Sampling Rare Events of Random Walks with Regularly Varying Increments: A Dichotomy

Guido Lagos

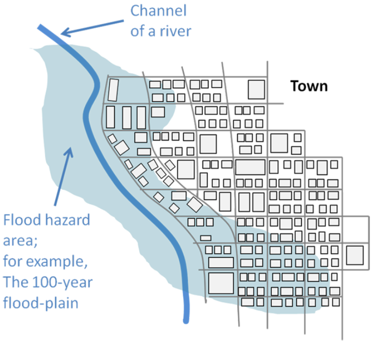
Joint work with Ton Dieker

H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

INFORMS 2014; San Francisco, CA

Motivation

- Random walks: fundamental mathematical models
- Maximum of a random walk:



- Subproblem: *level crossing* problem ⟵ this talk

Problem

Setting:

- Random walk $\left(S_n := \sum_{i=1}^{n} X_i\right)_{n\geq 0}$, $\quad X_i \sim$ iid
- Assume *negative drift*: $\mathbb{E}X_i < 0$. $\quad$ Hence $S_n \to -\infty$ a.s.
- Level $b > 0$, hitting time $\tau_b := \inf\{n \ : \ S_n > b\}$

ESTIMATION problem (70's+)

Estimate

$$\mathbb{P}(\tau_b < \infty)$$

i.e. probability that b is crossed

SAMPLING problem (2011+)

Conditional on $\tau_b < \infty$, sample

$$S_1, S_2, \ldots, S_{\tau_b}$$

i.e. paths crossing level b

Our problem / This talk

Sampling problem, when increments are of the form

$$\mathbb{P}(X > t) \approx \text{const}/t^{\alpha} \quad \text{as } t \to \infty \qquad \text{(Regularly Varying)}$$

Change of measure technique I

- Standard approach: change of measure technique

> "sample a path from another distribution $\mathbb{Q}$;
> decide output using likelihood ratio $\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}$"

- Example:

$$\mathbb{P}: \; X_i \sim N(-\mu, 1) \quad \rightsquigarrow \quad \text{instead use } \mathbb{Q}: \; X_i \sim N(+\mu, 1)$$

For $X_i \sim N(+\mu, 1)$, i.e. from $\mathbb{Q}$,

$$\begin{aligned} \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}(S_n : 0 \le n \le \tau_b) &= \frac{\phi(X_1+\mu)}{\phi(X_1-\mu)} \cdot \frac{\phi(X_2+\mu)}{\phi(X_2-\mu)} \cdots \frac{\phi(X_{\tau_b}+\mu)}{\phi(X_{\tau_b}-\mu)} \\ &= e^{-2\mu X_1} \cdot e^{-2\mu X_2} \cdots e^{-2\mu X_{\tau_b}} \\ &= e^{-2\mu S_{\tau_b}} \end{aligned}$$

General scheme

one-step transition kernel	$\mathrm{P}(s, s+\mathrm{d}x)$	$\rightsquigarrow$	changed one-step kernel	$\mathrm{Q}(s, s+\mathrm{d}x)$	$\rightsquigarrow$	
	one-step likelihood ratio	$\frac{\mathrm{dP}}{\mathrm{dQ}}(s, s+x)$	$\rightsquigarrow$	path likelihood ratio	$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}(S_n : 0 \le n \le \tau_b)$	

Change of measure technique II

Efficiency: does a $\mathbb{Q}$ makes sampling easier? What's "easier"?

Lemma (Linear running time)

If

$$\liminf_{b-s\to\infty} \quad \mathbb{E}_s^{\mathbb{Q}} X - \int_{b-s}^{\infty} \mathbb{Q}_s(X > u)\mathrm{d}u > 0$$

then $\mathbb{E}^{\mathbb{Q}}\tau_b = O(b)$*. In particular,* $\mathbb{Q}(\tau_b < \infty) = 1$*.*

Lemma (Efficient path sampling)

The following are equivalent:

- *For all* $B \subseteq \{\tau_b < \infty\}$, $\quad \mathbb{Q}\left((S_n)_n \in B\right) \geq \mathbb{P}\left((S_n)_n \in B\right)$
- $\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}(S_n : 0 \leq n \leq \tau_b)\mathbf{1}_{\{\tau_b < \infty\}} \leq 1 \quad \mathbb{Q}$*-a.s.*
- *Acceptance-Rejection algorithm does the following:*
  - *sample a Bernoulli with parameter* $\mathbb{P}(\tau_b < \infty)$*, without knowing value of* $\mathbb{P}(\tau_b < \infty)$
  - *if Bernoulli* $= 1$*, provide a sample* $(S_n)_{n=0}^{\tau_b} \sim \mathbb{P}(\,\cdot\,|\tau_b < \infty)$*.*

Our setting I: regularly varying distributions

Definition (Light- and heavy-tailed distributions)

- X is (right) light-tailed iff as $t \to \infty$

  "$\mathbb{P}(X > t)$ decays faster than an exponential"

- X is (right) heavy-tailed iff as $t \to \infty$

  "$\mathbb{P}(X > t)$ decays slower than any exponential"

Definition (Regularly Varying distributions)

Heuristically,

$$\mathbb{P}(X > t) \sim \frac{\text{const.}}{t^\alpha} \quad \text{as} \quad t \to \infty$$

and similar distributions.

- Regularly varying $\subset$ heavy-tailed
- We want $\mathbb{E}X_i$ finite $\Rightarrow$ necessarily $\alpha > 1$
- Problems with heavy-tails: "heavier tail $\Rightarrow$ harder problem"

Our setting II: $\mathbb{Q}$ of Blanchet Glynn 2008

- Motivation: transition kernel $\mathrm{P}^*(y, \mathrm{d}z)$ of $\mathbb{P}(\cdot | \tau_b < \infty)$ is

$$\mathrm{P}^*(y, \mathrm{d}z) = \mathrm{P}(y, \mathrm{d}z) \cdot \frac{u^*(z)}{u^*(y)} \text{ for } u^*(z) := \underbrace{\mathbb{P}_z(\tau_b < \infty)}_{\text{don't know this guy!}}$$

- Blanchet Glynn 2008: use

$$\mathrm{Q}(y, \mathrm{d}z) = \mathrm{P}(y, \mathrm{d}z) \cdot \frac{v(z)}{w(y)} \text{ for } v(z) \sim u^*(z) \text{ as } b - z \to \infty$$

where $w(y) =$ normalizing constant.

- Pakes-Veraberbeke Theorem (70's): for X_i's regularly varying,

$$\underbrace{\mathbb{P}_0(\tau_b < \infty)}_{u^*} \sim \underbrace{\frac{1}{|\mathbb{E}X|} \int_b^\infty \mathbb{P}(X > s)\mathrm{d}s}_{\text{approx. of } u^* \Rightarrow \text{ use as } v\,!} \quad b \to \infty.$$

Our result

Theorem

Consider

- $X_i \sim$ *regularly varying right-tails, with* $\alpha > 1$ *and* $\mathbb{E}|X_i| < \infty$
- $\mathbb{Q}$ *= change of measure of Blanchet Glynn 2008*
- $b > 0$ *arbitrary*

1. *The following are equivalent:*
   - $\mathbb{Q}((S_n)_n \in B) \geq \mathbb{P}((S_n)_n \in B)$ *for all* $B \subseteq \{\tau_b < \infty\}$
   - $\alpha \in (1, \frac{3}{2})$ $\leftarrow$ *i.e. only for "heaviest tailed" case!*
2. *Although* $\mathbb{Q}(\tau_b < \infty) = 1$,

   $\mathbb{E}^{\mathbb{Q}}\tau_b = \infty$ *when* $\alpha \in (1, \frac{3}{2})$ $\leftarrow$ *so it's impractical* ☹

But anyway, what's special about $\alpha_0 = 3/2$?

A recurrent threshold for α

Summary of requirements

- *"One can estimate probability* $\mathbb{P}(\tau_b < \infty)$ *efficiently"* if $\mathbb{E}^{\mathbb{Q}}\left(\frac{d\mathbb{P}}{d\mathbb{Q}}(S_n : 0 \le n \le \tau_b)^2 \mathbf{1}_{\{\tau_b<\infty\}}\right) \lesssim \mathbb{P}(\tau_b < \infty)^2 \cdot O(1)$
- *"One can sample path* $S_1, \ldots, S_{\tau_b} \sim \mathbb{P}(\cdot|\tau_b < \infty)$ *efficiently"* if $\frac{d\mathbb{P}}{d\mathbb{Q}}(S_n : 0 \le n \le \tau_b)\mathbf{1}_{\{\tau_b<\infty\}} \le 1 \quad \mathbb{Q}-a.s.$
- *"Linear running time"* if $\mathbb{E}^{\mathbb{Q}}\tau_b = O(b)$

Other changes of measure showing threshold $\alpha_0 = 3/2$:

Measure	Problem	Result	By
$\mathbb{Q}$ from Blanchet Glynn (2008)	Sampling	Efficiency $\Leftrightarrow$ $\alpha \in (1, \alpha_0)$. But ∞ run time.	Us
$\mathbb{Q}$ from Blanchet Liu (2012)	Estimation	Efficiency + linear run time $\Leftrightarrow$ $\alpha \in (\alpha_0, \infty)$.	[BL12]
$\mathbb{Q}$ from Murthy Juneja Blanchet (2013)	Estimation	Efficiency + linear run time $\Leftrightarrow$ $\alpha \in (\alpha_0, \infty)$.	[MJB13]

Summary & conclusions

- Problem: efficient path sampling of

$$(S_1, \ldots, S_{\tau_b}) \sim \mathbb{P}(\cdot|\tau_b < \infty)$$

  when increments $\sim$ regularly varying
- General framework for sampling efficiently rare events
- Showed that with Blanchet Glynn (2008) change of measure and with regularly varying tails:
  - Sampling paths efficiently $\Leftrightarrow \alpha \in (1,\ 3/2)$
  - i.e. only "works" for heaviest tails!
  - However $\mathbb{E}^{\mathbb{Q}}\tau_b = \infty$ for $\alpha \in (1,\ 3/2)$
- Threshold $\alpha_0 = 3/2$ also appears in other simulation problems using change of measure
- Surprising: these problems have different requirements, but same α_0 arises!

Thanks!