Importance Sampling Simulation of Queues with Time-Varying Rates

Ad Ridder, Vera Rensink (*Vrije University Amsterdam*)

StochMod14, Mannheim, June 30-July 2, 2014

Introduction

- Rare event in a queueing model: large queue;
- Estimate efficiently the probability;
- By application of importance sampling algorithms;
- Analysis of the rare event probability (large deviations);
- Analysis of complexity of algorithms;

The Queueing Model: $M_t/M/1$

- Arrivals according to nonhomogeneous Poisson process;

Rate function

$$\alpha(t) = \lambda + A \sin(2\pi t/\tau), \quad t \geq 0;$$

- Exponential services with rate μ;
- Single server; infinite waiting room; FCFS;
- $X(t)$ number of customers present at time t;
- Interarrival times $U_1, U_2, \ldots$;
- Service durations $V_1, V_2, \ldots$;
- Stability: $\lambda < \mu$;

Sequence of Queues

- Parameter $n = 1, 2, \ldots, n \to \infty$;
- Sequence $\{X_n(t) : t \geq 0;\ n = 1, 2, \ldots\}$ of $M_t/M/1$ queues;
- Interarrival times:
  - $U_1^{(n)}, U_2^{(n)}, \ldots$;
  - Such that $U_1, U_2, \ldots$ defined by $U_j = nU_j^{(n)}$ are as on previous slide;
- Service times:
  - $V_1^{(n)}, V_2^{(n)}, \ldots$;
  - Such that $V_1, V_2, \ldots$ defined by $V_j = nV_j^{(n)}$ are as on previous slide;
- Interpretation: rates are n-times faster;

Rare Event

- Let $X_n(t)$ number of customers present at time t in the n-queue;
- Denote

$$X^{(n)}(t) = \frac{1}{n} X_n(t), \quad t \geq 0;$$

Rare Event Problem

Fix $\bar{x}, \bar{y}, T$; compute the overflow probability

$$\ell_n = \mathbb{P}\left(X^{(n)}(T) \geq \bar{y} \,|\, X^{(n)}(0) = \bar{x}\right)$$

for large n.

T is called the overflow horizon.

Equivalent Formulation

- In the orginal (unscaled) $M_t/M/1$ model;

$$\ell_n = \mathbb{P}\big(X(nT) \geq n\bar{y} \,|\, X(0) = n\bar{x}\big)$$

- Scaling is the usual technique for analysis;
- And for showing figures;

Some Numbers

$\lambda = 1,\ A = 0.5,\ \tau = 1,\ \mu = 1.5,\ \ \bar{x} = 0.1,\ \bar{y} = 2.0,\ T = 3.0;$

n	$\ell_n \approx$
20	1.9e-08
40	1.8e-15
60	2.3e-22
80	3.1e-29
100	4.6e-36

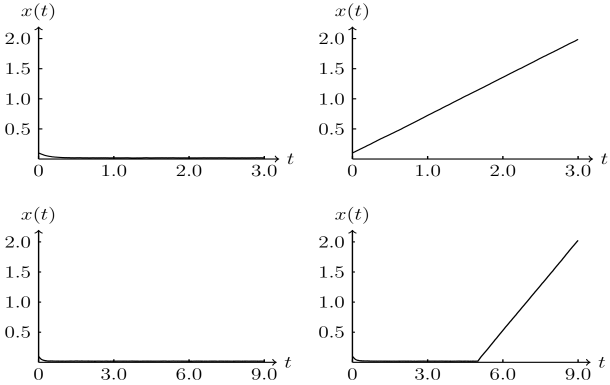
Exponential decay;

Reference Model: $M/M/1$

- Extensively studied;
- Same problem: $\mathbb{P}\big(X^{(n)}(T) \geq \bar{y} \,|\, X^{(n)}(0) = \bar{x}\big)$;
- Large deviations and most likely paths e.g. in classic work of Shwartz & Weiss 95;
- Idea is to adapt for $M_t/M/1$;

Most Likely Paths in $M/M/1$

- Limiting process $x(t) = \lim_{n\to\infty} X^{(n)}(t)$
- Plot of most likely path [left] and plot of optimal path to overflow at time T [right];
- Top: $T = 3$; Bottom $T = 9$;

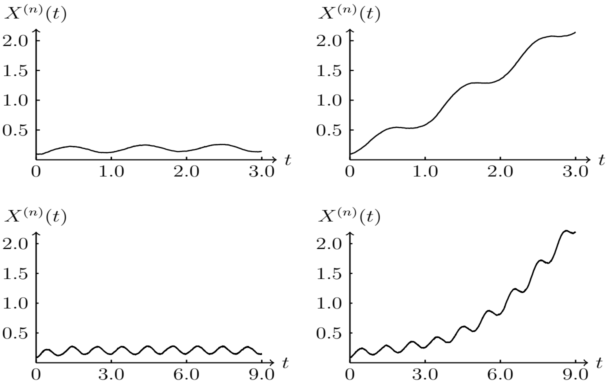


What about $M_t/M/1$?

- First, we use simulation do find empirically the typical paths to overflow;
- Then, we conjecture the optimal paths;
- And give a sketch of proof;
- Next, we apply importance sampling that follows these optimal path to overflow as its average (most likely) path;

Typical Paths in $M_t/M/1$ by Experiment

- Experiment: simulate $M_t/M/1$ queue and plot an average path and plot a typical path to overflow at time T; $n = 10$;
- Top: $T = 3$ small overflow horizon; Bottom $T = 9$ large overflow horizon;



Conjecture

Now let $n \to \infty$.

Conjecture

(i). The most likely path satisfies

$$x(t) = \Big(\bar{x} + (\lambda - \mu)\,t + \frac{A}{2\pi/\tau}\big(1 - \cos(2\pi t/\tau)\big)\Big)^+, \quad t \geq 0.$$

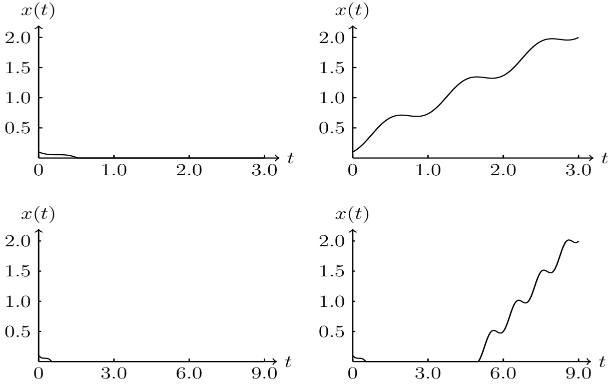
(ii). For small overflow horizon T, the optimal path to overflow satisfies

$$x(t) = \bar{x} + \frac{\bar{y} - \bar{x}}{T}\,t + \frac{Ae^{\theta}}{2\pi/\tau}\big(1 - \cos(2\pi t/\tau)\big), \quad 0 \leq t \leq T.$$

(iii). For large overflow horizon T, the optimal path to overflow is concatenation of paths of type (i) and type (ii).

The Plots

- Plot of most likely path and plot of optimal path to overflow at time T;
- Top: $T = 3$; Bottom $T = 9$;



Sketch of Proof

Most likely path.

- Let $A_n(t)$ = number of arrivals in $[0,t]$; $D_n(t)$ = number of departures;
- $A_n(t)$ is Poisson with mean $\int_0^t n\alpha(s)\,ds$;
- $A_n(t)$ can be considered to be a sum of n IID Poison RV's each with mean $\int_0^t \alpha(s)\,ds$;
- Applying SLLN we get

$$\lim_{n\to\infty} \frac{A_n(t)}{n} = \int_0^t \alpha(s)\,ds; \quad \lim_{n\to\infty} \frac{D_n(t)}{nt} = \mu \;\; (\mathbb{P} \text{ a.s.})$$

- Thus, the scaled queueing processes satisfy for large n

$$X^{(n)}(t) \approx \Big(\bar{x} + \frac{1}{n}\big(A_n(t) - D_n(t)\big)\Big)^+ \approx \Big(\bar{x} + \int_0^t \alpha(s)\,ds - \mu\, t\Big)^+$$
$$= \Big(\bar{x} + (\lambda - \mu)\,t + \frac{A}{2\pi/\tau}\big(1 - \cos(2\pi t/\tau)\big)\Big)^+, \quad t \geq 0.$$

Sketch of Proof (cont'd)

Optimal path to overflow at small overflow horizon;

- The heuristic is that the optimal path to overflow equals the most likely path under a change of measure (COM);
- The COM is the same as for the $M/M/1$: arrival and service rates

$$\alpha^*(t) = \alpha(t)e^{\theta}; \quad \mu^* = \mu e^{-\theta},$$

- With e^{θ} equals as in $M/M/1$ where it is known that

$$e^{\theta} = \frac{\bar{y} - \bar{x} + \sqrt{(\bar{y} - \bar{x})^2 + 4\lambda\mu T^2}}{2\lambda T}.$$

- Reason as previous slide that under the COM

$$\lim_{n\to\infty} \frac{A_n(t)}{n} = \int_0^t \alpha^*(s)\,ds$$
$$= e^{\theta}\left(\lambda t + \frac{A}{2\pi/\tau}\left(1 - \cos(2\pi t/\tau)\right)\right) \ (\mathbb{P}^* \text{ a.s.}),$$

Sketch of Proof (cont'd)

- Calculus:

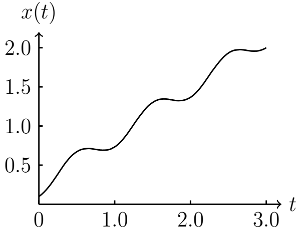
$$\lambda e^{\theta} - \mu e^{-\theta} = \frac{\bar{y} - \bar{x}}{T}.$$

- Thus for large n we get

$$\begin{aligned} X^{(n)}(t) &\approx \bar{x} + \frac{1}{n}\big(A_n(t) - D_n(t)\big) \\ &= \bar{x} + \lambda e^{\theta} t + \frac{A}{2\pi/\tau}\big(1 - \cos(2\pi t/\tau)\big)e^{\theta} - \mu e^{-\theta} t \\ &\approx \bar{x} + \frac{\bar{y} - \bar{x}}{T} t + \frac{Ae^{\theta}}{2\pi/\tau}\big(1 - \cos(2\pi t/\tau)\big), \quad 0 \le t \le T. \end{aligned}$$

Importance Sampling

- Based on the COM of the previous slide for small overflow horizon;
- Large overflow horizon: regular simulation until a time τ (same as in the $M/M/1$ queue); COM on $[\tau, T]$;
- Plot the average paths under this COM for $n = 100$;



Importance Sampling Estimator

- The (single-run) IS estimator of ℓ_n is

$$L_n = W\big(\mathbf{X}^{(n)}\big)\,\mathbb{1}\{X^{(n)}(T) \geq \bar{y}\},$$

- Where $\mathbf{X}^{(n)}$ is the path on $[0, T]$ simulated under COM;
- And where $W\big(\mathbf{X}^{(n)}\big)$ is the associated likelihood ratio;
- The final IS estimator is based om M repetitions:

$$\widehat{\ell}_n = \frac{1}{M}\sum_{k=1}^{M} L_n^{(k)}.$$

Performance of IS Estimator

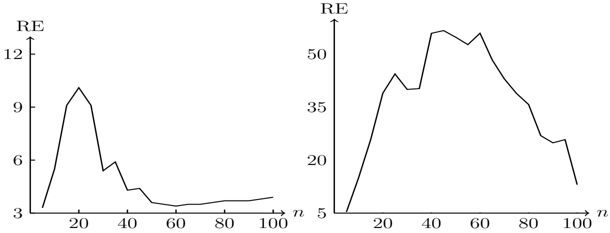
- Demand that estimator is accurate

$$\mathbb{P}(|\widehat{\ell}_n - \ell_n| < 0.1\ell_n) > 0.95.$$

- Determine the required sample size M for this;
- From Chebyshevs inequality: $M = O\big(\mathrm{RE}^2[L_n]\big)$;
- Where relative error $\mathrm{RE}[L_n] = \sqrt{\mathbb{V}ar[L_n]}/\ell_n$.
- Strong efficiency when relative error is bounded;
- Weak efficient when relative error is subexponential (e.g. polynomial);
- NB: regular Monte Carlo has exponential relative error.

Numerical Results

- Experiment: $n = 5, 10, \ldots, 100$;
- $T = 3$: sample size $M = 20000$; $T = 9$: sample size $M = 100000$;
- Repeated 100 times;
- Plot of the average estimated single-run relative errors $\widehat{\mathrm{RE}}[L_n]$;
- Left $T = 3$; right $T = 9$;



Conclusion and Outlook

- $M_t/M/1$: simple queue with time-dependent arrival rate;
- Rare event analysis and large deviations rely on $M/M/1$ results and heuristics;
    - Challenge for rigid analysis;
- Importance sampling algorithms show empirically to be efficient;
    - Large overflow horizons worse than small overflow times;
    - Needs a rigid analysis of efficiency;
- Possible extensions
    - More general models (varying service rates; more servers);
    - Other rare events (maximum in busy cycle);