Rare Event Simulation with Heavy Tails

.

Søren Asmussen[1]

Aarhus University, Denmark

Cherry Bud Workshop

Keio University, Yokohama

February 26, 2005

[1]URL: http://home.imf.au.dk/asmus

Problem; Key Examples

Efficient simulation of $\mathbb{P}(A)$ where $\mathbb{P}(A)$ is small $(10^{-2}, 10^{-3}, \ldots)$

VaR need $\mathbb{P}(A) = \mathbb{P}(L > x)$

$L \;=\; X_1 + \cdots + X_N$

N constant: # of assetts

N random: # of defaults in credit risk Ruin probability in insurance

Buffer overflow before T or in cycle

Large steady-state waiting times

Crude Monte Carlo inefficient

simulation generates $I(A)$

replicate R times, estimate $\mathbb{P}(A)$ by average

Variance $\dfrac{1}{R}\mathbb{P}(A)(1 - \mathbb{P}(A))$

If R moderate:

stdev of order $\sqrt{\mathbb{P}(A)} >> \mathbb{P}(A)$

Estimate tells only that $\mathbb{P}(A)$ is close to 0

but does not give order

More sophistication needed to keep R feasible

This lecture: Heavy-tailed risk factors

Set-Up

$A = A(x)$, x large parameter
(x order of 1% quantile, 0.1% quantile; buffer size; initial capital;...)
$\mathbb{P}(A(x)) \to 0$, $x \to \infty$
$Z(x)$ estimators:
$\mathbb{E}Z(x)) = \mathbb{P}(A(x))$, simulatable
e.g. CMC: $I(A(x))$

Relative error $\epsilon(x) \;=\; \dfrac{\mathrm{stdev}(Z(x))}{\mathbb{P}(A(x))}$

$$\text{CMC: } \epsilon(x) = \frac{\sqrt{\mathbb{P}(A(x))\big(1-\mathbb{P}(A(x))\big)}}{\mathbb{P}(A(x))} \sim \frac{1}{\sqrt{\mathbb{P}(A(x))}}$$

Bounded relative error: $\epsilon(x) = O(1)$

Logaritmic efficiency: $\dfrac{\mathrm{stdev}(Z(x))}{\mathbb{P}(A(x))^{1-\epsilon}} = O(1)$
for all $\epsilon > 0$ (almost as good)

Key question: how find such $Z(x)$?

Variance reduction methods:

- Antithetic variables
- Importance sampling
- Control variates
- Stratification
- Conditional MC
- Multilevel splitting

Importance sampling most established tool

Importance sampling

(change of measure)

$\mathbb{P}^* = \mathbb{P}^*(x)$ different prob measure,

$L = L(x) = \dfrac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{P}^*}$ likelihood ratio (R-N derivative)

Simulate from $\mathbb{P}^*$, estimator $Z(x) = L(x)I(A(x))$

In practice: change some density

e.g. log-return distribution: $N(\mu, \sigma^2) \longrightarrow N(\mu^*, \sigma^2)$

$$L \ = \ \prod_{i=1}^{\nu} \frac{f(X_i)}{f^*(X_i)};$$

. $\nu \ = \#\ X_i$ in simulation (stopping time)

In more complicated examples, change e.g.:

. diffusion coefficients (Girsanov)

. Markov chain transition probabilities

0-Variance IS: $\mathbb{P}^* = \mathbb{P}(\cdot \mid A(x))$

Not feasible since $L = \mathbb{P}(A(x))$

Key idea: take $\mathbb{P}^*$ close to $\mathbb{P}(\cdot \mid A(x))$

$\mathbb{P}^* \approx \mathbb{P}(\cdot \mid A(x))$: light–tailed examples

Ex 1 (n instead of x) $A(n) = \{S_n > n\kappa\}$
$S_n = X_1 + \cdots + X_n$ RW, $\kappa > \mathbb{E}(X)$
light tails: exponential moments exists
exponential family $f_\theta(y) \;=\; \dfrac{e^{\theta y}}{\mathbb{E}e^{\theta X}} f(y)$
Boltzmann's law:
density of $X_i \mid S_n > n\kappa \quad \to \quad f_{\theta^*}(y)$,
θ^* solves $\mathbb{E}_{\theta^*} X = \kappa$ (saddlepoint)
This exponential twisting is log eff

Ex 2
$M = \max S_n$ in RW with $\mathbb{E}(X) < 0$,
$A(x) = \{M > x\}$
Exponential twisting, θ^* solution of $\mathbb{E}e^{\theta^* X} = 1$
Bd relative error

Heavy tails?
Not much progress beyond $A(x) = \{S_n > x\}$

Heavy-tailed distributions

Often in finance: heavier than normal
Here: no exponential moments
Subexponential class (distrn's on $(0, \infty)$):

$$\frac{\overline{F}^{*n}(x)}{\overline{F}(x)} \longrightarrow n$$

Main examples:

- **Regular variation:** $\overline{F}(x) \sim \dfrac{L(x)}{x^\alpha}$
- **lognormal** $\mathrm{e}^{\mu + \sigma U}$
- **Weibull** $\overline{F}(x) \sim \mathrm{e}^{-x^\beta}$, $\beta < 1$

Intuition: S_n **large iff (asymptotically!) one** X_i **is so**
$S_n > x$ **iff (asymptotically!)** $\max(X_1, \ldots, X_n) > x$
$\{\max(X_1, \ldots, X_n) > x\} \ \subset \ \{S_n > x\}$
$P(\max(X_1, \ldots, X_n) > x) \ \sim \ n\overline{F}(x)$

$\mathbb{P}^* \approx \mathbb{P}(\cdot \mid A(x))$: first counterex for heavy tails

Key setting $A(x) = \{S_n > x\}$

$S_n > x$ iff (asymptotically!) $\max(X_1, \ldots, X_n) > x$

$\mathbb{P}(\cdot \mid A(x)) \approx$ one $X_i > x$, rest "normal"

IS: $\mathbb{P}^*$ forces one $X_i > x$, rest "normal"

Fails because of lack of absolute continuity

$A(x) \cap \{\text{all } X_i < x\}$ possible under $\mathbb{P}$, not $\mathbb{P}^*$

First efficient algorithm for heavy tails

(SA-Binswanger 1997); not IS!!

Key setting $A(x) = \{S_n > x\}$

Conditional MC: $Z(x) = P(A(x) \mid \mathcal{F})$

Always variance reduction:

. $\mathrm{Var}(Z) \;=\; E[\mathrm{Var}(Z \mid \mathcal{F})] + \mathrm{Var}[E(Z \mid \mathcal{F})]$

Good choice of $\mathcal{F}$?

First idea $\sigma(X_1, \ldots, X_{n-1})$, $Z(x) = \overline{F}(x - S_{n-1})$

Contribution to variance

from $\{\max(X_1, \ldots, X_{n-1}) > x\}$ too large

Get rid of big value by conditioning

Order statistics $X_{(1)} < \ldots < X_{(n)}$, $\mathcal{F} = \sigma\big(X_{(1)}, \ldots, X_{(n-1)}\big)$

$$Z(x) \;=\; \frac{\overline{F}\big((x - S_{(n-1)}) \vee X_{(n-1)}\big)}{\overline{F}(X_{(n-1)})}$$

$S_{(n-1)} \;=\; X_{(1)} + \cdots + X_{(n-1)}$

Log efficient!!! (RV case)

IS algorithms

Key setting $A(x) = \{S_n > x\}$
IS: $F \quad \longrightarrow \quad F^*$
SA-Binswanger-Højgaard 2000: $\overline{F}^*(y) \sim \dfrac{1}{\log y}$
as HT as can be!
Log efficient!!!
But works not well in practice
(for a given finite x)

Juneja-Shahabuddin 2002: hazard rate twisting key ingredient
hazard rate $\lambda(y) = \dfrac{f(y)}{\overline{F}(y)}$ (HT: $\longrightarrow 0$)

$$\overline{F}(y) \;=\; \exp\Bigl\{-\int_0^y \lambda(z)\,\mathrm{d}z\Bigr\}$$
$$\lambda(y) \;\longrightarrow\; \lambda^*(y) = \theta(x)\lambda(y)$$
$$\overline{F}(y) \;\longrightarrow\; \overline{F}(y)^{\theta(x)}, \quad \theta(x) \longrightarrow 0$$

Log efficient if $\theta(x) = \dfrac{b}{\log \Lambda(x)}$ (ABH: $b = 0$)

. Pareto: $\alpha \;\longrightarrow\; \alpha^* = b' / \log x$
. Lognormal: change of σ^2
. Weibull: change of scale

Cross-Entropy Approach to F^*

SA-Kroese-Rubinstein *Stochastic Models* 2005

http://www.maphysto.dk/publications

Back to $\mathbb{P}^* \approx \mathbb{P}(\cdot \mid A(x))$

Entropy measure of closeness (Kullback-Leibler)

Def'n $\mathcal{E}(\mathbb{P}, \mathbb{P}^*) = \mathbb{E} \log \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{P}^*} = \int \log f \mathrm{d}\mathbb{P} - \int \log f^* \mathrm{d}\mathbb{P}$

Minimal entropy = maximum likelihood

. $\mathbb{P}$: empirical distr. (data)

. $f^* = f_{\theta^*}$ fitted density

Here $\mathbb{P} = P(\cdot \mid A(x))$, $\mathbb{P}^* =$ IS measure

Minimize $\mathcal{E}\big(P(\cdot \mid A(x)), \mathbb{P}^*\big)$

Use subexponential asymptotics for $P(\cdot \mid A(x))$

(different from K-R adaptive implementation)

Example 1: Pareto, $f(x) = \alpha/(1+x)^{\alpha+1}$
Estimate $\mathbb{P}(X_1 + \cdots + X_n > x)$

Choose f^* of same parametric form; $\alpha^* =?$
MLE

$$\begin{aligned}\widehat{\alpha} &= \frac{n}{\log(1+y_1) + \cdots + \log(1+y_n)} \\ &= \Big(\int \log(1+y) F_n(\mathrm{d}y)\Big)^{-1}\end{aligned}$$

$F_n =$ empirical distribution
Here $\mathbb{P}(\cdot \mid A(x))$ takes role of F_n; asymptotically, one $X_i \sim X | X > x$, others unchanged
Replace for F_n in integral above:

$$\begin{aligned}&\frac{1}{n}\int_x^\infty \log(1+y)\frac{f(y)}{\overline{F}(x)}\,\mathrm{d}y \\ &\quad + \frac{n-1}{n}\int_0^\infty \log(1+y) f(y)\,\mathrm{d}y \\ = \ &\frac{1}{n}\big(1/\alpha + \log(1+x)\big) + \frac{n-1}{n}\alpha \ \sim \ \frac{\log x}{n}, \quad x \to \infty\end{aligned}$$

$$\alpha^* \ \sim \ \frac{n}{\log x}$$

More precise than
Juneja-Shahabuddin 02 ($\alpha^* \sim b/\log x$)
and SA-Binswanger-Højgaard 00 ($\alpha^* = 0$).

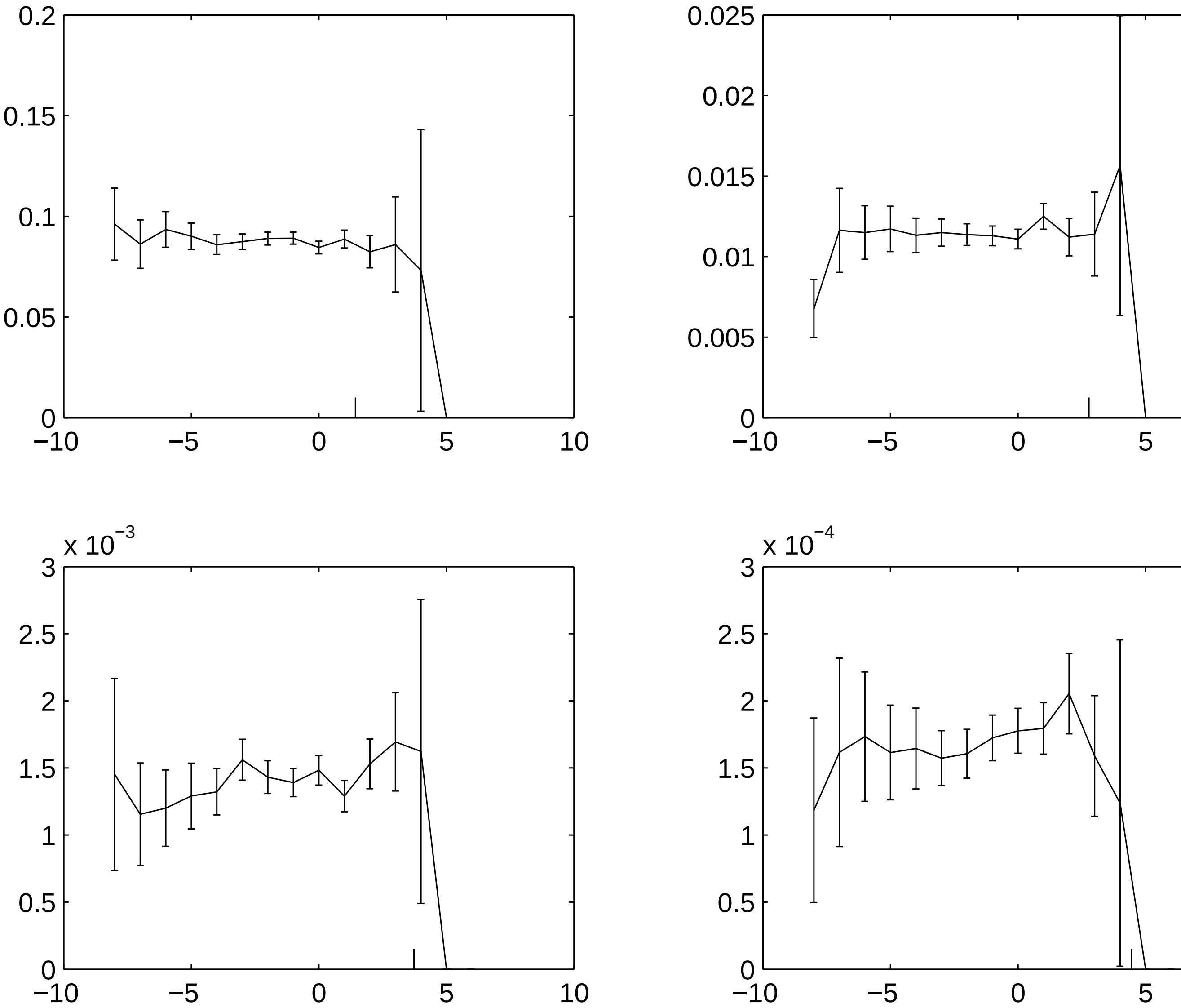

0.2
0.15
0.1
0.05
0
−10
−5
0
5
10
0.025
0.02
0.015
0.01
0.005
0
−10
−5
0
5
x 10^−3
3
2.5
2
1.5
1
0.5
0
−10
−5
0
5
10
x 10^−4
3
2.5
2
1.5
1
0.5
0
−10
−5
0
5

Figure 1:

Example 2: Weibull, $\overline{F}(x) = e^{-x^\beta}$, $\beta < 1$

$f(x) = \beta x^{\beta-1} e^{-x^\beta}$

Estimate $\mathbb{P}(X_1 + \cdots + X_n > x)$

Choose f^* of same parametric form; $\beta^* = ?$

MLE

$$\widehat{\beta} = \frac{n}{y_1^\beta + \cdots + y_n^\beta} = \Big(\int y^\beta F_n(\mathrm{d}y) \Big)^{-1}$$

F_n = empirical distribution

Here $\mathbb{P}(\cdot \mid A(x))$ takes role of F_n; asymptotically, one $X_i \sim X | X > x$, others unchanged

Replace for F_n above:

$$\int_x^\infty y^\beta \frac{1}{n} \frac{f(y)}{\overline{F}(x)} \, \mathrm{d}y + \int_0^\infty y^\beta \frac{n-1}{n} f(y) \, \mathrm{d}y$$

$$= \frac{1}{n}\big(x^\beta + 1\big) + \frac{n-1}{n} \sim \frac{x^\beta}{n}, \quad x \to \infty$$

$$\boxed{\beta^* \sim \frac{n}{x^\beta}}$$

More precise than

Juneja-Shahabuddin 00 ($\beta^* \sim b/x^\beta$)

and SA-Binswanger-Højgaard 00 ($\beta^* = 0$).

Example 3: Pareto, $f(x) = \alpha / x^{\alpha+1}$
Estimate $\mathbb{P}(X_1 + \cdots + X_n > x)$

Twist scale: choose f^* of form $\dfrac{\alpha/\gamma}{(1+x/\gamma)^{\alpha+1}}$
$\gamma^* =?$
MLE solves

$$\frac{1}{1+\alpha} = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i/\widehat{\gamma}}{1+y_i/\widehat{\gamma}} = \int \frac{y/\widehat{\gamma}}{1+y/\widehat{\gamma}} F_n(\mathrm{d}y)$$

Not explicit, but easy to see one unique solution
Further analysis two cases $n < 1+\alpha$, $n > 1+\alpha$;
First as before with variants, second surprise
Efficient simulation using twist of γ only possible in first
CE good guideline in both

New Algoritms

SA-Kroese 2004

Previous IS algorithms: why iid?
Particular role of one big X_i
New idea:

$$\mathbb{P}(S_n > x) = n\mathbb{P}(S_n > x, M_n = X_n)$$

$M_n = \max(X_1, \ldots, X_n)$

1. IS: change distr'n of X_n only

$$Z(x) \;=\; n\frac{f(X_n)}{f^*(X_n)}I(S_n > x, M_n = X_n)$$

2. Conditional MC:

$$\begin{aligned} Z(x) &= n\,\mathbb{P}\big(S_n > x, M_n = X_n \mid X_1, \ldots, X_{n-1}\big) \\ &= n\,\overline{F}\big(M_{n-1} \vee (x - S_{n-1})\big). \end{aligned}$$

Asymptotic properties:
RV: IS log eff, CondMC bd rel error
Weibull: Cond MC log eff (believe bd re)

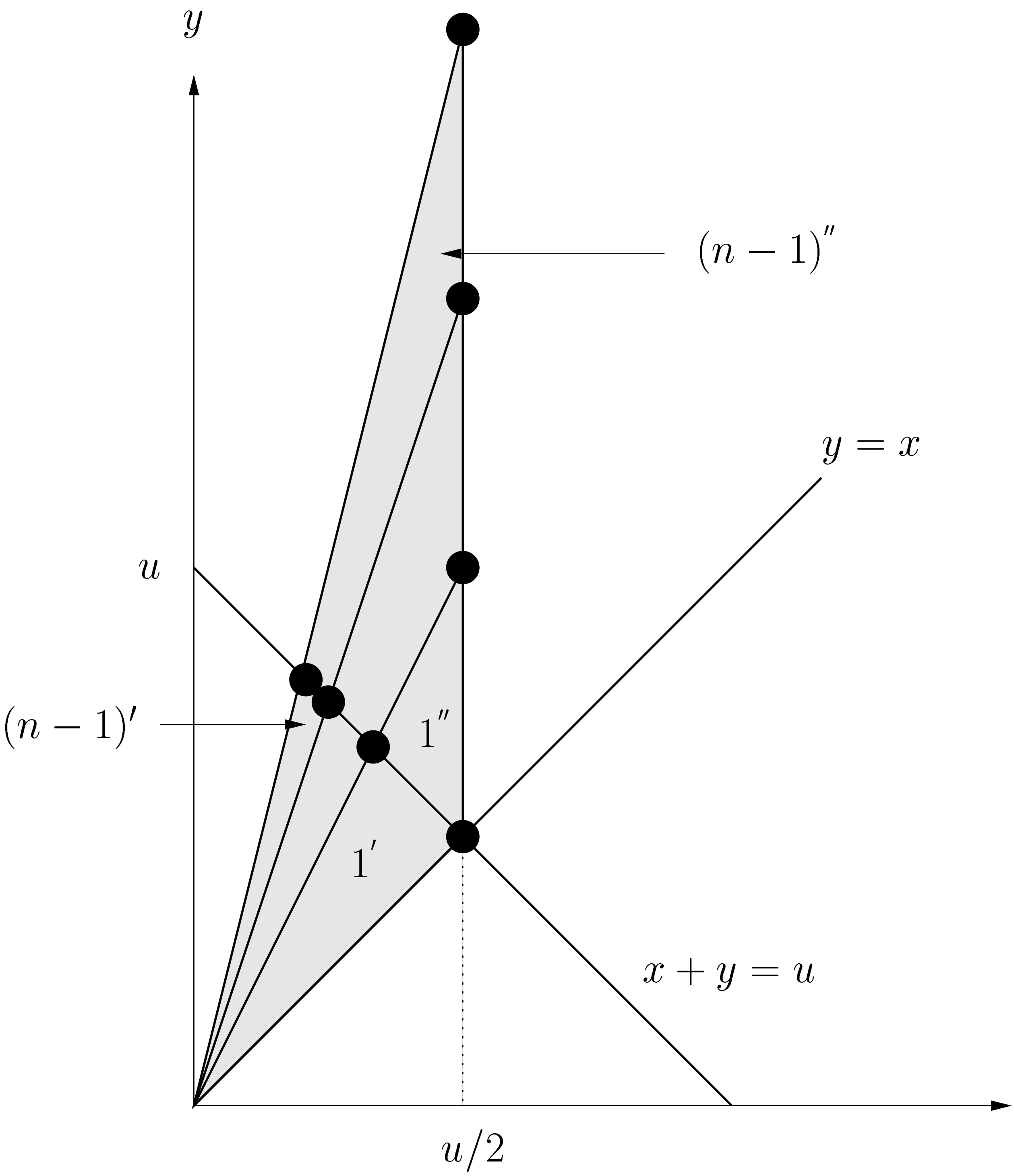


Figure 2:

Numerical Example

N geometric
M/G/1 queue $\longleftrightarrow$ ruin probability
$\mathbb{P}(W > x) = \psi(x)$
Pollaczeck-Khinchine-Beekman-Bowers:
common value $\mathbb{P}(X_1 + \cdots + X_N > x)$
$X \sim$ integrated tail

$R = 10, 1000, 000$ replications
95% conf int $\widehat{z}(1 \pm *\%)$

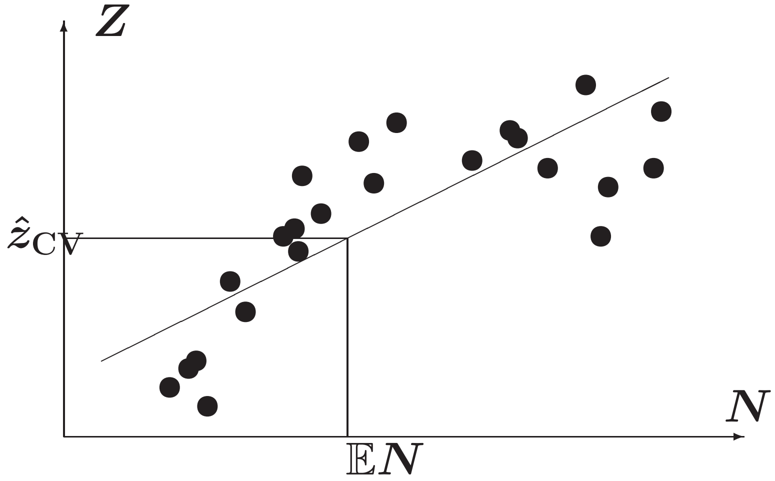
ρ	k	AB	CE	AK-IS	AK-Cond	JS	JS-AK
0.25	2	0.071	0.221	0.152	0.032	0.185	0.224
	5	0.105	0.436	0.260	0.031	0.421	0.506
	8	0.122	0.783	0.335	0.031	0.582	0.703
	11	0.115	1.856	0.397	0.031	0.713	0.859
0.5	2	0.111	1.475	0.192	0.045	0.253	0.402
	5	0.144	1.442	0.301	0.044	0.515	0.812
	8	0.146	8.180	0.380	0.044	0.702	1.098
	11	0.153	3.808	0.445	0.044	0.855	1.337
0.75	2	0.141	3.136	0.232	0.054	0.314	0.744
	5	0.205	7.433	0.341	0.054	0.591	1.381
	8	0.188	7.128	0.423	0.054	0.795	1.888
	11	0.180	17.787	0.494	0.054	0.960	2.272

Variance Reduction via N

Exploit that distr'n of N known

Control Variates:

regression of $Z = Z(x)$ upon N
uses $\mathbb{E}N$ only



Stratification : ...

Both guarantee variance reduction

ρ	k	AK-IS-CV	AK-Cond-CV	AK-Cond-Str	AK-IS-Str	JS-CV	JS-Str
0.25	2	0.150	0.008	0.008	0.121	0.132	0.181
	5	0.258	0.000	0.000	0.210	0.303	0.419
	8	0.335	0.000	0.000	0.272	0.420	0.583
	11	0.396	0.000	0.000	0.323	0.513	0.713
0.5	2	0.192	0.009	0.009	0.179	0.222	0.248
	5	0.301	0.000	0.000	0.285	0.458	0.512
	8	0.381	0.000	0.000	0.362	0.625	0.700
	11	0.445	0.000	0.000	0.424	0.761	0.855
0.75	2	0.232	0.009	0.011	0.225	0.301	0.308
	5	0.341	0.000	0.005	0.334	0.572	0.588
	8	0.423	0.000	0.005	0.419	0.773	0.793
	11	0.491	0.000	0.005	0.487	0.938	0.961

Explanation: Almost all variability in AK-Cond estimator comes from N

Ongoing Work

I: Dependent lognormals
(SA & Rojas-Nandayana)
Uses SA-Kroese ideas:

1. **IS: change distr'n of X_n only**

$$Z(x) \;=\; n\frac{f(X_n)}{f^*(Y_n)}I(S_n > x, M_n = X_n)$$

2. **Conditional MC:**

$$\begin{aligned} Z(x) &= n\,\mathbb{P}\big(S_n > x, M_n = X_n \,|\; X_1, \ldots, X_{n-1}\big) \\ &= n\,\overline{F}\big(M_{n-1} \vee (x - S_{n-1})\big). \end{aligned}$$

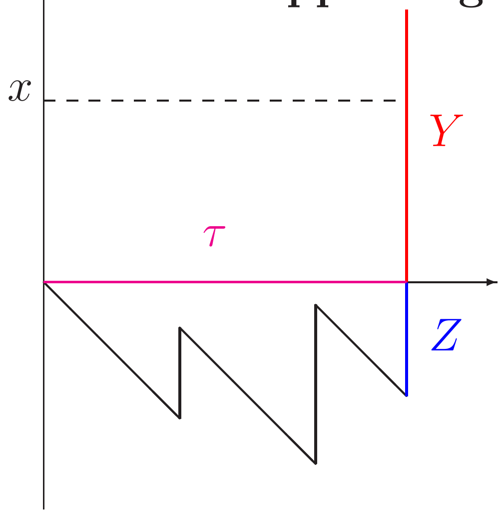
X_1, X_2 **positively dependent lognormals**
$Z(x) = \mathbb{P}(X_1 + X_2 > x)$
$X_1 = \mathrm{e}^{aU+bV_1}, \quad X_2 = \mathrm{e}^{aU+bV_2}$
U, V_1, V_2 **indep** $N(0,1)$
Simulate U **as** $N(\rho \log x, 1)$
Simulate V_1 **as** $N\big((0,1\big)$

$$Z(x) \;=\; 2\,\frac{\phi(U)}{\phi_{\rho\log x}(U)}\,\overline{\Phi}\Big(V_1 \vee \big(\log x\mathrm{e}^{-U} - \mathrm{e}^{bV_2}\big)/b\Big)$$

Different from Glasserman-Heidelberger-Shahabuddin

Ongoing Work

II: Transient M/G/1 = finite horizon ruin prob
(SA & Rolski)
Uses SA-Klüppelberg 97



Y **has cdf** $F_I(x) = \frac{1}{\mathbb{E}U}\int_0^x \mathbb{P}(U > y)\,\mathrm{d}y$
Z **given** $Y = y$ **distributed as** $U - y$ **given** $U > y$
path in $[0,\tau]$**: run backwards until 0 hit**

Algorithm for $\mathbb{P}\big(\tau(x) \le T(x)\big)$**:**
Generate N **with** $\mathbb{P}(N = n) = (1-\rho)\rho^n$
Generate $Y_1, \ldots, Y_{N-1}$ **from** F_I
Generate Y_N **from** F_I
given $Y_N > M_{N-1}$**,** $Y_1 + \ldots + Y_N > x$
Generate the Z_k **given the** Y_k **and** $Z_k \le T$
Generate τ_k **backwards**

$$Z(x) \;=\; N\, I\big(\tau_1+\cdots+\tau_N \le T\big) \prod_{k=1}^{N} \mathbb{P}\big(Z_k \le T \,\big|\, \cdots\big)\overline{F}_I(\cdots)$$