Convergence rates of Laplace-transform based estimators

Arnoud den Boer (University of Twente)

joint work with Michel Mandjes (University of Amsterdam),
Sindo Núñez Queija (University of Amsterdam),
Piotr Zuraniewski (TNO).

Project partially funded by Surfnet

Outline

1. Motivation from practical problem
2. Convergence rates of Laplace-transform based estimators
3. Applications

Outline

1. Motivation from practical problem
2. Convergence rates of Laplace-transform based estimators
3. Applications

Motivation

- Surfnet: network operator for universities and hospitals.

Motivation

- Surfnet: network operator for universities and hospitals.
- Contractual agreements about network usage & performance.

Motivation

- Surfnet: network operator for universities and hospitals.
- Contractual agreements about network usage & performance.
- Validation of these agreements via probing.

Motivation

- Surfnet: network operator for universities and hospitals.
- Contractual agreements about network usage & performance.
- Validation of these agreements via probing.
- More probes = higher accuracy. But: probes affect speed of regular network traffic.

Question from Surfnet: What is the optimal probing frequency?

Model

- Consider an isolated network link.

Model

- Consider an isolated network link.
- Model this as $M/G/1$ queue in stationarity.

Model

- Consider an isolated network link.
- Model this as $M/G/1$ queue in stationarity.
- Arrival rate λ and service time distribution B unknown.
  $\lambda\mathbb{E}[B] < 1$, jobs served at unit speed.
  Stationary workload/waiting time distribution W.

Model

- Consider an isolated network link.
- Model this as $M/G/1$ queue in stationarity.
- Arrival rate λ and service time distribution B unknown.
  $\lambda\mathbb{E}[B] < 1$, jobs served at unit speed.
  Stationary workload/waiting time distribution W.
- Send probing packets at equidistant time points to stationary queue.
  Record the delay.

Model

- Consider an isolated network link.
- Model this as $M/G/1$ queue in stationarity.
- Arrival rate λ and service time distribution B unknown.
  $\lambda\mathbb{E}[B] < 1$, jobs served at unit speed.
  Stationary workload/waiting time distribution W.
- Send probing packets at equidistant time points to stationary queue.
  Record the delay.

Goal: Estimate $P(W > w)$, for some predefined $w > 0$.

Empirical Estimator

- For now: assume probes have neglible job size.

Empirical Estimator

- For now: assume probes have neglible job size.
- Send probe i at time $i \cdot \delta$, $i = 1, \ldots, n$.
  Measured delay equals workload $W_{i\delta}$ at time $i\delta$.

Empirical Estimator

- For now: assume probes have neglible job size.
- Send probe i at time $i \cdot \delta$, $i = 1, \ldots, n$.
  Measured delay equals workload $W_{i\delta}$ at time $i\delta$.
- Empirical estimator

$$P(\widehat{W > } w) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{W_{i\delta} > w\}$$

Empirical Estimator

- For now: assume probes have neglible job size.
- Send probe i at time $i \cdot \delta$, $i = 1, \ldots, n$.
  Measured delay equals workload $W_{i\delta}$ at time $i\delta$.
- Empirical estimator

$$P(\widehat{W > w}) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{W_{i\delta} > w\}$$

Disadvantage of empirical estimator:

1. Hard to analyze finite-sample estimation error as function of δ, n (due to correlations).
2. No intervention at rare events $P(W > w)$.

Alternative Estimator

$$Q_i := W_{(i+1)\delta} - (W_{i\delta} - \delta)$$

If $W_{i\delta} \geq \delta$, then $Q_i =$ the amount of arrived work during $((i-1)\delta, i\delta]$.

If $W_{i\delta} < \delta$, then
$Q_i \geq$ the amount of arrived work during $((i-1)\delta, i\delta] \geq W_{i\delta}$.

Alternative Estimator

$$Q_i := W_{(i+1)\delta} - (W_{i\delta} - \delta)$$

If $W_{i\delta} \geq \delta$, then $Q_i =$ the amount of arrived work during $((i-1)\delta, i\delta]$.

If $W_{i\delta} < \delta$, then
$Q_i \geq$ the amount of arrived work during $((i-1)\delta, i\delta] \geq W_{i\delta}$.

The uncensored observations are i.i.d. distributed as

$$Q = \sum_{i=1}^{\text{Poiss}(\lambda\delta)} B_i$$

Alternative Estimator

$$Q_i := W_{(i+1)\delta} - (W_{i\delta} - \delta)$$

If $W_{i\delta} \geq \delta$, then $Q_i =$ the amount of arrived work during $((i-1)\delta, i\delta]$.

If $W_{i\delta} < \delta$, then
$Q_i \geq$ the amount of arrived work during $((i-1)\delta, i\delta] \geq W_{i\delta}$.

The uncensored observations are i.i.d. distributed as

$$Q = \sum_{i=1}^{\text{Poiss}(\lambda\delta)} B_i$$

Idea: estimate $P(W > w)$ based on this sample.

Alternative Estimator

$$Q_i := W_{(i+1)\delta} - (W_{i\delta} - \delta)$$

If $W_{i\delta} \geq \delta$, then $Q_i =$ the amount of arrived work during $((i-1)\delta, i\delta]$.

If $W_{i\delta} < \delta$, then
$Q_i \geq$ the amount of arrived work during $((i-1)\delta, i\delta] \geq W_{i\delta}$.

The uncensored observations are i.i.d. distributed as

$$Q = \sum_{i=1}^{\text{Poiss}(\lambda\delta)} B_i$$

Idea: estimate $P(W > w)$ based on this sample.

Advantage: i.i.d. rvs: easier to analyze.
Disadvantage: neglect censored observations.

Alternative Estimator

Write $\tilde{X}(s) = \mathbb{E}[\exp(-sX)]$, $s \in \mathbb{C}_+$.

Pollaczek-Khinchine formula.

$$\tilde{W}(s) = \frac{s(1-\lambda\mathbb{E}[B])}{s-\lambda+\lambda\tilde{B}(s)}, \quad (s \in \mathbb{C}_+)$$

Alternative Estimator

Write $\tilde{X}(s) = \mathbb{E}[\exp(-sX)]$, $s \in \mathbb{C}_+$.

Pollaczek-Khinchine formula.

$$\tilde{W}(s) = \frac{s(1 - \lambda\mathbb{E}[B])}{s - \lambda + \lambda\tilde{B}(s)}, \quad (s \in \mathbb{C}_+)$$

together with

$$\tilde{Q}(s) = \exp[-\lambda\delta + \lambda\delta\tilde{B}(s)], \quad \mathbb{E}[Q] = \lambda\delta\mathbb{E}[B],$$

Alternative Estimator

Write $\tilde{X}(s) = \mathbb{E}[\exp(-sX)]$, $s \in \mathbb{C}_+$.

Pollaczek-Khinchine formula.

$$\tilde{W}(s) = \frac{s(1 - \lambda\mathbb{E}[B])}{s - \lambda + \lambda\tilde{B}(s)}, \quad (s \in \mathbb{C}_+)$$

together with

$$\tilde{Q}(s) = \exp[-\lambda\delta + \lambda\delta\tilde{B}(s)], \quad \mathbb{E}[Q] = \lambda\delta\mathbb{E}[B],$$

yields

$$\tilde{W}(s) = \frac{s(1 - \delta^{-1}\mathbb{E}[Q])}{s + \delta^{-1}\mathrm{Log}\,\tilde{Q}(s)}, \quad (s \in \mathbb{C}_+)$$

Alternative Estimator

$$\tilde{W}(s) = \frac{s(1 - \delta^{-1}\mathbb{E}[Q])}{s + \delta^{-1}\mathrm{Log}\,\tilde{Q}(s)}, \quad (s \in \mathbb{C}_+)$$

Alternative Estimator

$$\tilde{W}(s) = \frac{s(1-\delta^{-1}\mathbb{E}[Q])}{s+\delta^{-1}\mathrm{Log}\,\tilde{Q}(s)}, \quad (s \in \mathbb{C}_+)$$

Estimator of $P(W \leq w)$:

1. Replace $\mathbb{E}[Q]$ and $\tilde{Q}(s)$ by $n^{-1}\sum_{i=1}^{n} Q_i$ and $n^{-1}\sum_{i=1}^{n} \exp(-sQ_i)$.

Alternative Estimator

$$\tilde{W}(s) = \frac{s(1 - \delta^{-1}\mathbb{E}[Q])}{s + \delta^{-1}\mathrm{Log}\,\tilde{Q}(s)}, \quad (s \in \mathbb{C}_+)$$

Estimator of $P(W \leq w)$:

1. Replace $\mathbb{E}[Q]$ and $\tilde{Q}(s)$ by $n^{-1}\sum_{i=1}^{n} Q_i$ and $n^{-1}\sum_{i=1}^{n} \exp(-sQ_i)$.
2. Divide by s to obtain estimated Laplace Transform of $P(W \leq w)$.

Alternative Estimator

$$\tilde{W}(s) = \frac{s(1 - \delta^{-1}\mathbb{E}[Q])}{s + \delta^{-1}\mathrm{Log}\,\tilde{Q}(s)}, \quad (s \in \mathbb{C}_+)$$

Estimator of $P(W \leq w)$:

1. Replace $\mathbb{E}[Q]$ and $\tilde{Q}(s)$ by $n^{-1}\sum_{i=1}^{n} Q_i$ and $n^{-1}\sum_{i=1}^{n}\exp(-sQ_i)$.
2. Divide by s to obtain estimated Laplace Transform of $P(W \leq w)$.
3. Apply an inverse Laplace transform procedure.

Alternative Estimator

$$\tilde{W}(s) = \frac{s(1-\delta^{-1}\mathbb{E}[Q])}{s+\delta^{-1}\mathrm{Log}\,\tilde{Q}(s)}, \quad (s \in \mathbb{C}_+)$$

Estimator of $P(W \leq w)$:

1. Replace $\mathbb{E}[Q]$ and $\tilde{Q}(s)$ by $n^{-1}\sum_{i=1}^{n} Q_i$ and $n^{-1}\sum_{i=1}^{n} \exp(-sQ_i)$.
2. Divide by s to obtain estimated Laplace Transform of $P(W \leq w)$.
3. Apply an inverse Laplace transform procedure.

Goal: find useful bounds on estimation error

Related literature

Decompounding: Buchmann and Grübel (2003,2004), Van Es et al. (2007), Bogsted and Pitts (2010), Duval (2013), Comte et al. (2014abc).

Estimation in queueing and insurance models: Glynn and Torres (1996, 1997), Zeevi and Glynn (2004), Hall and Park (2004), Hansen and Pitts (2006), Mnatsakov et al. (2008), Zhang and Xu (2010), Shimizu (2012), Blanchet et al. (2013).

Probing: Bacelli et al. (2009), Antunes and Pipiras (2001).

Related literature

Decompounding: Buchmann and Grübel (2003,2004), Van Es et al. (2007), Bogsted and Pitts (2010), Duval (2013), Comte et al. (2014abc).

Estimation in queueing and insurance models: Glynn and Torres (1996, 1997), Zeevi and Glynn (2004), Hall and Park (2004), Hansen and Pitts (2006), Mnatsakov et al. (2008), Zhang and Xu (2010), Shimizu (2012), Blanchet et al. (2013).

Probing: Bacelli et al. (2009), Antunes and Pipiras (2001).

Outline

Laplace-transform based estimation

General setting:

- $\mathcal{X}$ a collection of nonnegative random variables.
- $\tilde{\mathcal{X}}$ Laplace transforms of $\mathcal{X}$.
- Known map

$$\Psi : \tilde{\mathcal{X}} \to \{g : \mathbb{C}_+ \to \mathbb{C}\},$$

with $\tilde{Y} = \Psi \tilde{X}$ for some unknown X, Y nonnegative.

Laplace-transform based estimation

General setting:

- $\mathcal{X}$ a collection of nonnegative random variables.
- $\tilde{\mathcal{X}}$ Laplace transforms of $\mathcal{X}$.
- Known map

$$\Psi : \tilde{\mathcal{X}} \to \{g : \mathbb{C}_+ \to \mathbb{C}\},$$

  with $\tilde{Y} = \Psi \tilde{X}$ for some unknown X, Y nonnegative.
- Estimate cdf $F^Y(w)$ of Y, for given $w > 0$, based on i.i.d. sample $X_1, \ldots, X_n$ of X.

Laplace-transform based estimation

1) Plug-in estimator of $\tilde{X}$:

$$\tilde{X}_n(s) = n^{-1} \sum_{i=1}^{n} \exp(-sX_i).$$

Laplace-transform based estimation

1) Plug-in estimator of $\tilde{X}$:

$$\tilde{X}_n(s) = n^{-1} \sum_{i=1}^{n} \exp(-sX_i).$$

2) Plug-in estimator of $\tilde{F}^Y(s)$:

$$\tilde{F}_n^Y(s) = s^{-1}(\Psi \tilde{X}_n)(s).$$

Laplace-transform based estimation

1) Plug-in estimator of $\tilde{X}$:

$$\tilde{X}_n(s) = n^{-1} \sum_{i=1}^{n} \exp(-sX_i).$$

2) Plug-in estimator of $\tilde{F}^Y(s)$:

$$\tilde{F}_n^Y(s) = s^{-1}(\Psi \tilde{X}_n)(s).$$

3) Plug-in estimator of $F^Y(w)$:

$$F_n^Y(w) = \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{(c+\mathrm{i}y)w} \tilde{F}_n^Y(c+\mathrm{i}y)\mathrm{d}y \qquad \text{for some } c > 0.$$

Laplace-transform based estimation

1) Plug-in estimator of $\tilde{X}$:

$$\tilde{X}_n(s) = n^{-1} \sum_{i=1}^{n} \exp(-sX_i).$$

2) Plug-in estimator of $\tilde{F}^Y(s)$:

$$\tilde{F}_n^Y(s) = s^{-1}(\Psi \tilde{X}_n)(s).$$

3) Plug-in estimator of $F^Y(w)$:

$$F_n^Y(w) = \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{(c+\mathrm{i}y)w} \tilde{F}_n^Y(c+\mathrm{i}y)\mathrm{d}y \qquad \text{for some } c > 0.$$

1. Is $\tilde{X}_n$ in the domain of Ψ?
2. Is $\tilde{F}_n^Y$ a Laplace transform?

Laplace-transform based estimation

Let $E_n = \{\tilde{X}_n \in \mathrm{dom}(\Psi)\}$.

Laplace-transform based estimation

Let $E_n = \{\tilde{X}_n \in \mathrm{dom}(\Psi)\}$.

On E_n,

$$1)\; \tilde{X}_n(s) = n^{-1} \sum_{i=1}^{n} \exp(-sX_i),$$

Laplace-transform based estimation

Let $E_n = \{\tilde{X}_n \in \mathrm{dom}(\Psi)\}$.

On E_n,

$$1)\ \tilde{X}_n(s) = n^{-1} \sum_{i=1}^{n} \exp(-sX_i),$$

$$2)\ \tilde{F}_n^Y(s) = s^{-1}(\Psi \tilde{X}_n)(s),$$

Laplace-transform based estimation

Let $E_n = \{\tilde{X}_n \in \mathrm{dom}(\Psi)\}$.

On E_n,

$$1)\ \tilde{X}_n(s) = n^{-1} \sum_{i=1}^{n} \exp(-sX_i),$$

$$2)\ \tilde{F}_n^Y(s) = s^{-1}(\Psi \tilde{X}_n)(s),$$

$$3)\ F_n^Y(w) = \int_{-\sqrt{n}}^{\sqrt{n}} \frac{1}{2\pi} e^{(c+\mathrm{i}y)w} \tilde{F}_n^Y(c+\mathrm{i}y)\mathrm{d}y.$$

Laplace-transform based estimation

Let $E_n = \{\tilde{X}_n \in \mathrm{dom}(\Psi)\}$.

On E_n,

1) $\tilde{X}_n(s) = n^{-1} \sum_{i=1}^{n} \exp(-sX_i)$,

2) $\tilde{F}_n^Y(s) = s^{-1}(\Psi \tilde{X}_n)(s)$,

3) $F_n^Y(w) = \int_{-\sqrt{n}}^{\sqrt{n}} \frac{1}{2\pi} e^{(c+\mathrm{i}y)w} \tilde{F}_n^Y(c + \mathrm{i}y)\mathrm{d}y$.

On $E_n^{\mathfrak{c}}$, define $F_n^Y(w) \in [0, 1]$ arbitrarily.

Main result: conditions

Assume

(A1) $\mathbb{P}\left(E_n^{\mathsf{c}}\right) = O(n^{-1/2})$,

Main result: conditions

Assume

(A1) $\mathbb{P}(E_n^{\mathfrak{c}}) = O(n^{-1/2})$,

(A2) $F^Y(y)$ is C^1 on $[0, \infty)$, and twice differentiable at $y = w$,

Main result: conditions

Assume

(A1) $\mathbb{P}(E_n^{\mathfrak{c}}) = O(n^{-1/2})$,

(A2) $F^Y(y)$ is C^1 on $[0, \infty)$, and twice differentiable at $y = w$,

(A3) There are k_1, k_2, and nonnegative random variables $(Z_n)_{n \in \mathbb{N}}$ with $\sup_{1<p<2} \mathbb{E}[|Z_n|^p] \leq k_2 n^{-1/2}$ for all $n \in \mathbb{N}$, such that, on E_n,

$$|(\Psi \tilde{X}_n)(s) - (\Psi \tilde{X})(s)| \leq k_1 |\tilde{X}_n(s) - \tilde{X}(s)| + Z_n \text{ a.s.}$$

for all $s \in c + \mathrm{i}y$, $n \in \mathbb{N}$ and $y \in [-\sqrt{n}, \sqrt{n}]$.

Main result

Convergence rate

Suppose (A1)-(A3). Then there is a $C > 0$ such that, for all $n \in \mathbb{N}$,

$$\mathbb{E}[|F_n^Y(w) - F^Y(w)|] \leq Cn^{-1/2}\log(n+1).$$

Main result

Convergence rate

Suppose (A1)-(A3). Then there is a $C > 0$ such that, for all $n \in \mathbb{N}$,

$$\mathbb{E}[|F_n^Y(w) - F^Y(w)|] \leq Cn^{-1/2} \log(n+1).$$

Note: $n^{-1/2}$ optimal rate (take Ψ identity, apply Cramér Rao bound).

Remarks

Proof based on standard probabilistic inequalities and results on convergence rate of characteristic functions.

Remarks

Proof based on standard probabilistic inequalities and results on convergence rate of characteristic functions.

Ill-posedness of the problem (Laplace inversion not continuous operator) does not play a rôle when estimating cdf.

Outline

1. Motivation from practical problem
2. Convergence rates of Laplace-transform based estimators
3. Applications

Application 1: Queueing example

The M/G/1 queueing example, with

- $\mathcal{X} = \{X = \sum_{i=1}^{\text{Poiss}(\lambda\delta)} B_i \mid 0 < \mathbb{E}[X] < \delta, E[B_1^2] < \infty\}$;
- $E_n = \{0 \leq n^{-1} \sum_{i=1}^{n} X_i < \delta\}$;
- $(\Psi\tilde{X})(s) = \frac{s(1-\delta^{-1}\mathbb{E}[X])}{s+\delta^{-1}\mathrm{Log}(\tilde{X}(s))}$;
- Y: $\tilde{Y} = (\Psi\tilde{X})$ has F^Y cont. diff. and twice diff. in w;

Application 1: Queueing example

The M/G/1 queueing example, with

- $\mathcal{X} = \{X = \sum_{i=1}^{\text{Poiss}(\lambda\delta)} B_i \mid 0 < \mathbb{E}[X] < \delta, E[B_1^2] < \infty\}$;
- $E_n = \{0 \leq n^{-1} \sum_{i=1}^{n} X_i < \delta\}$;
- $(\Psi\tilde{X})(s) = \frac{s(1-\delta^{-1}\mathbb{E}[X])}{s+\delta^{-1}\mathrm{Log}(\tilde{X}(s))}$;
- Y: $\tilde{Y} = (\Psi\tilde{X})$ has F^Y cont. diff. and twice diff. in w;

satisfies the assumptions (A1)-(A3), and thus yields

$$\mathbb{E}[|F_n^Y(w) - F^Y(w)|] \leq Cn^{-1/2}\log(n+1)$$

for all n and some $C > 0$.

Application 2: Decompounding

Decompounding, with

- $\mathcal{X} = \{X = \sum_{i=1}^{\text{Poiss}(\lambda)} Y_i \mid 0 < \mathbb{E}[X^2] < \infty, P(Y_1 = 0) = 0\}$;

Application 2: Decompounding

Decompounding, with

- $\mathcal{X} = \{X = \sum_{i=1}^{\text{Poiss}(\lambda)} Y_i \mid 0 < \mathbb{E}[X^2] < \infty, P(Y_1 = 0) = 0\}$;
- $E_n = \{0 < n^{-1} \sum_{i=1}^{n} \mathbf{1}\{X_i = 0\} < 1\}$;
- $(\Psi \tilde{X})(s) = 1 - \mathrm{Log}(\tilde{X}(s)) / \lim_{t \to \infty} \log(\tilde{X}(t))$;
- Y: $\tilde{Y} = (\Psi \tilde{X})$ has F^Y cont. diff. and twice diff. in w;

Application 2: Decompounding

Decompounding, with

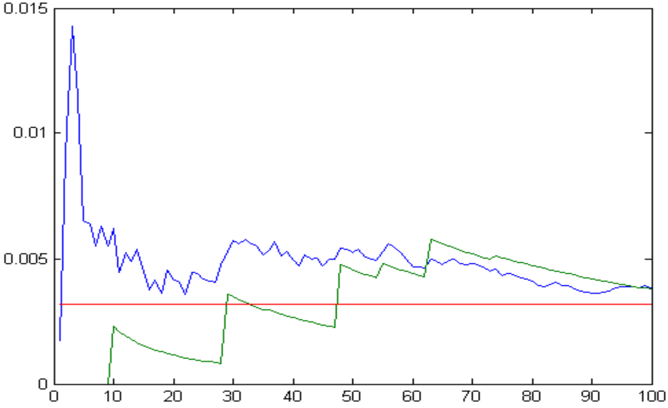
- $\mathcal{X} = \{X = \sum_{i=1}^{\text{Poiss}(\lambda)} Y_i \mid 0 < \mathbb{E}[X^2] < \infty, P(Y_1 = 0) = 0\}$;
- $E_n = \{0 < n^{-1} \sum_{i=1}^{n} \mathbf{1}\{X_i = 0\} < 1\}$;
- $(\Psi \tilde{X})(s) = 1 - \mathrm{Log}(\tilde{X}(s)) / \lim_{t \to \infty} \log(\tilde{X}(t))$;
- Y: $\tilde{Y} = (\Psi \tilde{X})$ has F^Y cont. diff. and twice diff. in w;

satisfies the assumptions (A1)-(A3), and thus yields

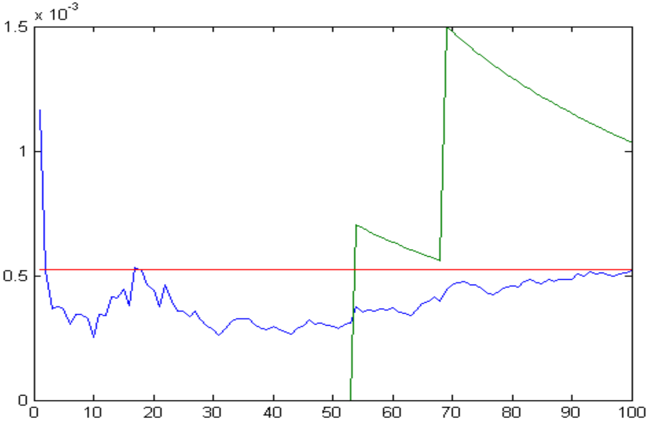
$$\mathbb{E}[|F_n^Y(w) - F^Y(w)|] \leq C n^{-1/2} \log(n+1)$$

for all n and some $C > 0$.

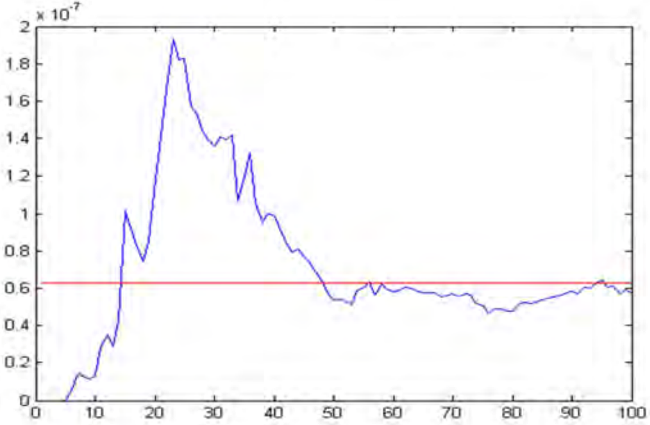
Numerical: *P(W>w)* for *w=90*



Numerical: *P(W>w)* for *w=120*



Numerical: *P(W>w)* for *w=150*



Main message:

If you estimate $P(Y \leq w)$ from a sample $X_1, \ldots, X_n$
and you know a relation between the Laplace transforms:

$$\tilde{Y} = \Psi \tilde{X}$$

then, under mild conditions and if Ψ is "almost Lipschitz continuous", the obvious plug-in estimator has near-optimal convergence rate.

Main message:
If you estimate $P(Y \leq w)$ from a sample $X_1, \ldots, X_n$
and you know a relation between the Laplace transforms:

$$\tilde{Y} = \Psi \tilde{X}$$

then, under mild conditions and if Ψ is "almost Lipschitz continuous", the obvious plug-in estimator has near-optimal convergence rate.

Nice future challenge: incorporate interval-observations.