# 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:** Let's see an example to see whether this is a valid answer.

## Rare-Event Simulation

Let's see whether this works... Let $Z = \mathbb{1}_{\{A\}}$, so that indeed $z = \mathbb{E}[Z] = \mathbb{P}(A)$. $Z$ is then Bernoulli distributed with parameter $z$.

## Rare-Event Simulation

Let's see whether this works... Let $Z = \mathbb{1}_{\{A\}}$, so that indeed $z = \mathbb{E}[Z] = \mathbb{P}(A)$. $Z$ is then Bernoulli distributed with parameter $z$.
We know that $\text{Var}[Z] = z(1-z)$. This means that

$$\frac{\sigma_Z}{z} = \sqrt{\frac{1-z}{z}}$$

which behaves like $z^{-\frac{1}{2}}$ for small $z$.

Recall the 95%-confidence interval for $\hat{z}_R$:

$$\left(\hat{z}_R - 1.96\frac{\sigma_Z}{\sqrt{R}}, \hat{z}_R + 1.96\frac{\sigma_Z}{\sqrt{R}}\right)$$

Suppose that we want to acquire a precision such that the width of our confidence interval is about 20% of the value of $z$. In other words:

$$0.1 = 1.96\frac{\sigma_Z}{z\sqrt{R}}$$

or

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

Recall the 95%-confidence interval for $\hat{z}_R$:

$$\left(\hat{z}_R - 1.96\frac{\sigma_Z}{\sqrt{R}}, \hat{z}_R + 1.96\frac{\sigma_Z}{\sqrt{R}}\right)$$

Suppose that we want to acquire a precision such that the width of our confidence interval is about 20% of the value of $z$. In other words:

$$0.1 = 1.96\frac{\sigma_Z}{z\sqrt{R}}$$

or

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

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

**A:** No, when $z$ becomes small enough, there is always a point at which crude Monte Carlo will fail as $R$ gets too large.

We will need other machinery in this setting. Importance sampling will turn out to be a useful tool.

# Rare-Event Simulation

Formal setup:

- Let $A(x)$ be a family of rare events, with e.g. $x \in \mathbb{R}_+$ or $x \in \mathbb{N}$.

# Rare-Event Simulation

Formal setup:

- Let $A(x)$ be a family of rare events, with e.g. $x \in \mathbb{R}_+$ or $x \in \mathbb{N}$.
- $z(x) := \mathbb{P}(A(x)) \to 0$ as $x \to \infty$

# Rare-Event Simulation

Formal setup:

- Let $A(x)$ be a family of rare events, with e.g. $x \in \mathbb{R}_+$ or $x \in \mathbb{N}$.
- $z(x) := \mathbb{P}(A(x)) \to 0$ as $x \to \infty$
- Let $Z(x)$ be an unbiased estimator for $z(x)$: $\mathbb{E}[Z(x)] = z(x)$.

# Rare-Event Simulation

Formal setup:

- Let $A(x)$ be a family of rare events, with e.g. $x \in \mathbb{R}_+$ or $x \in \mathbb{N}$.
- $z(x) := \mathbb{P}\left(A(x)\right) \to 0$ as $x \to \infty$
- Let $Z(x)$ be an unbiased estimator for $z(x)$: $\mathbb{E}\left[Z(x)\right] = z(x)$.
- An algorithm is defined as a family $Z(x)$ of such estimators.

## Rare-Event Simulation

Formal setup:

- Let $A(x)$ be a family of rare events, with e.g. $x \in \mathbb{R}_+$ or $x \in \mathbb{N}$.
- $z(x) := \mathbb{P}(A(x)) \to 0$ 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.

Quest: find well-performing algorithms such that the required $R$ does not explode.

## Rare-Event Simulation

Formal setup:

- Let $A(x)$ be a family of rare events, with e.g. $x \in \mathbb{R}_+$ or $x \in \mathbb{N}$.
- $z(x) := \mathbb{P}\left(A(x)\right) \to 0$ as $x \to \infty$
- Let $Z(x)$ be an unbiased estimator for $z(x)$: $\mathbb{E}\left[Z(x)\right] = z(x)$.
- An algorithm is defined as a family $Z(x)$ of such estimators.

Quest: find well-performing algorithms such that the required $R$ does not explode. What does this mean?

Formal setup:

- Let $A(x)$ be a family of rare events, with e.g. $x \in \mathbb{R}_+$ or $x \in \mathbb{N}$.
- $z(x) := \mathbb{P}(A(x)) \to 0$ 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.

Quest: find well-performing algorithms such that the required $R$ does not explode. What does this mean?

Recall the expression we just had:

$$R = \frac{100 \cdot 1.96^2 \mathrm{Var}[Z(x)]}{z(x)^2}$$

We wish $R$ to stay finite as $z(x) \downarrow 0$. This happens, when $Z(x)$ has a *bounded relative error*:

$$\limsup_{x \to \infty} \frac{\mathrm{Var}[Z(x)]}{z(x)^2} < \infty.$$

Bounded relative error:

$$\limsup_{x\to\infty} \frac{\mathrm{Var}\left[Z(x)\right]}{z(x)^2} < \infty.$$

In practice, we often check whether a variant of this condition holds, *logarithmic efficiency*:

$$\limsup_{x\to\infty} \frac{\mathrm{Var}\left[Z(x)\right]}{z(x)^{2-\epsilon}} = 0.$$

for any $\epsilon > 0$.
This condition is slightly weaker.

Bounded relative error:

$$\limsup_{x\to\infty} \frac{\mathrm{Var}\left[Z(x)\right]}{z(x)^2} < \infty.$$

In practice, we often check whether a variant of this condition holds, *logarithmic efficiency*:

$$\limsup_{x\to\infty} \frac{\mathrm{Var}\left[Z(x)\right]}{z(x)^{2-\epsilon}} = 0.$$

for any $\epsilon > 0$.
This condition is slightly weaker. Example: if $z(x) \sim Ce^{-\gamma x}$ and $\mathrm{Var}\left[Z(x)\right] \sim x^p e^{-2\gamma x}$, then we have logarithmic efficiency, but no bounded relative error.

Bounded relative error:

$$\limsup_{x\to\infty} \frac{\mathrm{Var}\left[Z(x)\right]}{z(x)^2} < \infty.$$

In practice, we often check whether a variant of this condition holds, *logarithmic efficiency*:

$$\limsup_{x\to\infty} \frac{\mathrm{Var}\left[Z(x)\right]}{z(x)^{2-\epsilon}} = 0.$$

for any $\epsilon > 0$.
This condition is slightly weaker. Example: if $z(x) \sim Ce^{-\gamma x}$ and $\mathrm{Var}\left[Z(x)\right] \sim x^p e^{-2\gamma x}$, then we have logarithmic efficiency, but no bounded relative error.

Why logarithmic efficiency?

- The difference is minor from a practical point of view.

Bounded relative error:

$$\limsup_{x\to\infty} \frac{\mathrm{Var}\left[Z(x)\right]}{z(x)^2} < \infty.$$

In practice, we often check whether a variant of this condition holds, *logarithmic efficiency*:

$$\limsup_{x\to\infty} \frac{\mathrm{Var}\left[Z(x)\right]}{z(x)^{2-\epsilon}} = 0.$$

for any $\epsilon > 0$.
This condition is slightly weaker. Example: if $z(x) \sim Ce^{-\gamma x}$ and $\mathrm{Var}\left[Z(x)\right] \sim x^p e^{-2\gamma x}$, then we have logarithmic efficiency, but no bounded relative error.

Why logarithmic efficiency?

- The difference is minor from a practical point of view.
- Logarithmic efficiency is often easier to verify.

Hence, our quest: find logarithmically efficient estimators.

# Rare-Event Simulation

Example. Let $N$ be a geometric r.v. with success parameter $\pi$, i.e. $\mathbb{P}(N = n) = \pi(1-\pi)^{n-1}$, and consider $z := \mathbb{P}(N \leq m) = 1 - (1-\pi)^m$.

Example. Let $N$ be a geometric r.v. with success parameter $\pi$, i.e. $\mathbb{P}(N = n) = \pi(1-\pi)^{n-1}$, and consider $z := \mathbb{P}(N \leq m) = 1 - (1-\pi)^m$.

Now, let $\pi = \pi(x)$ depend on a parameter $x$ such that $\pi(x) \downarrow 0$ as $x \to \infty$ so that $z = z(x) \downarrow 0$.

Example. Let $N$ be a geometric r.v. with success parameter $\pi$, i.e. $\mathbb{P}(N = n) = \pi(1-\pi)^{n-1}$, and consider $z := \mathbb{P}(N \le m) = 1 - (1-\pi)^m$.

Now, let $\pi = \pi(x)$ depend on a parameter $x$ such that $\pi(x) \downarrow 0$ as $x \to \infty$ so that $z = z(x) \downarrow 0$.
For simulating $z$, we consider importance sampling such that $N$ is instead simulated from a geometric distribution with a success parameter $\tilde{\pi}$, which does not depend on $x$.

## Rare-Event Simulation

Example. Let $N$ be a geometric r.v. with success parameter $\pi$, i.e. $\mathbb{P}(N = n) = \pi(1-\pi)^{n-1}$, and consider $z := \mathbb{P}(N \le m) = 1 - (1-\pi)^m$.

Now, let $\pi = \pi(x)$ depend on a parameter $x$ such that $\pi(x) \downarrow 0$ as $x \to \infty$ so that $z = z(x) \downarrow 0$.
For simulating $z$, we consider importance sampling such that $N$ is instead simulated from a geometric distribution with a success parameter $\tilde{\pi}$, which does not depend on $x$.

A family $\{Z(x)\}$ of estimators is now given by

$$Z(x) = \frac{1}{R}\sum_{i=1}^{R} \mathbb{1}_{\{N_i \le m\}} \frac{\pi(x)(1-\pi(x))^{N_i-1}}{\tilde{\pi}(1-\tilde{\pi})^{N_i-1}},$$

The book shows that

$$\limsup_{x\to\infty} \frac{\mathrm{Var}\left[Z(x)\right]}{z(x)^2} \le \limsup_{x\to\infty} \frac{\mathbb{E}\left[Z(x)^2\right]}{z(x)^2} = e - 1.$$

Hence, logarithmically efficient and bounded relative error!

## Rare-Event Simulation

We now proceed to the study of sums of light-tailed random variables (i.e. the relevant tails decay at least at an exponential rate).

Recall the idea of exponential tilting: suppose that $X_1, \ldots, X_n$ are i.i.d with common density $f(x)$. The importance distribution then preserves the i.i.d. property but changes $f(x)$ to $g_\theta(x) = \frac{e^{\theta x}}{\mathbb{E}[e^{\theta X}]} f(x)$. Then,

$$L_{n,\theta} = \prod_{i=1}^{n} \frac{f(X_i)}{g_\theta(X_i)} = e^{-\theta S_n} \hat{F}[\theta]^n = e^{-\theta S_n + n\kappa(\theta)},$$

where $S_n = \sum_{i=1}^{n} X_i$, $\hat{F}[\theta] = \mathbb{E}\left[e^{\theta X}\right]$, $\kappa(\theta) = \log \hat{F}[\theta]$.

## Rare-Event Simulation

We now proceed to the study of sums of light-tailed random variables (i.e. the relevant tails decay at least at an exponential rate).

Recall the idea of exponential tilting: suppose that $X_1, \ldots, X_n$ are i.i.d with common density $f(x)$. The importance distribution then preserves the i.i.d. property but changes $f(x)$ to $g_\theta(x) = \frac{e^{\theta x}}{\mathbb{E}[e^{\theta X}]} f(x)$. Then,

$$L_{n,\theta} = \prod_{i=1}^{n} \frac{f(X_i)}{g_\theta(X_i)} = e^{-\theta S_n} \hat{F}[\theta]^n = e^{-\theta S_n + n\kappa(\theta)},$$

where $S_n = \sum_{i=1}^n X_i$, $\hat{F}[\theta] = \mathbb{E}\left[e^{\theta X}\right]$, $\kappa(\theta) = \log \hat{F}[\theta]$. Moreover,

$$\begin{aligned} \mathbb{E}\left[h(X_1, \ldots, X_n)\right] &= \mathbb{E}_\theta\left[h(X_1, \ldots, X_n) L_{n,\theta}\right] \\ &= \mathbb{E}_\theta\left[h(X_1, \ldots, X_n) e^{-\theta S_n + n\kappa(\theta)}\right]. \end{aligned}$$

## Rare-Event Simulation

We now proceed to the study of sums of light-tailed random variables (i.e. the relevant tails decay at least at an exponential rate).

Recall the idea of exponential tilting: suppose that $X_1, \ldots, X_n$ are i.i.d with common density $f(x)$. The importance distribution then preserves the i.i.d. property but changes $f(x)$ to $g_\theta(x) = \frac{e^{\theta x}}{\mathbb{E}[e^{\theta X}]} f(x)$. Then,

$$L_{n,\theta} = \prod_{i=1}^{n} \frac{f(X_i)}{g_\theta(X_i)} = e^{-\theta S_n} \hat{F}[\theta]^n = e^{-\theta S_n + n\kappa(\theta)},$$

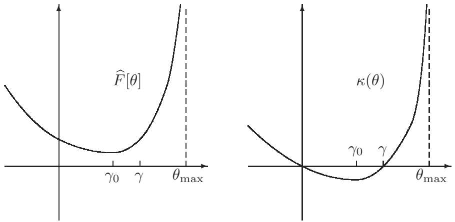
where $S_n = \sum_{i=1}^n X_i$, $\hat{F}[\theta] = \mathbb{E}\left[e^{\theta X}\right]$, $\kappa(\theta) = \log \hat{F}[\theta]$. Moreover,

$$\begin{aligned}
\mathbb{E}\left[h(X_1, \ldots, X_n)\right] &= \mathbb{E}_\theta\left[h(X_1, \ldots, X_n) L_{n,\theta}\right] \\
&= \mathbb{E}_\theta\left[h(X_1, \ldots, X_n) e^{-\theta S_n + n\kappa(\theta)}\right].
\end{aligned}$$

What do $\hat{F}[\theta]$ and $\kappa(\theta)$ typically look like?

# Rare-Event Simulation

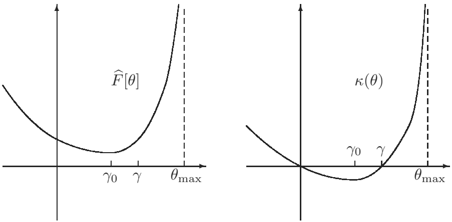
These are $\hat{F}[\theta]$ and $\kappa(\theta)$ for a distribution $F$ with negative mean and $F(0) < 1$:




- For $\theta \geq \theta_{max}$, $\hat{F}(\theta) = \infty$.
- $\gamma_0$ is the solution of $\hat{F}'[\theta] = \kappa'(\theta) = 0$.
- $\gamma$ is the solution to $\hat{F}[\gamma] = 1 - \kappa(\gamma) = 1$.

# Rare-Event Simulation

These are $\hat{F}[\theta]$ and $\kappa(\theta)$ for a distribution $F$ with negative mean and $F(0) < 1$:




- For $\theta \geq \theta_{max}$, $\hat{F}(\theta) = \infty$.
- $\gamma_0$ is the solution of $\hat{F}'[\theta] = \kappa'(\theta) = 0$.
- $\gamma$ is the solution to $\hat{F}[\gamma] = 1 - \kappa(\gamma) = 1$.

We also define the notion of changed drift:

$$\mu_\theta = \mathbb{E}_\theta[X] = \mathbb{E}\left[\frac{Xe^{\theta X}}{\hat{F}[\theta]}\right] = \frac{\hat{F}'[\theta]}{\hat{F}[\theta]} = \kappa'(\theta).$$

Thus, $\mu_\theta < 0$ when $\theta < \gamma_0$, $\mu_{\gamma_0} = 0$ and $\mu_\theta > 0$ when $\theta > \gamma_0$.

Suppose that the support of $X$ is not contained in $(-\infty, 0]$, but that still, $\mathbb{E}[X] < 0$.

Now: **Siegmund's algorithm** to consider the problem of estimating, for large $x$,

$$z(x) = \mathbb{P}(\tau(x) < \infty), \text{ where } \tau(x) := \inf\{n : S_n > x\}.$$

Idea: use exponential tilting to obtain

$$\begin{aligned} z(x) &= \mathbb{E}\left[\mathbb{1}_{\{\tau(x)<\infty\}}\right] = \mathbb{E}_\theta\left[\mathbb{1}_{\{\tau(x)<\infty\}} L_{\tau(x),\theta}\right] \\ &= \mathbb{E}_\theta\left[\mathbb{1}_{\{\tau(x)<\infty\}} e^{-\theta S_{\tau(x)} + \tau(x)\kappa(\theta)}\right]. \end{aligned}$$

Suppose that the support of $X$ is not contained in $(-\infty, 0]$, but that still, $\mathbb{E}[X] < 0$.

Now: **Siegmund's algorithm** to consider the problem of estimating, for large $x$,

$$z(x) = \mathbb{P}(\tau(x) < \infty), \text{ where } \tau(x) := \inf\{n : S_n > x\}.$$

Idea: use exponential tilting to obtain

$$\begin{aligned} z(x) &= \mathbb{E}\left[\mathbb{1}_{\{\tau(x)<\infty\}}\right] = \mathbb{E}_\theta\left[\mathbb{1}_{\{\tau(x)<\infty\}} L_{\tau(x),\theta}\right] \\ &= \mathbb{E}_\theta\left[\mathbb{1}_{\{\tau(x)<\infty\}} e^{-\theta S_{\tau(x)} + \tau(x)\kappa(\theta)}\right]. \end{aligned}$$

**Q:** How to choose $\theta$?

**A:** Well, should have at the very least that $\mathbb{P}_\theta(\tau(x) < \infty) = 1$,

Suppose that the support of $X$ is not contained in $(-\infty, 0]$, but that still, $\mathbb{E}[X] < 0$.

Now: **Siegmund's algorithm** to consider the problem of estimating, for large $x$,

$$z(x) = \mathbb{P}(\tau(x) < \infty), \text{ where } \tau(x) := \inf\{n : S_n > x\}.$$

Idea: use exponential tilting to obtain

$$\begin{aligned} z(x) &= \mathbb{E}\left[\mathbb{1}_{\{\tau(x)<\infty\}}\right] = \mathbb{E}_\theta\left[\mathbb{1}_{\{\tau(x)<\infty\}} L_{\tau(x),\theta}\right] \\ &= \mathbb{E}_\theta\left[\mathbb{1}_{\{\tau(x)<\infty\}} e^{-\theta S_{\tau(x)} + \tau(x)\kappa(\theta)}\right]. \end{aligned}$$

**Q:** How to choose $\theta$?

**A:** Well, should have at the very least that $\mathbb{P}_\theta(\tau(x) < \infty) = 1$, which happens when $\mathbb{E}_\theta[X] > 0$, or $\theta > \gamma_0$. Then, we have that

$$z(x) = \mathbb{E}_\theta\left[e^{-\theta S_{\tau(x)} + \tau(x)\kappa(\theta)}\right].$$

## Rare-Event Simulation

Actually, when a positive solution $\gamma$ to $\hat{F}[\gamma] = 1 - \kappa(\theta) = 1$ exists, $\theta = \gamma$ turns out to be optimal. Then, we have that

$$z(x) = \mathbb{E}_\theta \left[ e^{-\theta S_{\tau(x)} + \tau(x)\kappa(\theta)} \right] = \mathbb{E}_\gamma \left[ e^{-\gamma S_{\tau(x)}} \right] = e^{-\gamma x} \mathbb{E}_\gamma \left[ e^{-\gamma \xi(x)} \right],$$

where $\xi(x) = S_{\tau(x)} - x$ is the overshoot.

## Rare-Event Simulation

Actually, when a positive solution $\gamma$ to $\hat{F}[\gamma] = 1 - \kappa(\theta) = 1$ exists, $\theta = \gamma$ turns out to be optimal. Then, we have that

$$z(x) = \mathbb{E}_\theta \left[ e^{-\theta S_{\tau(x)} + \tau(x)\kappa(\theta)} \right] = \mathbb{E}_\gamma \left[ e^{-\gamma S_{\tau(x)}} \right] = e^{-\gamma x} \mathbb{E}_\gamma \left[ e^{-\gamma \xi(x)} \right],$$

where $\xi(x) = S_{\tau(x)} - x$ is the overshoot.

**Theorem:** Siegmund's algorithm given by

$$Z(x) = e^{-\gamma x} e^{-\gamma \xi(x)}$$

(simulated in a setting where the distribution of the $X$'s are exponentially tilted at rate $\gamma$) has bounded relative error.

Example of this particular change of measure: let $X = U - T$, where $U, T$ are independent exponentially distributed with rate $\delta, \beta$, where $\beta < \delta$. Then, $\gamma$ is found by solving

$$1 = \hat{F}[\gamma] = \mathbb{E}\left[e^{\gamma U}\right]\mathbb{E}\left[e^{-\gamma T}\right] = \frac{\delta}{\delta - \gamma}\frac{\beta}{\beta + \gamma}.$$

This leads to $\gamma = \delta - \beta$.

Example of this particular change of measure: let $X = U - T$, where $U, T$ are independent exponentially distributed with rate $\delta, \beta$, where $\beta < \delta$. Then, $\gamma$ is found by solving

$$1 = \hat{F}[\gamma] = \mathbb{E}\left[e^{\gamma U}\right]\mathbb{E}\left[e^{-\gamma T}\right] = \frac{\delta}{\delta - \gamma}\frac{\beta}{\beta + \gamma}.$$

This leads to $\gamma = \delta - \beta$. Note that

$$\mathbb{E}\left[e^{sX}\right] = \mathbb{E}_\gamma\left[e^{sX}L_{1,\gamma}\right] = \mathbb{E}_\gamma\left[e^{sX}e^{-\gamma X}\right] = \mathbb{E}_\gamma\left[e^{(s-\gamma)X}\right].$$

This implies that

$$\mathbb{E}_\gamma\left[e^{sX}\right] = \mathbb{E}\left[e^{(s+\gamma)X}\right] = \mathbb{E}\left[e^{(s+\gamma)U}\right]\mathbb{E}\left[e^{-(s+\gamma)V}\right] = \frac{\beta}{\beta - s}\frac{\delta}{\delta + s}.$$

Example of this particular change of measure: let $X = U - T$, where $U$, $T$ are independent exponentially distributed with rate $\delta, \beta$, where $\beta < \delta$. Then, $\gamma$ is found by solving

$$1 = \hat{F}[\gamma] = \mathbb{E}\left[e^{\gamma U}\right] \mathbb{E}\left[e^{-\gamma T}\right] = \frac{\delta}{\delta - \gamma} \frac{\beta}{\beta + \gamma}.$$

This leads to $\gamma = \delta - \beta$. Note that

$$\mathbb{E}\left[e^{sX}\right] = \mathbb{E}_\gamma\left[e^{sX} L_{1,\gamma}\right] = \mathbb{E}_\gamma\left[e^{sX} e^{-\gamma X}\right] = \mathbb{E}_\gamma\left[e^{(s-\gamma)X}\right].$$

This implies that

$$\mathbb{E}_\gamma\left[e^{sX}\right] = \mathbb{E}\left[e^{(s+\gamma)X}\right] = \mathbb{E}\left[e^{(s+\gamma)U}\right] \mathbb{E}\left[e^{-(s+\gamma)V}\right] = \frac{\beta}{\beta - s} \frac{\delta}{\delta + s}.$$

This represents the distribution of the difference between two exponential random variables with rates $\beta$ and $\delta$!
**Conclusion**: $z(x)$ can be computed be estimated by interchanging parameters. Applications in M/M/1 queue (see book).

**Theorem:** Siegmund's algorithm given by

$$Z(x) = e^{-\gamma x} e^{-\gamma \xi(x)}$$

(simulated in a setting where the $X$'s are exponentially tilted at rate $\gamma$) has bounded relative error.

**Sketch of proof:**

**Theorem:** Siegmund's algorithm given by

$$Z(x) = e^{-\gamma x} e^{-\gamma \xi(x)}$$

(simulated in a setting where the $X$'s are exponentially tilted at rate $\gamma$) has bounded relative error.

**Sketch of proof:**

- One can reason that under the exponentially tilted measure, $\xi(x)$ is a regenerative process, regenerating each time $S_{\tau(x)}$ has a partial maximum.
- From this, we can derive that $\{\xi(x), x \geq 0\}$ has a stationary distribution, i.e.

  $$\lim_{x \to \infty} \mathbb{E}_\gamma \left[ e^{-\gamma \xi(x)} \right] = \mathbb{E}_\gamma \left[ e^{-\gamma \xi(\infty)} \right] =: C$$

  so that

  $$z(x) = \mathbb{P}\left(\tau(x) < \infty\right) \sim C e^{-\gamma x},$$

  which is called the *Cramér-Lundberg* approximation.

## Rare-Event Simulation

**Proof cntd.:** We furthermore have that

$$\mathbb{E}_\gamma\left[Z^2(x)\right] = e^{-2\gamma x}\mathbb{E}_\gamma\left[e^{-2\gamma\xi(x)}\right] \sim C_1 e^{-2\gamma x},$$

where $C_1 = \mathbb{E}_\gamma\left[e^{-2\gamma\xi(x)}\right]$. Thus,

$$\mathrm{Var}_\gamma\left[Z(x)\right] \sim C_1 e^{-2\gamma x} - C^2(e^{-2\gamma x}) = C_2 e^{-2\gamma x}$$

with $C_2 = C_1 - C^2$. Thus, we have that the relative error satisfies

$$\frac{\mathrm{Var}_\gamma\left[Z(x)\right]}{z(x)} \sim \frac{C_2 e^{-2\gamma x}}{C^2 e^{-2\gamma x}} = C_3 < \infty.$$

**Q:** Is $g_\gamma(x)$ the only importance sampling density which admits a bounded relative error in Siegmund's algorithm?

**Q:** Is $g_\gamma(x)$ the only importance sampling density which admits a bounded relative error in Siegmund's algorithm?

**A:** Yes, actually, it's the only one that even admits logarithmic efficiency!

To see this, suppose that we regard another importance sampling density $\tilde{h}(x)$ such that $\mathbb{E}_{\tilde{h}}[X] > 0$ (why?). Then, we would have

that $z(x) = \mathbb{E}\left[\mathbb{1}_{\{\tau(x)<\infty\}}\right] = \mathbb{E}_{\tilde{h}}\left[L_{\tau(x)}(f|\tilde{h})\right]$, so that

$$Z(x) := L_{\tau(x)}(f|\tilde{h}) = \prod_{i=1}^{\tau(x)} \frac{f(X_i)}{\tilde{h}(X_i)}.$$

**Q:** Is $g_\gamma(x)$ the only importance sampling density which admits a bounded relative error in Siegmund's algorithm?

**A:** Yes, actually, it's the only one that even admits logarithmic efficiency!

To see this, suppose that we regard another importance sampling density $\tilde{h}(x)$ such that $\mathbb{E}_{\tilde{h}}[X] > 0$ (why?). Then, we would have

that $z(x) = \mathbb{E}\left[\mathbb{1}_{\{\tau(x)<\infty\}}\right] = \mathbb{E}_{\tilde{h}}\left[L_{\tau(x)}(f|\tilde{h})\right]$, so that

$$Z(x) := L_{\tau(x)}(f|\tilde{h}) = \prod_{i=1}^{\tau(x)} \frac{f(X_i)}{\tilde{h}(X_i)}.$$

**Theorem:** This importance sampling algorithm is logarithmically efficient if and only if $\tilde{h}(x) = g_{\gamma(x)}$.

**Q:** Is $g_\gamma(x)$ the only importance sampling density which admits a bounded relative error in Siegmund's algorithm?

**A:** Yes, actually, it's the only one that even admits logarithmic efficiency!

To see this, suppose that we regard another importance sampling density $\tilde{h}(x)$ such that $\mathbb{E}_{\tilde{h}}[X] > 0$ (why?). Then, we would have

that $z(x) = \mathbb{E}\left[\mathbb{1}_{\{\tau(x)<\infty\}}\right] = \mathbb{E}_{\tilde{h}}\left[L_{\tau(x)}(f|\tilde{h})\right]$, so that

$$Z(x) := L_{\tau(x)}(f|\tilde{h}) = \prod_{i=1}^{\tau(x)} \frac{f(X_i)}{\tilde{h}(X_i)}.$$

**Theorem:** This importance sampling algorithm is logarithmically efficient if and only if $\tilde{h}(x) = g_{\gamma(x)}$.

**Proof:** We have already shown sufficiency.

**Q:** Is $g_\gamma(x)$ the only importance sampling density which admits a bounded relative error in Siegmund's algorithm?

**A:** Yes, actually, it's the only one that even admits logarithmic efficiency!

To see this, suppose that we regard another importance sampling density $\tilde{h}(x)$ such that $\mathbb{E}_{\tilde{h}}[X] > 0$ (why?). Then, we would have

that $z(x) = \mathbb{E}\left[\mathbb{1}_{\{\tau(x)<\infty\}}\right] = \mathbb{E}_{\tilde{h}}\left[L_{\tau(x)}(f|\tilde{h})\right]$, so that

$$Z(x) := L_{\tau(x)}(f|\tilde{h}) = \prod_{i=1}^{\tau(x)} \frac{f(X_i)}{\tilde{h}(X_i)}.$$

**Theorem:** This importance sampling algorithm is logarithmically efficient if and only if $\tilde{h}(x) = g_{\gamma(x)}$.

**Proof:** We have already shown sufficiency. So we assume now that $\tilde{h}(x) \neq g_{\gamma(x)}$ and show that the algorithm can not be logarithmically efficient in that case.

# Rare-Event Simulation

Suppose $\tilde{h}(x) \neq g_{\gamma(x)}$. Then,

$$\begin{aligned}\mathbb{E}_{\tilde{h}}\left[Z^2(x)\right] &= \mathbb{E}_{\tilde{h}}\left[L^2_{\tau(x)}(f|\tilde{h})\right] = \mathbb{E}_{\tilde{h}}\left[L^2_{\tau(x)}(f|g_\gamma)L^2_{\tau(x)}(g_\gamma|\tilde{h})\right] \\ &= \mathbb{E}_{\gamma}\left[L^2_{\tau(x)}(f|g_\gamma)L_{\tau(x)}(g_\gamma|\tilde{h})\right] = \mathbb{E}_{\gamma}\left[e^{\sum_{i=1}^{\tau(x)} K_i}\right],\end{aligned}$$

## Rare-Event Simulation

Suppose $\tilde{h}(x) \neq g_{\gamma(x)}$. Then,

$$\begin{aligned}\mathbb{E}_{\tilde{h}}\left[Z^2(x)\right] &= \mathbb{E}_{\tilde{h}}\left[L^2_{\tau(x)}(f|\tilde{h})\right] = \mathbb{E}_{\tilde{h}}\left[L^2_{\tau(x)}(f|g_\gamma)L^2_{\tau(x)}(g_\gamma|\tilde{h})\right] \\ &= \mathbb{E}_{\gamma}\left[L^2_{\tau(x)}(f|g_\gamma)L_{\tau(x)}(g_\gamma|\tilde{h})\right] = \mathbb{E}_{\gamma}\left[e^{\sum_{i=1}^{\tau(x)} K_i}\right],\end{aligned}$$

where

$$K_i = \log\left(\left(\frac{f(X_i)}{g_\gamma(X_i)}\right)^2 \frac{g_\gamma(X_i)}{\tilde{h}(X_i)}\right) = -2\gamma X_i - \log\left(\frac{\tilde{h}(X_i)}{g_\gamma(X_i)}\right).$$

## Rare-Event Simulation

Suppose $\tilde{h}(x) \neq g_{\gamma(x)}$. Then,

$$\begin{aligned}\mathbb{E}_{\tilde{h}}\left[Z^2(x)\right] &= \mathbb{E}_{\tilde{h}}\left[L^2_{\tau(x)}(f|\tilde{h})\right] = \mathbb{E}_{\tilde{h}}\left[L^2_{\tau(x)}(f|g_\gamma)L^2_{\tau(x)}(g_\gamma|\tilde{h})\right] \\ &= \mathbb{E}_{\gamma}\left[L^2_{\tau(x)}(f|g_\gamma)L_{\tau(x)}(g_\gamma|\tilde{h})\right] = \mathbb{E}_{\gamma}\left[e^{\sum_{i=1}^{\tau(x)} K_i}\right],\end{aligned}$$

where

$$K_i = \log\left(\left(\frac{f(X_i)}{g_\gamma(X_i)}\right)^2 \frac{g_\gamma(X_i)}{\tilde{h}(X_i)}\right) = -2\gamma X_i - \log\left(\frac{\tilde{h}(X_i)}{g_\gamma(X_i)}\right).$$

Further,

$$\mathbb{E}_\gamma\left[K_i\right] = -2\gamma\mathbb{E}_\gamma\left[X_i\right] + \epsilon',$$

where $\mathbb{E}_\gamma\left[X_i\right] > 0$ and $\epsilon' = -\mathbb{E}_\gamma\left[\log\left(\frac{\tilde{h}(X_1)}{g_\gamma(X_1)}\right)\right]$. Further, we can prove that $\epsilon' > 0$.

Further, due to Jensen's inequality and Wald's equality,

$$\begin{aligned}\mathbb{E}_{\tilde{h}}\left[Z^2(x)\right] &= \mathbb{E}_\gamma\left[e^{\sum_{i=1}^{\tau(x)} K_i}\right] \\ &\geq e^{\mathbb{E}_\gamma\left[\sum_{i=1}^{\tau(x)} K_i\right]} \\ &= e^{\mathbb{E}_\gamma[\tau(x)](\epsilon' - 2\gamma\mathbb{E}_\gamma[X_i])}\end{aligned}$$

Further, due to Jensen's inequality and Wald's equality,

$$\begin{aligned}\mathbb{E}_{\tilde{h}}\left[Z^2(x)\right] &= \mathbb{E}_\gamma\left[e^{\sum_{i=1}^{\tau(x)} K_i}\right] \\ &\geq e^{\mathbb{E}_\gamma\left[\sum_{i=1}^{\tau(x)} K_i\right]} \\ &= e^{\mathbb{E}_\gamma[\tau(x)](\epsilon' - 2\gamma\mathbb{E}_\gamma[X_i])}\end{aligned}$$

Note that $\mathbb{E}_\gamma\left[\frac{\tau(x)}{x}\right] \to \frac{1}{\mathbb{E}_\gamma[X_1]}$ a.s. as $x \to \infty$. Thus, using $z(x) \sim Ce^{-\gamma x}$, we have that for $0 < \epsilon < \frac{\epsilon'}{\gamma\mathbb{E}_\gamma[X_1]}$,

Further, due to Jensen's inequality and Wald's equality,

$$\begin{aligned}\mathbb{E}_{\tilde{h}}\left[Z^2(x)\right] &= \mathbb{E}_\gamma\left[e^{\sum_{i=1}^{\tau(x)} K_i}\right] \\ &\geq e^{\mathbb{E}_\gamma\left[\sum_{i=1}^{\tau(x)} K_i\right]} \\ &= e^{\mathbb{E}_\gamma[\tau(x)](\epsilon' - 2\gamma\mathbb{E}_\gamma[X_i])}\end{aligned}$$

Note that $\mathbb{E}_\gamma\left[\frac{\tau(x)}{x}\right] \to \frac{1}{\mathbb{E}_\gamma[X_1]}$ a.s. as $x \to \infty$. Thus, using $z(x) \sim Ce^{-\gamma x}$, we have that for $0 < \epsilon < \frac{\epsilon'}{\gamma\mathbb{E}_\gamma[X_1]}$,

$$\begin{aligned}\liminf_{x\to\infty} \frac{\mathbb{E}_{\tilde{h}}\left[Z^2(x)\right]}{z(x)^{2-\epsilon}} &= \liminf_{x\to\infty} \frac{\mathbb{E}_{\tilde{h}}\left[Z^2(x)\right]}{C^{2-\epsilon}e^{-2\gamma x+\epsilon\gamma x}} = \liminf_{x\to\infty} \frac{e^{x(\frac{\epsilon'}{\mathbb{E}_\gamma[X_1]} - 2\gamma)}}{C^{2-\epsilon}e^{-2\gamma x+\epsilon\gamma x}} \\ &= \liminf_{x\to\infty} \frac{1}{C^{2-\epsilon}} \frac{e^{\frac{\epsilon' x}{\mathbb{E}_\gamma[x_1]}}}{e^{\epsilon\gamma x}} = \infty.\end{aligned}$$

Further, due to Jensen's inequality and Wald's equality,

$$\begin{aligned}\mathbb{E}_{\tilde{h}}\left[Z^2(x)\right] &= \mathbb{E}_\gamma\left[e^{\sum_{i=1}^{\tau(x)} K_i}\right]\\ &\geq e^{\mathbb{E}_\gamma\left[\sum_{i=1}^{\tau(x)} K_i\right]}\\ &= e^{\mathbb{E}_\gamma[\tau(x)](\epsilon'-2\gamma\mathbb{E}_\gamma[X_i])}\end{aligned}$$

Note that $\mathbb{E}_\gamma\left[\frac{\tau(x)}{x}\right] \to \frac{1}{\mathbb{E}_\gamma[X_1]}$ a.s. as $x \to \infty$. Thus, using $z(x) \sim Ce^{-\gamma x}$, we have that for $0 < \epsilon < \frac{\epsilon'}{\gamma\mathbb{E}_\gamma[X_1]}$,

$$\begin{aligned}\liminf_{x\to\infty} \frac{\mathbb{E}_{\tilde{h}}\left[Z^2(x)\right]}{z(x)^{2-\epsilon}} &= \liminf_{x\to\infty} \frac{\mathbb{E}_{\tilde{h}}\left[Z^2(x)\right]}{C^{2-\epsilon}e^{-2\gamma x+\epsilon\gamma x}} = \liminf_{x\to\infty} \frac{e^{x(\frac{\epsilon'}{\mathbb{E}_\gamma[X_1]}-2\gamma)}}{C^{2-\epsilon}e^{-2\gamma x+\epsilon\gamma x}}\\ &= \liminf_{x\to\infty} \frac{1}{C^{2-\epsilon}}\frac{e^{\frac{\epsilon' x}{\mathbb{E}_\gamma[x_1]}}}{e^{\epsilon\gamma x}} = \infty.\end{aligned}$$

Conclusion: when $\tilde{h} \neq g_\gamma$, we will have no logarithmic efficiency... theorem proved!

## Rare-Event Simulation

Well, not entirely just yet. In passing, we assumed that $\epsilon' = -\mathbb{E}_\gamma \left[ \log \left( \frac{\tilde{h}(X_1)}{g_\gamma(X_1)} \right) \right] > 0$.

## Rare-Event Simulation

Well, not entirely just yet. In passing, we assumed that $\epsilon' = -\mathbb{E}_\gamma\left[\log\left(\frac{\tilde{h}(X_1)}{g_\gamma(X_1)}\right)\right] > 0$.

To see this, note that, again due to Jensen's inequality

$$\begin{aligned}
\mathbb{E}_\gamma\left[\log\left(\frac{\tilde{h}(X_1)}{g_\gamma(X_1)}\right)\right] &< \log \mathbb{E}_\gamma\left[\frac{\tilde{h}(X_1)}{g_\gamma(X_1)}\right] \\
&= \log \int_{x:g_\gamma(x)>0} \frac{\tilde{h}(x)}{g_\gamma(x)} g_\gamma(x)dx \\
&= \log \int_{x:g_\gamma(x)>0} \tilde{h}(x)dx \\
&\leq \log 1 = 0.
\end{aligned}$$

## Rare-Event Simulation

Another problem. Consider again a random walk $S_n := \sum_{i=1}^{n} X_n$, where the summands are independently distributed with common distribution $F$ and mean $\mu$ (sign unimportant this time).

## Rare-Event Simulation

Another problem. Consider again a random walk $S_n := \sum_{i=1}^{n} X_n$, where the summands are independently distributed with common distribution $F$ and mean $\mu$ (sign unimportant this time).

We are now interested in the family of rare events

$$A(n) = \{S_n > n(\mu + \delta)\},$$

where $\delta > 0$.

## Rare-Event Simulation

Another problem. Consider again a random walk $S_n := \sum_{i=1}^{n} X_n$, where the summands are independently distributed with common distribution $F$ and mean $\mu$ (sign unimportant this time).

We are now interested in the family of rare events

$$A(n) = \{S_n > n(\mu + \delta)\},$$

where $\delta > 0$.

Due to the weak law of large numbers we have that

$$z(n) = \mathbb{P}\left(A(n)\right) \to 0$$

The main question: how to simulate the value of $z(n)$ efficiently as $n$ gets large?

## Rare-Event Simulation

To efficiently estimate $z(n) = \mathbb{P}(A(n))$, we again employ exponential change of measure, so that our algorithm becomes

$$Z(n) = e^{-\theta S_n + n\kappa(\theta)} \mathbb{1}_{\{S_n > n(\mu+\delta)\}},$$

where the $X_i$ are sampled from the tilted distribution.

## Rare-Event Simulation

To efficiently estimate $z(n) = \mathbb{P}(A(n))$, we again employ exponential change of measure, so that our algorithm becomes

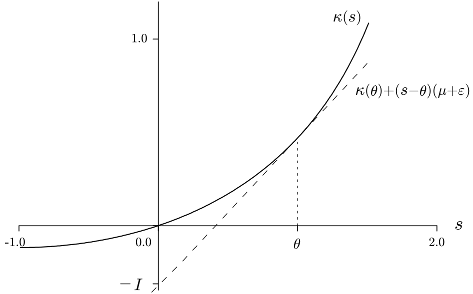
$$Z(n) = e^{-\theta S_n + n\kappa(\theta)} \mathbb{1}_{\{S_n > n(\mu+\delta)\}},$$

where the $X_i$ are sampled from the tilted distribution.
It turns out that $\theta$ should be chosen such that

$$\mathbb{E}_\theta[X] = \kappa'(\theta) = \mu + \delta.$$

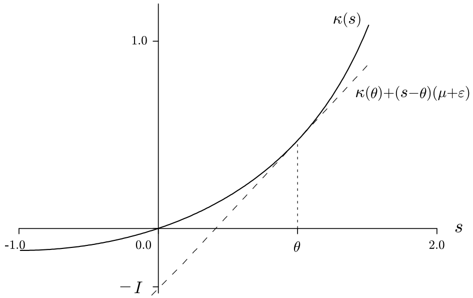
$$\mathbb{E}_\theta[X] = \kappa'(\theta) = \mu + \delta.$$

Under this constraint, $\theta > 0$, since $\kappa'(0) = \mu$ and $\kappa$ is a convex function. Furthermore, we have that $I = \theta(\mu + \delta) - \kappa(\theta) > 0$.



$$\mathbb{E}_\theta\left[X\right] = \kappa'(\theta) = \mu + \delta.$$

Under this constraint, $\theta > 0$, since $\kappa'(0) = \mu$ and $\kappa$ is a convex function. Furthermore, we have that $I = \theta(\mu + \delta) - \kappa(\theta) > 0$.




**Theorem:** The algorithm using the exponentially $(\theta)$ tilted measure is logarithmically efficient, and this measure is the only one to make the algorithm logarithmically efficient.

**Proof:** Structure similar to Siegmund's algorithm.

First, note that, since $\theta > 0$

$$\begin{aligned} z(n) = \mathbb{P}\left(A(n)\right) &= \mathbb{E}_\theta\left[e^{-\theta S_n + n\kappa(\theta)}\mathbb{1}_{\{S_n > n(\mu+\delta)\}}\right] \\ &= e^{-nI}\mathbb{E}_\theta\left[e^{-\theta(S_n - n(\mu+\delta))}\mathbb{1}_{\{S_n > n(\mu+\delta)\}}\right] \\ &\le e^{-nI} \end{aligned}$$

Furthermore,

$$\begin{aligned} \mathrm{Var}_\theta\left[Z(n)\right] \le \mathbb{E}_\theta\left[Z^2(n)\right] &= \mathbb{E}_\theta\left[e^{-2\theta S_n + 2n\kappa(\theta)}\mathbb{1}_{\{S_n > n(\mu+\delta)\}}\right] \\ &= e^{-2nI}\mathbb{E}_\theta\left[e^{-2\theta(S_n - n(\mu+\delta))}\mathbb{1}_{\{S_n > n(\mu+\delta)\}}\right] \\ &\le e^{-2nI}. \end{aligned}$$

Thus, to establish logarithmic efficiency, we will require a lower bound of some sort on the $z(n)$.

Note that, when the density of $X_n$ is exponentially tilted with parameter $\theta$,

$$\frac{S_n - n(\mu + \delta)}{\sqrt{n}} \rightarrow \mathcal{N}(0, \sigma_\theta^2) \text{ and}$$

Thus, to establish logarithmic efficiency, we will require a lower bound of some sort on the $z(n)$.

Note that, when the density of $X_n$ is exponentially tilted with parameter $\theta$,

$$\frac{S_n - n(\mu+\delta)}{\sqrt{n}} \to \mathcal{N}(0, \sigma_\theta^2) \text{ and}$$

$$\lim_{n\to\infty} \mathbb{P}_\theta\left(\frac{S_n - n(\mu+\delta)}{\sqrt{n}} \in (0,1)\right) = \Phi(\frac{1}{\sigma_\theta}) - \Phi(0) =: c > 0.$$

Thus,

Thus, to establish logarithmic efficiency, we will require a lower bound of some sort on the $z(n)$.

Note that, when the density of $X_n$ is exponentially tilted with parameter $\theta$,

$$\frac{S_n - n(\mu+\delta)}{\sqrt{n}} \to \mathcal{N}(0, \sigma_\theta^2) \text{ and}$$

$$\lim_{n\to\infty} \mathbb{P}_\theta\left(\frac{S_n - n(\mu+\delta)}{\sqrt{n}} \in (0,1)\right) = \Phi(\frac{1}{\sigma_\theta}) - \Phi(0) =: c > 0.$$

Thus,

$$\liminf_{n\to\infty}(e^{nI+\theta\sqrt{n}}z(n)) = \liminf_{n\to\infty} e^{\theta\sqrt{n}}\mathbb{E}_\theta\left[e^{-\theta(S_n-n(\mu+\delta))}\mathbb{1}_{\{S_n>n(\mu+\delta)\}}\right]$$

$$\geq \liminf_{n\to\infty} e^{\theta\sqrt{n}}\mathbb{E}_\theta\left[e^{-\theta(S_n-n(\mu+\delta))}\mathbb{1}_{\{\frac{S_n-n(\mu+\delta)}{\sqrt{n}}\in(0,1)\}}\right]$$

$$\geq \liminf_{n\to\infty} e^{\theta\sqrt{n}}e^{-\theta\sqrt{n}}\mathbb{P}_\theta\left(\mathbb{1}_{\{\frac{S_n-n(\mu+\delta)}{\sqrt{n}}\in(0,1)\}}\right) = c > 0.$$

## Rare-Event Simulation

Thus, to show logarithmic efficiency, note that

$$\limsup_{n\to\infty} \frac{\mathrm{Var}_\theta\left[Z(n)\right]}{z(n)^{2-\epsilon}} = \limsup_{n\to\infty} \frac{\mathrm{Var}_\theta\left[Z(n)\right] e^{(2-\epsilon)(nI+\theta\sqrt{n})}}{z(n)^{2-\epsilon} e^{(2-\epsilon)(nI+\theta\sqrt{n})}}$$

$$\leq \frac{\limsup_{n\to\infty} e^{-\epsilon nI+(2-\epsilon)\theta\sqrt{n}}}{c^{2-\epsilon}} = 0.$$

To show that there is no other density which allows logarithmic efficiency, a proof using similar arguments as the one for Siegmund's algorithm can be given.