Stochastic Simulation

Jan-Pieter Dorsman[1] & Michel Mandjes[1,2,3]

[1]Korteweg-de Vries Institute for Mathematics, University of Amsterdam

[2]CWI, Amsterdam

[3]Eurandom, Eindhoven

University of Amsterdam,
Fall, 2017

Chapter VI
Rare-Event Simulation

Rare-Event Simulation

In rare-event simulation, we consider the problem of estimating a rare event A. I.e., think of $z := \mathbb{P}(A)$ as being in the order of 10^{-3} or less.

Rare-Event Simulation

In rare-event simulation, we consider the problem of estimating a rare event A. I.e., think of $z := \mathbb{P}(A)$ as being in the order of 10^{-3} or less.

Q: How to estimate these probabilities efficiently?

Rare-Event Simulation

In rare-event simulation, we consider the problem of estimating a rare event A. I.e., think of $z := \mathbb{P}(A)$ as being in the order of 10^{-3} or less.

Q: How to estimate these probabilities efficiently?

A: Naive thought: just do a crude Monte Carlo simulation using $Z = \mathbb{1}_{\{A\}}$, and let it last very long?

Rare-Event Simulation

In rare-event simulation, we consider the problem of estimating a rare event A. I.e., think of $z := \mathbb{P}(A)$ as being in the order of 10^{-3} or less.

Q: How to estimate these probabilities efficiently?

A: Naive thought: just do a crude Monte Carlo simulation using $Z = \mathbb{1}_{\{A\}}$, and let it last very long?

A: But we have seen this may take prohibitively long.

Rare-Event Simulation

As we have seen, number of runs required is roughly

$$R = \frac{100 \cdot 1.96^2 z(1-z)}{z^2}.$$

Rare-Event Simulation

As we have seen, number of runs required is roughly

$$R = \frac{100 \cdot 1.96^2 z(1-z)}{z^2}.$$

This number increases like z^{-1} towards ∞ as $z \downarrow 0$.

Rare-Event Simulation

As we have seen, number of runs required is roughly

$$R = \frac{100 \cdot 1.96^2 z(1-z)}{z^2}.$$

This number increases like z^{-1} towards ∞ as $z \downarrow 0$.

This is why we look at importance sampling.

Rare-Event Simulation

Formal setup (short recap):

- Let $A(x)$ be a family of rare events.

Rare-Event Simulation

Formal setup (short recap):

- Let $A(x)$ be a family of rare events.
- $z(x) := \mathbb{P}(A(x)) \to 0$ as $x \to \infty$

Rare-Event Simulation

Formal setup (short recap):

- Let $A(x)$ be a family of rare events.
- $z(x) := \mathbb{P}\left(A(x)\right) \to 0$ as $x \to \infty$
- Let $Z(x)$ be an unbiased estimator for $z(x)$: $\mathbb{E}\left[Z(x)\right] = z(x)$.

Rare-Event Simulation

Formal setup (short recap):

- Let $A(x)$ be a family of rare events.
- $z(x) := \mathbb{P}(A(x)) \to 0$ as $x \to \infty$
- Let $Z(x)$ be an unbiased estimator for $z(x)$: $\mathbb{E}[Z(x)] = z(x)$.
- An algorithm is defined as a family $Z(x)$ of such estimators.

Rare-Event Simulation

Formal setup (short recap):

- Let $A(x)$ be a family of rare events.
- $z(x) := \mathbb{P}(A(x)) \to 0$ as $x \to \infty$
- Let $Z(x)$ be an unbiased estimator for $z(x)$: $\mathbb{E}[Z(x)] = z(x)$.
- An algorithm is defined as a family $Z(x)$ of such estimators.

Q: find algorithm s such that the required R does not explode.
Different flavors: *bounded relative error*, *logarithmic efficiency*, ...

Two generic examples

Let S_n be defined as $\sum_{i=1}^{n} X_i$ for i.i.d. X_i.

Two generic examples

Let S_n be defined as $\sum_{i=1}^{n} X_i$ for i.i.d. X_i.

Example 1: partial sums. We wish to estimate

$$\pi_n := \mathbb{P}\left(\frac{S_n}{n} > a\right)$$

for n large and $a > \mathbb{E}X_1$.

Two generic examples

Let S_n be defined as $\sum_{i=1}^{n} X_i$ for i.i.d. X_i.

Example 1: partial sums. We wish to estimate

$$\pi_n := \mathbb{P}\left(\frac{S_n}{n} > a\right)$$

for n large and $a > \mathbb{E}X_1$.

Example 2: hitting probabilities. We wish to estimate

$$\alpha(B) := \mathbb{P}(\exists n \in \mathbb{N} : S_n \geq B),$$

for B large and $\mathbb{E}X_1 < 0$.

Two generic examples

These examples were dealt with by Jan-Pieter last week.

As they are the most important examples, I'll say a bit more about them, and explain at an intuitive level why exponential twisting works so nicely.

Two generic examples

These examples were dealt with by Jan-Pieter last week.

As they are the most important examples, I'll say a bit more about them, and explain at an intuitive level why exponential twisting works so nicely.

(It can be seen as my personal view on Section VI.6.)

Standard example 1: partial sums

Consider, with $S_n := \sum_{i=1}^n X_i$, the probability

$$\pi_n := \mathbb{P}\left(\frac{S_n}{n} > a\right)$$

for n large and $a > \mathbb{E}X_1$.

Standard example 1: partial sums

Consider, with $S_n := \sum_{i=1}^{n} X_i$, the probability

$$\pi_n := \mathbb{P}\left(\frac{S_n}{n} > a\right)$$

for n large and $a > \mathbb{E}X_1$.

This probability tends to 0 by virtue of the law of large numbers.

Standard example 1: partial sums

Bahadur & Rao:

Standard example 1: partial sums

Bahadur & Rao:

Assume moment generating function $\hat{F}(\theta) := \mathbb{E}e^{\theta X} < \infty$ for some $\theta > 0$ ($\sim$ light tails!)

Standard example 1: partial sums

Bahadur & Rao:

Assume moment generating function $\hat{F}(\theta) := \mathbb{E}e^{\theta X} < \infty$ for some $\theta > 0$ ($\sim$ light tails!)

Define Legendre transform

$$I(a) = \sup_{\theta}(\theta a - \log \hat{F}(\theta)).$$

Then

$$\pi_n \Big/ \left(\frac{C_a}{\sqrt{n}} e^{-nI(a)} \right) = \mathbb{P}\left(\frac{S_n}{n} > a \right) \Big/ \left(\frac{C_a}{\sqrt{n}} e^{-nI(a)} \right) \to 1$$

(if X non-lattice; if X lattice similar formula applies).

Standard example 1: partial sums

This formula only holds for $n \to \infty$, so there is still a need to estimate π_n!

How can that be efficiently done?

Naive simulation: as we have seen, # runs $\sim 400/\pi_n$, that is, grows exponentially in n...
Idea: *mimic conditional distribution.*

Intermezzo: conditional distribution

Of course, want would like to pick $\mathbb{Q}$ such that variance of new estimator is minimal. Say p is rare-event probability we wish to estimate.

Intermezzo: conditional distribution

Of course, want would like to pick $\mathbb{Q}$ such that variance of new estimator is minimal. Say p is rare-event probability we wish to estimate.

Proposal: let $\mathbb{Q}$ coincide with the distribution *conditional on* the event A:

$$\mathbb{Q}(\omega) = \mathbb{P}(\omega \,|\, \omega \in A) = \frac{\mathbb{P}(\omega)}{p}$$

if $\omega \in A$ and 0 else.

Intermezzo: conditional distribution

Then,

$$\mathbb{V}\mathrm{ar}_{\mathbb{Q}}(LI) = \mathbb{E}_{\mathbb{Q}}(L^2 I) - p^2 = \int_{\omega \in A} L^2 \mathrm{d}\mathbb{Q}(\omega) - p^2.$$

But

$$\int_{\omega \in A} L^2 \mathrm{d}\mathbb{Q}(\omega) = \int_{\omega \in A} p^2 \frac{\mathrm{d}\mathbb{P}(\omega)}{p} = p^2,$$

so that $\mathbb{V}\mathrm{ar}_{\mathbb{Q}}(LI) = 0$.

Intermezzo: conditional distribution

So we have found a zero-variance estimator!!

Intermezzo: conditional distribution

So we have found a zero-variance estimator!!

Drawback: *it requires knowledge of p*, and p is unknown...

Intermezzo: conditional distribution

So we have found a zero-variance estimator!!

Drawback: *it requires knowledge of p*, and p is unknown...

But we could try to mimic the conditional distribution!

Standard example 1: partial sums

Consider distribution of some X_i, and condition on $S_n \geq na$:

$$g(x) := \mathbb{P}(X_i \in [x, x + \mathrm{d}x) \,|\, S_n \geq na) = \frac{f(x)\,\mathbb{P}(S_{n-1} \geq na - x)}{\mathbb{P}(S_n \geq na)},$$

with $f(x)$ density of X_i under $\mathbb{P}$.

Standard example 1: partial sums

Consider distribution of some X_i, and condition on $S_n \geq na$:

$$g(x) := \mathbb{P}(X_i \in [x, x + \mathrm{d}x) \,|\, S_n \geq na) = \frac{f(x)\,\mathbb{P}(S_{n-1} \geq na - x)}{\mathbb{P}(S_n \geq na)},$$

with $f(x)$ density of X_i under $\mathbb{P}$.

Bahadur-Rao:

$$\mathbb{P}(S_n \geq na) \sim \frac{C_a}{\sqrt{n}} e^{-nI(a)},$$

whereas

$$\mathbb{P}(S_{n-1} \geq na - x) \sim \frac{C_a}{\sqrt{n}} \exp\left(-(n-1) \cdot I\left(\frac{na - x}{n-1}\right)\right).$$

(Use that $\sqrt{n} \sim \sqrt{n-1}$.)

Standard example 1: partial sums

Let us consider, for $n \to \infty$,

$$\exp\left(-(n-1)\cdot I\left(\frac{na-x}{n-1}\right)\right).$$

Well, it equals

$$\exp\left(-(n-1)\cdot I\left(\frac{(n-1)a+a-x}{n-1}\right)\right) \approx -(n-1)I(a)-(a-x)I'(a).$$

Standard example 1: partial sums

Let us consider, for $n \to \infty$,

$$\exp\left(-(n-1)\cdot I\left(\frac{na-x}{n-1}\right)\right).$$

Well, it equals

$$\exp\left(-(n-1)\cdot I\left(\frac{(n-1)a+a-x}{n-1}\right)\right) \approx -(n-1)I(a)-(a-x)I'(a).$$

As a consequence,

$$g(x) = \frac{f(x)\,\mathbb{P}(S_{n-1} \geq na-x)}{\mathbb{P}(S_n \geq na)} \approx f(x)\cdot\exp(I(a)+(x-a)I'(a)).$$

Standard example 1: partial sums

Now we verify that

$$g(x) = \frac{f(x)\,\mathbb{P}(S_{n-1} \geq na - x)}{\mathbb{P}(S_n \geq na)} \approx f(x) \cdot \exp(I(a) + (x-a)I'(a))$$

is a genuine density.

To this end, we need some basic convex analysis.

Standard example 1: partial sums

Now we verify that $f(x) \cdot \exp(I(a) + (x-a)I'(a))$ is a genuine density.

Define by $\theta(a)$ the optimizer in the definition of $I(a)$:

$$I(a) = \sup_\theta(\theta a - \log \hat{F}(\theta)) = \theta(a)a - \log \hat{F}(\theta(a)).$$

Hence $\theta(a)$ satisfies

$$a = \frac{\hat{F}'(\theta(a))}{\hat{F}(\theta(a))}.$$

This entails that

$$I'(a) = \theta(a) + \theta'(a)\left(a - \frac{\hat{F}'(\theta(a))}{\hat{F}(\theta(a))}\right) = \theta(a).$$

Standard example 1: partial sums

We conclude that we have to consider

$$f(x) \cdot \exp(I(a) + (x-a)\theta(a)).$$

Notice that this expression is non-negative, and in addition

$$\begin{aligned}
&\int f(x) \cdot \exp(I(a) + (x-a)\theta(a))\mathrm{d}x \\
&= \exp(I(a) - \theta(a)a) \cdot \int e^{\theta(a)x} f(x)\mathrm{d}x \\
&= \exp(I(a) - \theta(a)a)\hat{F}(\theta(a)) = e^{I(a)-I(a)} =
\end{aligned}$$

Standard example 1: partial sums

We conclude that we have to consider

$$f(x) \cdot \exp(I(a) + (x-a)\theta(a)).$$

Notice that this expression is non-negative, and in addition

$$\begin{aligned}
&\int f(x) \cdot \exp(I(a) + (x-a)\theta(a))\mathrm{d}x \\
&\quad = \exp(I(a) - \theta(a)a) \cdot \int e^{\theta(a)x} f(x)\mathrm{d}x \\
&\quad = \exp(I(a) - \theta(a)a)\hat{F}(\theta(a)) = e^{I(a)-I(a)} = 1.
\end{aligned}$$

Hence

$$f(x) \cdot \exp(I(a) + (x-a)\theta(a)) = f(x)\frac{e^{\theta(a)x}}{\hat{F}(\theta(a))}$$

is a density.

Standard example 1: partial sums

What did we do so far?

Standard example 1: partial sums

What did we do so far?

We determined a proxy for the distribution of the increments, conditional on $S_n \geq na$.

Standard example 1: partial sums

What did we do so far?

We determined a proxy for the distribution of the increments, conditional on $S_n \geq na$.

Now we analyze the performance of the resulting estimator. Recall that under the original measure

$$\mathbb{V}\mathrm{ar}(I) = \pi_n(1 - \pi_n) \approx \pi_n \approx (C_a/\sqrt{n}) \cdot e^{-nI(a)}.$$

Standard example 1: partial sums

Under the original measure

$$\mathbb{V}\mathrm{ar}(I) = \pi_n(1-\pi_n) \approx \pi_n \approx (C_a/\sqrt{n}) \cdot e^{-nI(a)}.$$

Under the new measure,

$$\mathbb{V}\mathrm{ar}_{\mathbb{Q}}(LI) = \mathbb{E}_{\mathbb{Q}}(L^2 I) - \pi_n^2.$$

So we have to analyze $\mathbb{E}_{\mathbb{Q}}(L^2 I)$ for large n.

Standard example 1: partial sums

Under the original measure

$$\mathbb{V}\mathrm{ar}(I) = \pi_n(1-\pi_n) \approx \pi_n \approx (C_a/\sqrt{n}) \cdot e^{-nI(a)}.$$

Under the new measure,

$$\mathbb{V}\mathrm{ar}_{\mathbb{Q}}(LI) = \mathbb{E}_{\mathbb{Q}}(L^2 I) - \pi_n^2.$$

So we have to analyze $\mathbb{E}_{\mathbb{Q}}(L^2 I)$ for large n.

Observe:

$$L = \prod_{i=1}^{n} \frac{f(X_i)}{g(X_i)} = \prod_{i=1}^{n} \frac{\hat{F}(\theta(a))}{e^{\theta(a)X_i}} = (\hat{F}(\theta(a)))^n \cdot e^{-\theta(a)S_n}.$$

Standard example 1: partial sums

If $I = 1$ (that is, $S_n \geq na$),

$$L = (\hat{F}(\theta(a)))^n \cdot e^{-\theta(a)S_n} \leq (\hat{F}(\theta(a)))^n \cdot e^{-\theta(a)a} = e^{-2nI(a)}.$$

Standard example 1: partial sums

If $I = 1$ (that is, $S_n \geq na$),

$$L = (\hat{F}(\theta(a)))^n \cdot e^{-\theta(a)S_n} \leq (\hat{F}(\theta(a)))^n \cdot e^{-\theta(a)a} = e^{-2nI(a)}.$$

In other words,

$$\mathbb{V}\mathrm{ar}_{\mathbb{Q}}(LI) = \mathbb{E}_{\mathbb{Q}}(L^2 I) - \pi_n^2 \leq \mathbb{E}_{\mathbb{Q}}(L^2 I) \leq e^{-2nI(a)}.$$

Recall that under $\mathbb{P}$ we had

$$\mathbb{V}\mathrm{ar}(I) = \pi_n(1 - \pi_n) \approx \pi_n \approx (C_a/\sqrt{n}) \cdot e^{-nI(a)}.$$

Standard example 1: partial sums

If $I = 1$ (that is, $S_n \geq na$),

$$L = (\hat{F}(\theta(a)))^n \cdot e^{-\theta(a) S_n} \leq (\hat{F}(\theta(a)))^n \cdot e^{-\theta(a) a} = e^{-2nI(a)}.$$

In other words,

$$\mathbb{V}\mathrm{ar}_{\mathbb{Q}}(LI) = \mathbb{E}_{\mathbb{Q}}(L^2 I) - \pi_n^2 \leq \mathbb{E}_{\mathbb{Q}}(L^2 I) \leq e^{-2nI(a)}.$$

Recall that under $\mathbb{P}$ we had

$$\mathbb{V}\mathrm{ar}(I) = \pi_n(1 - \pi_n) \approx \pi_n \approx (C_a/\sqrt{n}) \cdot e^{-nI(a)}.$$

$\Longrightarrow$ Substantial variance reduction!

Standard example 1: partial sums

As an aside: more refined analysis is possible:

$$\mathbb{V}\mathrm{ar}_{\mathbb{Q}}(LI) \sim \frac{\tilde{C}_a}{\sqrt{n}} e^{-2nI(a)}.$$

Straightforward computation: number of runs R required grows roughly proportional to $\sqrt{n}$ (rather than $1/\pi_n$, that is proportional to $\sqrt{n}e^{nI(a)}$).

Standard example 1: partial sums

As an aside: more refined analysis is possible:

$$\mathbb{V}\mathrm{ar}_{\mathbb{Q}}(LI) \sim \frac{\tilde{C}_a}{\sqrt{n}} e^{-2nI(a)}.$$

Straightforward computation: number of runs R required grows roughly proportional to $\sqrt{n}$ (rather than $1/\pi_n$, that is proportional to $\sqrt{n}e^{nI(a)}$).

$\Longrightarrow$ Procedure is logarithmically efficient

Standard example II: hitting probabilities

Focus now on

$$\alpha(B) := \mathbb{P}(\exists n \in \mathbb{N} : S_n \geq B),$$

where it is assumed that $\mathbb{E}X_1 < 0$.

Again *light-tailed* regime, that is, $\hat{F}(\theta) = \mathbb{E}e^{\theta X} = 1$ has a positive solution, say $\theta^\star$.

Standard example II: hitting probabilities

Again the recipe is: mimic distribution random walk, conditional on the rare event.

Standard example II: hitting probabilities

Again the recipe is: mimic distribution random walk, conditional on the rare event.

Cramér-Lundberg: $\alpha(B) \sim Ce^{-\theta^\star B}$.
Consider distribution of some X_i, and condition on $\{\exists n \in \mathbb{N} : S_n \geq B\}$: informally,

$$g(x) := \mathbb{P}(X_i \in [x, x + \mathrm{d}x) \,|\, \exists n \in \mathbb{N} : S_n \geq B) = \frac{f(x)\alpha(B-x)}{\alpha(B)},$$

with $f(x)$ density of X_i under $\mathbb{P}$.

Standard example II: hitting probabilities

Again the recipe is: mimic distribution random walk, conditional on the rare event.

Cramér-Lundberg: $\alpha(B) \sim Ce^{-\theta^\star B}$.
Consider distribution of some X_i, and condition on $\{\exists n \in \mathbb{N} : S_n \geq B\}$: informally,

$$g(x) := \mathbb{P}(X_i \in [x, x + \mathrm{d}x) \,|\, \exists n \in \mathbb{N} : S_n \geq B) = \frac{f(x)\alpha(B-x)}{\alpha(B)},$$

with $f(x)$ density of X_i under $\mathbb{P}$.

With Cramér-Lundberg:

$$g(x) \approx f(x) \cdot e^{\theta^\star x}.$$

This is a density as $\theta^\star$ solves $\hat{F}(\theta^\star) = 1$.

Standard example II: hitting probabilities

It can be proven that $\mathbb{E}_{\mathbb{Q}} X_1 > 0$, so event is not rare anymore. Hence, the hitting time T is finite a.s., so $I = 1$ a.s.

Standard example II: hitting probabilities

It can be proven that $\mathbb{E}_{\mathbb{Q}} X_1 > 0$, so event is not rare anymore. Hence, the hitting time T is finite a.s., so $I = 1$ a.s.

Also,

$$L = e^{-\theta^\star S_T}.$$

Thus

$$\alpha(B) = \mathbb{E}_{\mathbb{Q}}(L) = e^{-\theta^\star B}\mathbb{E}_{\mathbb{Q}} e^{\theta^\star (B - S_T)} \approx C e^{-\theta^\star B}.$$

Standard example II: hitting probabilities

This entails

$$\alpha(B) = \mathbb{E}_{\mathbb{Q}}(L) = e^{-\theta^\star B}\mathbb{E}_{\mathbb{Q}} e^{\theta^\star (B - S_T)} \approx C e^{-\theta^\star B}.$$

In addition

$$\mathbb{V}\mathrm{ar}_{\mathbb{Q}}(L) \sim \tilde{C} e^{-2\theta^\star B},$$

so that number of runs needed remains bounded when B grows!

Standard example II: hitting probabilities

$$\alpha(B) = \mathbb{E}_{\mathbb{Q}}(L) = e^{-\theta^\star B}\mathbb{E}_{\mathbb{Q}} e^{\theta^\star(B-S_T)} \approx C e^{-\theta^\star B}.$$

In addition

$$\mathbb{V}\mathrm{ar}_{\mathbb{Q}}(L) \sim \tilde{C} e^{-2\theta^\star B},$$

so that number of runs needed remains bounded when B grows!

Standard example II: hitting probabilities

$$\alpha(B) = \mathbb{E}_{\mathbb{Q}}(L) = e^{-\theta^\star B}\mathbb{E}_{\mathbb{Q}} e^{\theta^\star(B-S_T)} \approx Ce^{-\theta^\star B}.$$

In addition

$$\mathbb{V}\mathrm{ar}_{\mathbb{Q}}(L) \sim \tilde{C} e^{-2\theta^\star B},$$

so that number of runs needed remains bounded when B grows!

$\Longrightarrow$ Procedure has bounded relative error

Markov modulation

(Cf. Section VI.7.)

Markov modulation

(Cf. Section VI.7.)

Idea: often X_i are affected by external (background) process. Let background process be a discrete-time Markov chain on $\{1, \ldots, d\}$.

For instance, X_i is distributed as a sample from the random variable Y_j (with density $f_j(\cdot)$) when $J_i = j$.

Transition probabilities of Markov chain J are given by p_{ij}; equilibrium probabilities are π_i.

Markov modulation

We again focus on

$$\alpha(B) := \mathbb{P}(\exists n \in \mathbb{N} : S_n \geq B),$$

where it is assumed that

$$\sum_{j=1}^{d} \pi_j \mathbb{E} Y_j < 0.$$

Markov modulation

Idea: solve eigensystem

$$x_j = \mathbb{E}e^{\theta Y_j} \sum_{k=1}^{d} p_{jk} x_k.$$

Markov modulation

Idea: solve eigensystem

$$x_j = \mathbb{E}e^{\theta Y_j} \sum_{k=1}^{d} p_{jk} x_k.$$

Perron-Frobenius: there is a positive eigenvalue such that x is positive componentwise and $\theta > 0$.

Markov modulation

With (x, θ) solving the eigensystem

$$x_j = \mathbb{E}e^{\theta Y_j} \sum_{k=1}^{d} p_{jk} x_k,$$

clearly

$$q_{jk} = p_{jk} \frac{x_k}{x_j} \mathbb{E}e^{\theta Y_j}$$

form a transition probability matrix.

Markov modulation

With (x, θ) solving the eigensystem

$$x_j = \mathbb{E} e^{\theta Y_j} \sum_{k=1}^{d} p_{jk} x_k,$$

clearly

$$q_{jk} = p_{jk} \frac{x_k}{x_j} \mathbb{E} e^{\theta Y_j}$$

form a transition probability matrix.

In addition, define the twisted increments by

$$g_j(y) = f_j(y) \frac{e^{\theta y}}{\mathbb{E}\, e^{\theta Y_j}}.$$

Markov modulation

Observe

$$\alpha(B) := \mathbb{P}(\exists n \in \mathbb{N} : S_n \geq B) = \mathbb{P}(T_B < \infty),$$

with T_B the first hitting time of $[B, \infty)$.

Markov modulation

Observe

$$\alpha(B) := \mathbb{P}(\exists n \in \mathbb{N} : S_n \geq B) = \mathbb{P}(T_B < \infty),$$

with T_B the first hitting time of $[B, \infty)$.

Can we evaluate likelihood ratio?

Markov modulation

Observe

$$\alpha(B) := \mathbb{P}(\exists n \in \mathbb{N} : S_n \geq B) = \mathbb{P}(T_B < \infty),$$

with T_B the first hitting time of $[B, \infty)$.

Can we evaluate likelihood ratio?

Observe that $T_B < \infty$ with probability 1 under $\mathbb{Q}$.

Markov modulation

By definition,

$$L = \prod_{i=1}^{T_B} \left(\frac{p_{J_i J_{i+1}}}{q_{J_i J_{i+1}}} \right) \left(\frac{f_{J_i}(X_i)}{g_{J_i}(X_i)} \right).$$

Markov modulation

By definition,

$$L = \prod_{i=1}^{T_B} \left(\frac{p_{J_i J_{i+1}}}{q_{J_i J_{i+1}}} \right) \left(\frac{f_{J_i}(X_i)}{g_{J_i}(X_i)} \right) .$$

This can be simplified to

$$L = \prod_{i=1}^{T_B} \left(\frac{x_{J_i}}{x_{J_{i+1}}} \frac{1}{\mathbb{E}\, e^{\theta Y_{J_i}}} \right) \left(\frac{\mathbb{E}\, e^{\theta Y_{J_i}}}{e^{\theta Y_{J_i}}} \right) ,$$

which equals

$$\frac{x_{J_0}}{x_{J_{T_B+1}}} e^{-\theta S_{T_B}} .$$

Markov modulation

From

$$L = \frac{x_{J_0}}{x_{J_{\tau_B+1}}} e^{-\theta S_{\tau_B}},$$

we conclude that, for a finite contant C,

$$\alpha(B) = \mathbb{E}_{\mathbb{Q}}(LI) = \mathbb{E}_{\mathbb{Q}}(L) \leq Ce^{-\theta S_{\tau_B}}.$$

Actually in can be proven that, as $B \to \infty$,

$$\alpha(B) \sim \tilde{C} e^{-\theta S_{\tau_B}},$$

and that the procedure has bounded relative error.

Caveats

Slight variations to the 'leading examples' model may already cause serious problems... For $a_- < \mu := \mathbb{E}X_1 < a_+$,

$$\pi_n := \mathbb{P}\left(\frac{S_n}{n} \notin (a_-, a_+)\right).$$

Assume w.l.o.g. $I(a_+) < I(a_-)$.

Caveats

Slight variations to the 'leading examples' model may already cause serious problems... For $a_- < \mu := \mathbb{E}X_1 < a_+$,

$$\pi_n := \mathbb{P}\left(\frac{S_n}{n} \notin (a_-, a_+)\right).$$

Assume w.l.o.g. $I(a_+) < I(a_-)$. This intuitively means that most likely point in set is a_+.

Define, as before,

$$\theta(a) := \arg\sup_\theta(\theta a - \log \hat{F}(\theta)).$$

Caveats

Naïvely, one would use

$$g(x) = f(x)\frac{e^{\theta(a_+)x}}{\hat{F}(\theta(a_+))}.$$

To check how this change of measure performs, we analyze

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}}(L^2 I) &= \Big(\mathbb{E}_{\mathbb{Q}}(e^{-2\theta(a_+)S_n}1\{S_n \geqslant na_+\}) \\ &\quad + \mathbb{E}_{\mathbb{Q}}(e^{-2\theta(a_+)S_n}1\{S_n \leqslant na_-\})\Big) \times (\hat{F}(\theta(a_+)))^{2n}.\end{aligned}$$

Caveats

Consider only exponential terms (in n).

Caveats

Consider only exponential terms (in n).

First term:

$$\lim_{n\to\infty} \frac{1}{n} \log \left(\mathbb{E}_{\mathbb{Q}}(e^{-2\theta(a_+)S_n} 1\{S_n \geqslant na_+\})(\hat{F}(\theta(a_+)))^{2n} \right) = -2I(a_+).$$

Caveats

Consider only exponential terms (in n).

First term:

$$\lim_{n\to\infty} \frac{1}{n} \log \left(\mathbb{E}_{\mathbb{Q}}(e^{-2\theta(a_+)S_n} 1\{S_n \geqslant na_+\})(\hat{F}(\theta(a_+)))^{2n} \right) = -2I(a_+).$$

Second term is more difficult. Observe: undershoot below level na_- will be modest. Hence

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} \log \left(\mathbb{E}_{\mathbb{Q}}(e^{-2\theta(a_+)S_n} 1\{S_n \leqslant na_-\}) \times (\hat{F}(\theta(a_+)))^{2n} \right) \\ &= \; -2\theta(a_+)a_- + \lim_{n\to\infty} \frac{1}{n} \log \mathbb{Q}\left(\frac{S_n}{n} \leqslant a_- \right) + 2 \log \hat{F}(\theta(a_+)). \end{aligned}$$

Caveats

To find decay rate of $\mathbb{Q}(S_n/n \leqslant a_-)$, let $J(\cdot)$ be the Legendre-Fenchel transform under $\mathbb{Q}$, that is

$$
\begin{aligned}
J(a_-) = \sup_\theta & \left[\theta a_- - \log \mathbb{E}_{\mathbb{Q}} e^{\theta X_1}\right] \\
= & \sup_\theta \left[\theta a_- - \log \phi(\theta + \theta(a_+)) + \log \hat{F}(\theta(a_+))\right] \\
= & -\theta(a_+) a_- + \sup_\theta \left[(\theta + \theta(a_+)) a_- - \log \hat{F}(\theta + \theta(a_+))\right] \\
& \qquad + \log \hat{F}(\theta(a_+)) \\
= & -\theta(a_+) a_- + I(a_-) + \log \hat{F}(\theta(a_+)).
\end{aligned}
$$

We obtain that

$$
\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \log & \left(\mathbb{E}_{\mathbb{Q}}(e^{-2\theta(a_+)S_n} 1\{S_n \leqslant n a_-\}) \times (\hat{F}(\theta(a_+)))^{2n}\right) \\
& = -\theta(a_+) a_- - I(a_-) + \log \hat{F}(\theta(a_+)).
\end{aligned}
$$

Caveats

Hence logarithmic efficiency if

$$-2I(a_+) \geqslant -\theta(a_+)a_- - I(a_-) + \log \hat{F}(\theta(a_+)),$$

or

$$I(a_+) + \theta(a_+)a_+ \leqslant I(a_-) + \theta(a_+)a_-.$$

This condition is not always met!

Caveats

Example. Let the X_i be i.i.d. samples from an exponential distribution with mean 1. We take $a_+ = 2$ and $a_- < 1$. We have that $\theta(a_+) = 1 - 1/a_+$ and $I(a_+) = a_+ - 1 - \log a_+$. The inequality to be verified is therefore

$$3 - \log 2 \leqslant \frac{3}{2}a_- - \log a_-, \quad \text{or} \quad \log a_- \leqslant \frac{3}{2}a_- - 3 + \log 2.$$

Numerical search: condition is met when $0 < a_- < 0.119$.

Caveats & remedies

This type of problems arises often when ‘overflow set’ has ‘unfavorable’ shape.

Remedy 1: Split rare event of interest in disjoint parts:

$$\alpha(n) = \mathbb{P}(\bar{X}_n \notin (a_-, a_+)) = \mathbb{P}\left(\frac{S_n}{n} \geqslant a_+\right) + \mathbb{P}\left(\frac{S_n}{n} \leqslant a_-\right),$$

and estimate each separately.

Caveats & remedies

Remedy 2: Adaptive change-of-measure. Main problem: rare event can be reached through path that is 'far away' from most-likely path, leading to large likelihood ratio. Can be avoided by updating the change-of-measure during simulation run, depending on current position.

Caveats & remedies

Remedy 3: Random change-of-measure. In this approach, one flips a coin, and the outcome decides from which twisted distribution one samples. Assume $p \in (0,1)$; measure $\mathbb{Q}$: tilt with parameter $\theta(a_+) > 0$ ($\theta(a_-) < 0$) with probability p ($1-p$, respectively). Likelihood equals

$$L = \left(p e^{\theta(a_+) S_n} \hat{F}(\theta(a_+))^n + (1-p) e^{\theta(a_-) S_n} \hat{F}(\theta(a_-))^n \right)^{-1}.$$

Caveats & remedies

Then $\mathbb{E}_{\mathbb{Q}}(L^2 I)$ is smaller than

$$\mathbb{E}_{\mathbb{Q}}\left(\left(\frac{1\{S_n \geqslant na_+\}}{pe^{\theta(a_+)S_n}\hat{F}(\theta(a_+))}\right)^2 + \left(\frac{1\{S_n \leqslant na_-\}}{(1-p)e^{\theta(a_-)S_n}\hat{F}(\theta(a_-))}\right)^2\right)$$

$$\leqslant \quad \frac{1}{p^2}e^{-2nI(a_+)} + \frac{1}{(1-p)^2}e^{-2nI(a_-)},$$

which has decay rate $-2I(a_+)$. Hence procedure is logarithmically efficient.

EPILOGUE

EPILOGUE

- We have treated a series of techniques to be used in discrete-event simulations;

EPILOGUE

- We have treated a series of techniques to be used in discrete-event simulations;
- We have described how to statistically assess simulation output;

EPILOGUE

- We have treated a series of techniques to be used in discrete-event simulations;
- We have described how to statistically assess simulation output;
- We have treated a number of relevant (and more advanced) topics, such as derivative estimation and rare-event simulation;

EPILOGUE

- We have treated a series of techniques to be used in discrete-event simulations;
- We have described how to statistically assess simulation output;
- We have treated a number of relevant (and more advanced) topics, such as derivative estimation and rare-event simulation;
- ... and we wish you all the best in preparing the exam!