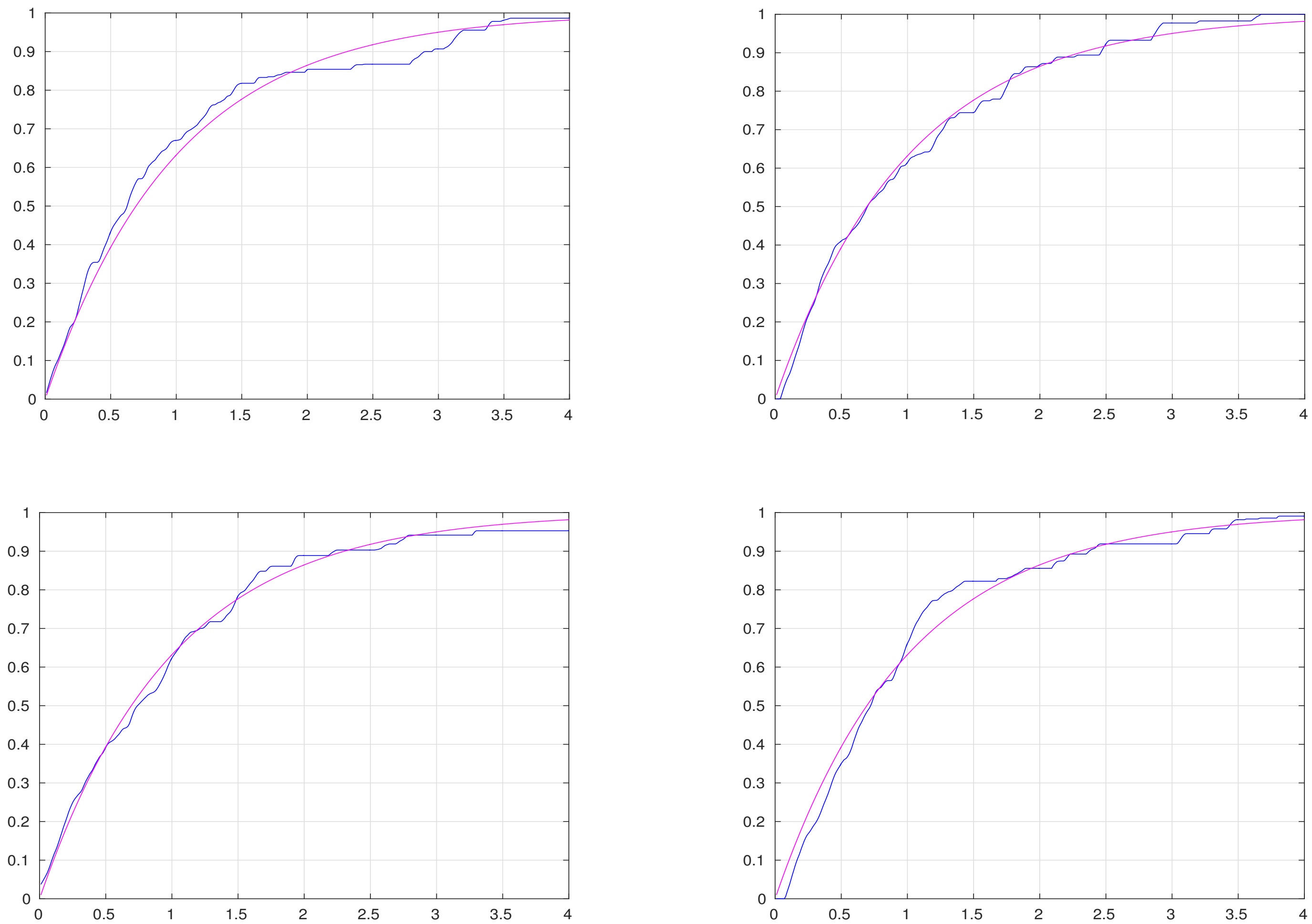
- The rate of convergence is nonparametric:

$$\inf_{\tilde{G}}\mathrm{Risk}_{x_0}[\tilde{G};\Sigma_\beta] \le O\big(T^{-\beta/(2\beta+2)}\big),\quad T\to\infty,$$

but dependence on λ and x_0 is "weak".

Numerical illustration

- Setting: $\sigma \sim \exp(1)$, $T = 1000$, $\delta = 0.01$ $h = 3\delta$, $\lambda \in \{1, 10, 40, 100\}$



Comparison of the estimators

- Different regimes in terms of the convergence rate:
  - *light traffic regime*: $x_0\sqrt{\frac{\rho}{T}} \ll C(T-x_0)^{-\beta/(2\beta+2)}$
  - *heavy traffic regime*: $x_0\sqrt{\frac{\rho}{T}} \gg C(T-x_0)^{-\beta/(2\beta+2)}$
- Numerical behavior:
  - *dependence on λ:* even undersmoothed estimator $\tilde{G}_h(x_0)$ is much better than $\hat{G}(x_0)$ already for moderate λ;
  - *heavy tails:* $\tilde{G}_h(x_0)$ is more sensitive to heavy tails of G than $\hat{G}(x_0)$.
- Lower bounds on the risk?

  Difficult because of the dependence structure; the distribution of observations is not available explicitly.

Gaussian approximation in heavy traffic

- Corollary to Proposition 2:

  Let $\{M_\ell/G/\infty,\ \ell = 1, 2, \ldots\}$ be a sequence of the $M/G/\infty$ systems with fixed G and arrival rates $\lambda_\ell = \ell\lambda$, $\lambda > 0$. Let $X_\ell^n = (X_\ell(t_1), \ldots, X_\ell(t_n))$ be observations of the queue–length process in the ℓth system; then

$$\frac{X_\ell^n - \ell\rho e_n}{\sqrt{\ell\rho}} \xrightarrow{d} \mathcal{N}_n(0, V(H)), \quad \ell \to \infty,$$

  where $\rho = \frac{\lambda}{\mu}$, $e_n = (1, \ldots, 1) \in \mathbb{R}^n$, $V(H) = \{H((i-j)\delta)\}_{i,j=1,\ldots,n}$.

- This result is in line with general results of Borovkov (1967), Iglehart (1973) and Whitt (1974) on weak convergence for queues.

A Gaussian model

- In heavy traffic $\{X(t)\}$ is close to a stationary Gaussian process. By $(**)$, G is proportional to the derivative of the covariance function.

- A problem for stationary Gaussian process:

  Let $\{X(t),\ t \in \mathbb{R}\}$ be a stationary Gaussian process with zero mean and covariance function γ. We observe $X^n = (X(t_1), \dots, X(t_n))$, $t_i = i\delta$, $i = 1, \dots, n$, $n\delta = T$.

- The goal is to estimate $\theta = \gamma'(x_0)$ using observation X^n. We are mainly interested in lower bounds on the minimax risk

$$\mathrm{Risk}^*_{x_0}[\Gamma] = \inf_{\hat{\theta}} \sup_{\gamma \in \Gamma} \left[\mathrm{E}_\gamma \left| \hat{\theta} - \gamma'(x_0) \right|^2 \right]^{1/2},$$

  where Γ is a suitable class of covariance functions.

Lower bound in the Gaussian problem

- Definition: Let $I = [x_0 - d, x_0 + d]$, $L > 0$ and $\beta > 0$. We say that $\gamma \in \Gamma_\beta := \Gamma_\beta(L, I, K)$ if
  (i) $\int_{-\infty}^{\infty} |\gamma(t)| \mathrm{d}t \le K < \infty$;
  (ii) γ is $\ell := \max\{k \in \mathbb{N} : k < \beta + 1\}$ times continuously differentiable on I and
$$|\gamma^{(\ell)}(x) - \gamma^{(\ell)}(y)| \le L|x - y|^{\beta+1-\ell}, \ \ \forall x, y \in I.$$
- Theorem 2:

  There exist positive constants C_1, C_2 and c depending on β, x_0, d and K only such that if
$$C_1 \delta^{-2} \le T, \qquad L^2 T \le C_2 \delta^{-2\beta-2}$$
  then
$$\liminf_{T\to\infty} \left\{ L^{-1/(\beta+1)} T^{\beta/(2\beta+2)} \, \mathrm{Risk}^*_{x_0}[\Gamma_\beta] \right\} \ge c > 0.$$

Concluding remarks

- Two different estimators with different properties...
- The lower bound in the Gaussian model strongly suggests that the "queue–length"–based estimator is rate optimal in the heavy traffic regime...
- In the light traffic regime it is not clear if estimators with better dependence on λ and x_0 can be constructed...
- A fundamental question: how to judge optimality of estimators when the joint distribution of observations is intractable?