# RARE EVENT SIMULATION
# IN THE
# PRESENCE OF HEAVY TAILS

Perwez Shahabuddin

Columbia University
New York, U.S.A.
(perwez@ieor.columbia.edu)

- Parts of this work were done jointly with:
  - N. K. Boots (Ph.D. student, Vrije University)
  - S. Juneja (Asst. Prof., I.I.T., Delhi, India)
- Partially supported by
  - U.S. National Science Foundation Career Award
  - Tinbergen Institute/Vrije University
  - IBM University Partnership Award

# The Classical Random Walk Problem

- $S_0 = 0$,
  $S_i = S_{i-1} + Y_i \quad \text{for} \quad i \geq 1$
  where $Y_i$'s are i.i.d

- $E(Y) < 0$, i.e., negative drift.

- Let $u$ be some level

- Problem: $\boxed{\text{Estimate } \gamma(u) = P(\max_{i \geq 0} S_i \geq u)}$
  (i.e., probability of eventually hitting $u$)

- Clearly $\gamma(u) \to 0$ as $u \to \infty$.

- Let $\tau(u)$ be hitting time of $u$
  (let $\tau(u) = \infty$ if the random walk never hits)

- Then $\gamma(u) = P(\tau(u) < \infty)$

# Application:GI/GI/1 Queue

- Let $A_i$'s be interarrival times
- Let $X_i$'s be service times.
- Let $W_i$ be waiting time of $i$th customer

- Lindley's recursion (starting from system empty):
  $W_1 = 0$,
  $W_i = [W_{i-1} + X_i - A_i]^+ = [W_{i-1} + Y_i]^+$
  where $Y_i \equiv X_i - A_i$.

- As $i \to \infty$, $W_i \Rightarrow W$ (steady state)

- Problem: $\boxed{\text{Estimate } P(W > u)}$

- Consider the non-reflected random walk:
  $S_1 = 0$,
  $S_i = S_{i-1} + Y_i$

- It turns out that
  $P(W > u) = P(\max_{i \geq 0} S_i > u) \equiv \gamma(u)$

# Application: Insurance Risk

- Let $c$ be the premium collection rate
- Let $\xi_i$'s be the interarrival times of claims
- Let $X_i$'s be the claim sizes
- Let $u$ be the initial capital

- Let $\tilde{S}_i$ be the capital at the time of $i$th claim arrival
  $\tilde{S}_0 = u$,
  $\tilde{S}_i = \tilde{S}_{i-1} + c\xi_i - X_i = \tilde{S}_i - Y_i$.
  where $Y_i \equiv X_i - c\xi_i$.

- Problem: $\boxed{\text{Estimate } P(\min_{i\geq 0} \tilde{S}_i < 0)}$
  (i.e., probability of eventual ruin)

- It is easy to see that this is equal to $P(\max_{i\geq 0} S_i > u)$
  where $S_1 = 0$, $S_i = S_{i-1} + Y_i$.

# Classical Large Deviation Results: Light Tail

- Let $f(y)$ be the probability density function (pdf) of $Y$
- Let $F(y)$ be the cumulative distribution function (cdf).
- Let $M(\theta) = E(e^{\theta Y})$ be the moment generating function (mgf)
- $M(\theta)$ is convex with $M(0) = 1$, $M'(0) = E(Y) < 0$

- Let $\theta^*$ solve
  $M(\theta) = 1 \quad$ and $\quad \theta > 0$
- Then $\gamma(u) = P(\tau(u) < \infty) = Ce^{-\theta^* u}$ where $C$ is constant. (Feller, etc.)
- Still need simulation as $C$ is unknown.

## Difficulties with Naive Simulation

- A simulation replication may never end, with positive probability, i.e., $P(\tau(u) = \infty) > 0$.

- The usual difficulty in simulating rare events.

# Importance Sampling

- Importance sampling:
  $\gamma(u) = E_p(I(\tau(u) < \infty)) = E_{\tilde{p}}(I(\tau(u) < \infty)L)$
  where
  - $p$ is the original measure
  - $\tilde{p}$ is the new measure
  - $L = p/\tilde{p}$ is the likelihood ratio

- Recall: $\tilde{p}(\cdot) = p(\cdot|\tau(u) < \infty)$ gives zero variance. Use this intuition.

- Sample path large deviations says that the most likely path to $u$ under some scaling) is:

- All $Y_i$'s contribute by becoming larger.

# Importance Sampling through Exponential Twisting (Siegmund, Asmussen, Lehtonen and Nyrhinen, etc.)

- One can achieve an approximation to this path through "exponentially twisting" each $Y$ by amount $\theta^* > 0$,i.e.,

$$\tilde{f}_{\theta^*}(y) = \frac{f(y)e^{\theta^* y}}{\int_{-\infty}^{\infty} f(s)e^{\theta^* s}ds} = \frac{f(y)e^{\theta^* y}}{M(\theta^*)} = f(y)e^{\theta^* y}$$

- $E_{\tilde{p}}(Y) = M'(\theta^*) > 0$, i.e., new random walk has a positive drift.

- Hence $\tilde{P}(\tau(u) < \infty) = 1$, i.e., simulation always ends.

- On the set $\{\tau(u) < \infty)\}$, $L$ has a nice form

$$L = \frac{f(X_1)\dots f(X_{\tau(u)})}{\tilde{f}_{\theta^*}(X_1)\dots \tilde{f}_{\theta^*}(X_{\tau(u)})} = e^{-\theta^*(X_1+\dots+X_{\tau(u)})} \leq e^{-\theta^* u}$$

# Variance Bounds

- Recall: $\gamma(u) = E_{\tilde{p}}(I(\tau(u) < \infty)L)$

- Note:

$$\begin{aligned} Var_{\tilde{p}}(I(\tau(u) < \infty)L) &= E_{\tilde{p}}(I(\tau(u) < \infty)L^2) - \gamma^2(u) \\ &\leq e^{-2\theta^* u} - C^2 e^{-2\theta^* u} \\ &= \Omega(e^{-2\theta^* u}) \end{aligned}$$

- Compare with variance of naive simulation
  $Var_p(I(\tau(u) < \infty)) = \gamma(u) - \gamma^2(u) \approx \gamma(u) = \Omega(e^{-\theta^* u})$

- Implies Bounded Relative Error.

# From Light Tails to Heavy Tails

Question: What happens when the increment random variable $Y_i$ is heavy tailed to the right (i.e., if $Y_i = Y_i^+ - Y_i^-$ and $Y_i^+$ is heavy tailed)?

Motivation

- Heavy tailed service time in GI/GI/1 queues
  Lot of interest in this these days due to telecom.

- Heavy tailed on-times of source in fluid queues
  Again, lot of interest due to telecom.

- Heavy tailed claims in insurance risk processes.

From now on, denote $Y^+$ by $X$ (non-negative and heavy-tailed).

# Rare Event Simulation in the Presence of Heavy Tails

- Very difficult.

- Only problem for which techniques have been developed so far: (Asmussen, Binswanger, Hojgaard [1997, 1998]):

  $\boxed{\text{Estimate } P(X_1 + X_2 + \ldots + X_N > u)}$

  where

  – $X_i$'s are subexponential and i.i.d.

  – $N$ is geometric and independent of $X_i$'s

- Even then, importance sampling very specific, and asymptotic optimality results only for particular cases

Question: What makes it difficult?

# Heavy Tails: Definition and Properties

Definition: $X \geq 0$ is heavy tailed (in particular, subexponential) iff for all $\theta > 0$, $E(e^{\theta X}) = \infty$

**Property 1:**

$$\lim_{u \to \infty} \frac{P(X_1 + X_2 \ldots X_n > u)}{P(\max(X_1, X_2, \ldots X_n) > u)} \to 1$$

**Property 2:**

$$\lim_{t \to \infty} P(X > t + h | X > t) = 1$$

(Compare Property 2 to exponential case where

$$P(X > t + h | X > t) = P(X > h) = e^{-h} < 1\Big)$$

## Difficulty 1:Mechanics of Importance Sampling

- The framework for importance sampling in the light tailed case, i.e., exponential twisting, cannot be used.

- Reason: The normalizing constant $M(\theta) = \infty$ for all $\theta > 0$.

## Difficulty 2: Conceptual

- Most likely paths are much more complex

- Mainly one $Y_i$ contributes.

- Which one should we change?

## Solution(Mechanics): Subexponential-twisting

- Let $\lambda(x) = \frac{f(x)}{\overline{F}(x)}$ be the hazard rate of $X$.

- Let $\Lambda(x) = \int_0^x \lambda(s)ds$ be the hazard function. It is increasing, and concave for $X$ heavy tailed.

- Example: For heavy tailed Weibull, $\Lambda(x) = x^\alpha$ where $0 < \alpha < 1$.

- Known Facts:
  - The $\overline{F}(x) = e^{-\Lambda(x)}$ (sub-exponential decrease of tail)
  - The $f(x) = \lambda(x)e^{-\Lambda(x)}$

- The "subexponentially twisted density by amount $\theta$" is given by:

$$\tilde{f}_\theta(x) = \frac{f(x)e^{\theta\Lambda(x)}}{\int_0^\infty f(s)e^{\theta\Lambda(s)}ds}$$

  for $0 < \theta < 1$.

# Crude Motivations for Subexponential Twisting

- Normalization constant is finite and easily computable.
- Infact $\int_0^\infty f(s)e^{\theta\Lambda(s)}ds = \frac{1}{1-\theta}$
- Also, subexponential random variables have a momentum of their own (Property 2)
- So do <u>not</u> need to twist at the high rate of $e^{\theta x}$ as in "exponential twisting"

# Delayed Subexponential Twisting

$$\tilde{f}_{\theta,x^*}(x) = \begin{array}{ll} f(x) & \text{for} \quad 0 \leq x \leq x^* \\ C * \tilde{f}_\theta(x) & \text{for} \quad x \geq x^*. \end{array}$$

($C$ is a normalizing constant)

Motivations:

- Twist only $X_i$'s that are promising, i.e., higher than $x^*$.

- Process behaves normally until a big jump.

Theorem:[Juneja and Shahabuddin (1999)] Delayed subexponential twisting with

- $\theta \equiv \theta_u = 1 - \frac{1}{\Lambda(u)}$
- $x \equiv x_u^*$ such that $\Lambda(x_u^*) = \log(\Lambda(u))$

is *asymptotically optimal* (a.o.) for the problem of estimating $P(X_1 + \ldots + X_N > u)$ where $X_i$'s are i.i.d and $N$ is geometric and independent of $X_i$'s.

(Asmussen, Binswanger, Hojgaard (2000) have another a.o. importance sampling approach for the same problem)

Question: What about the random walk problem?

## Solution (for conceptual difficulty)

Two options:

1. Find and implement a state and/or time dependent change of measure (Too difficult!)

2. Compromise slightly on the efficiency criteria usually used, i.e., asymptotic optimality

We do No. 2.

## Compromise: Live with some controllable bias in the estimate

- Instead of estimating $P(\tau(u) < \infty)$, we estimate $P(\tau(u) \leq k_0)$

- Choose $k_0 \equiv k_0(u)$ high so that the bias (i.e., $P(k_0(u) < \tau(u) < \infty)$) is "low".

- In particular, using some large deviations theory developed by Asmussen and Kluppelberg(1996), we are able to select $k_0(u)$ such that the bias is of the same order as the *desired* CI half width. Hence in practical terms the bias does not matter.

- Developed techniques to estimate $P(\tau(u) \leq k_0(u))$ in an *asymptotically optimal* fashion. (We use a version of delayed sub-exponential twisting).

- We call such an estimator "large-set asymptotically optimal".

(All this done in Boots and Shahabuddin (2000))

## Selection of $k_0(u)$

- The relative bias is given by

$$\frac{P(k_0(u) \leq \tau(u) < \infty)}{P(\tau(u) < \infty)}$$

- Let $\delta$ be the *desired* relative bias. Example: 1%, 5% or 10%. (usually chosen of the same order as the desired relative error, $RE \equiv HW/\gamma$)

- Choose $k_0(u)$ such that

$$\lim_{u \to \infty} \frac{P(k_0(u) < \tau(u) < \infty)}{P(\tau(u) < \infty)} \leq \delta$$

- How? Use some asymptotic theory developed about the distribution of $\tau(u)$ by Asmussen and Kluppelberg(1996)

- We then get that

$$k_0(u) = -\frac{a(u)\log(\delta)}{C}$$

where $a(u)$ is some known (concave) function of $u$, and $C$ is a constant.

# Experiments

| u | | $\rho = 0.25$ | $\rho = 0.5$ | $\rho = 0.75$ |
|---|---|---|---|---|
| 100 | A1 | $2.31E-4 \pm 1.7\%$ | $1.38E-3 \pm 3.4\%$ | $1.68E-2 \pm 10.8\%$ |
| | | (ER = 2.4E2) | ( ER = 3.8) | ( ER = 0.07) |
| | AA | $1.17E-4(49.1\%)$ | $5.00E-4(64.5\%)$ | $1.50E3(92.1\%)$ |
| 200 | A1 | $4.71E-6 \pm 2.0\%$ | $2.46E-5 \pm 3.5\%$ | $6.41E-4 \pm 38.8\%$ |
| | | (ER = 5.9E3) | (ER = 1.5E2) | (ER = 0.46) |
| | AA | $3.64E-6(21.0\%)$ | $1.09E-5(57.3\%)$ | $3.28E-5(95.5\%)$ |
| 400 | A1 | $1.65E-8 \pm 2.5\%$ | $7.12E-8 \pm 3.1\%$ | $1.53E-6 \pm 69.5\%$ |
| | | (ER = 9.2E5) | ( ER = 1.3E5) | ( ER = 13.7) |
| | AA | $1.44E-8(13.3\%)$ | $4.33E-8(39.1\%)$ | $1.30E-7(92.0\%)$ |
| 800 | A1 | $5.54E-12 \pm 3.0\%$ | $2.04E-11 \pm 3.1\%$, | $1.27E-10 \pm 9.4\%$ |
| | | (ER = 1.8E9) | (ER = 4.5E8) | (ER = 1.5E6) |
| | AA | $5.08E-12(6.8\%)$ | $1.52E-12(25.5\%)$ | $4.57E-11(66.4\%)$ |

Estimates of $P(W > u)$ for the M/GI/1 Queue with Weibull$(1, 1/2)$ Service Times.

Glossary:

- A1: the algorithm developed in this work
- ER: efficiency ratio of A1 to naive simulation.
- AA: is a well known asymptotic approximation

$$P(W > u) \sim \frac{\rho}{(1-\rho)} \overline{F}_I(u)$$

where $F_I(x)$ is the integrated tail distribution of the service time and and $\rho$ is the traffic intensity

# Experiments(Contd.)

| Method | Result |
|---|---|
| A1 | $1.12E-4 \pm 2.5\%$ (ER $= 80.7$) |
| AA | $6.36E-5 (42.2\%)$ |

Estimates of $P(W > u)$ for the $D/GI/1$ Queue with Weibull$(1, 1/2)$ service times. Here $u = 50$ and $\rho = 0.5$.

# Summary

1. Introduced sub-exponenial twisting

2. Proved it to be asymptotically optimal for estimating tails of geometric sum of heavy tailed r.v.'s

3. Introduced the criterion: large set asymptotic optimality

4. Developed a simulation algorithm for estimating small hitting probabilities of heavy tailed random walks.

5. Empirically works for "Weibull type" heavy tails and "lognormal type" heavy tails (not for regularly varying tails, e.g., Pareto)

6. Proved large-set asymptotic optimality for Weibull type tails

# Open Problems

- Techniques and proofs for regularly-varying tails (e.g., Pareto)

- Generalizations to tandem queues and networks