Estimation of the probability of the union of rare events

Mathis Le Bail

August 19, 2022

Introduction

$X \in \mathbb{R}^d$ a random variable of distribution F_X. We want to estimate $\pi = \mathbb{P}(X \in A)$ where A is the event of interest.

Crude Monte Carlo Estimator with $(X_i) \overset{\text{i.i.d}}{\sim} F_X$

$$\hat{\pi}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_A(X_i) \tag{1}$$

Introduction

$X \in \mathbb{R}^d$ a random variable of distribution F_X. We want to estimate $\pi = \mathbb{P}(X \in A)$ where A is the event of interest.

Crude Monte Carlo Estimator with $(X_i) \overset{\text{i.i.d}}{\sim} F_X$

$$\hat{\pi}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_A(X_i) \tag{1}$$

$$\frac{\sigma_{\pi_n}}{\pi} = \frac{\sqrt{\pi(1-\pi)}}{\sqrt{n}\pi} \underset{\pi \to 0}{\sim} \frac{1}{\sqrt{n\pi}} \underset{\pi \to 0}{\to} \infty \tag{2}$$

Introduction

$X \in \mathbb{R}^d$ a random variable of distribution F_X. We want to estimate $\pi = \mathbb{P}(X \in A)$ where A is the event of interest.

Crude Monte Carlo Estimator with $(X_i) \overset{\text{i.i.d}}{\sim} F_X$

$$\hat{\pi}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_A(X_i) \tag{1}$$

$$\frac{\sigma_{\pi_n}}{\pi} = \frac{\sqrt{\pi(1-\pi)}}{\sqrt{n}\pi} \underset{\pi \to 0}{\sim} \frac{1}{\sqrt{n\pi}} \underset{\pi \to 0}{\to} \infty \tag{2}$$

Sample size N needed to acquire a 10% relative precision for the half-width of the 95% confidence interval

$$\begin{aligned} &\frac{1.96\sqrt{\pi(1-\pi)}}{\pi\sqrt{N}} \approx 0.1 \\ &N \approx \frac{100 \times 1.96^2 \pi(1-\pi)}{\pi^2} \sim \frac{100 \times 1.96^2}{\pi} \\ &\pi \sim 10^{-8} \Rightarrow N \sim 10^{10} \end{aligned} \tag{3}$$

Family of rare events $\{R(\tau)\}$, $\pi(\tau) = \mathbb{P}(R(\tau)) \to 0$ when $\tau \to \infty$.

R$(\tau) = (X > \tau)$ where X is a random variable

Suppose that we have an unbiased estimator Π_τ of $\pi(\tau)$ for each τ.

Family of rare events $\{R(\tau)\}$, $\pi(\tau) = \mathbb{P}(R(\tau)) \to 0$ when $\tau \to \infty$.

R$(\tau) = (X > \tau)$ where X is a random variable

Suppose that we have an unbiased estimator Π_τ of $\pi(\tau)$ for each τ.

Definition

Π_τ is *strongly efficient* if

$$\sup_\tau \frac{Var(\Pi_\tau)}{\pi(\tau)^2} < \infty \tag{4}$$

Efficiency properties in rare-event simulation

Family of rare events $\{R(\tau)\}$, $\pi(\tau) = \mathbb{P}(R(\tau)) \to 0$ when $\tau \to \infty$.

R$(\tau) = (X > \tau)$ where X is a random variable

Suppose that we have an unbiased estimator Π_τ of $\pi(\tau)$ for each τ.

Definition

Π_τ is *strongly efficient* if

$$\sup_\tau \frac{Var(\Pi_\tau)}{\pi(\tau)^2} < \infty \tag{4}$$

Definition

Π_τ is *asymptotically optimal or weakly efficient* if

$$\frac{\log(\mathbb{E}(\Pi_\tau^2))}{2\log(\pi(\tau))} \underset{\tau\to\infty}{\to} 1 \tag{5}$$

$\widehat{\Pi_n(\tau)} = \frac{1}{n}\sum_{i=1}^{n} \Pi_\tau^i$ the average of the i.i.d n random variables $\Pi_\tau^1, ..., \Pi_\tau^n$ to estimate $\pi(\tau)$.

$$\mathbb{P}\left(\left|\frac{\widehat{\Pi_n(\tau)} - \pi(\tau)}{\pi(\tau)}\right| > \epsilon\right) < \delta \tag{6}$$

- Chebyshev's inequality gives us

$$\mathbb{P}\left(\left|\frac{\widehat{\Pi_n(\tau)} - \pi(\tau)}{\pi(\tau)}\right| > \epsilon\right) \leq \frac{Var(\frac{\widehat{\Pi_n(\tau)}}{\pi(\tau)})}{\epsilon^2} = \frac{Var(\Pi_\tau)}{n\pi(\tau)^2\epsilon^2} \tag{7}$$

$\widehat{\Pi_n(\tau)} = \frac{1}{n}\sum_{i=1}^{n}\Pi_\tau^i$ the average of the i.i.d n random variables $\Pi_\tau^1, ..., \Pi_\tau^n$ to estimate $\pi(\tau)$.

$$\mathbb{P}\left(\left|\frac{\widehat{\Pi_n(\tau)} - \pi(\tau)}{\pi(\tau)}\right| > \epsilon\right) < \delta \tag{6}$$

- Chebyshev's inequality gives us

$$\mathbb{P}\left(\left|\frac{\widehat{\Pi_n(\tau)} - \pi(\tau)}{\pi(\tau)}\right| > \epsilon\right) \leq \frac{Var(\frac{\widehat{\Pi_n(\tau)}}{\pi(\tau)})}{\epsilon^2} = \frac{Var(\Pi_\tau)}{n\pi(\tau)^2\epsilon^2} \tag{7}$$

- Strongly efficient : There is a $K \in \mathbb{N}$ such that

$$\frac{Var(\Pi_\tau)}{\pi(\tau)^2} \leq K \Rightarrow \mathbb{P}\left(\left|\frac{\widehat{\Pi_n(\tau)} - \pi(\tau)}{\pi(\tau)}\right| > \epsilon\right) \leq \frac{K}{n\epsilon^2} \Rightarrow n \approx \mathcal{O}(1)$$

$\widehat{\Pi_n(\tau)} = \frac{1}{n}\sum_{i=1}^{n}\Pi_\tau^i$ the average of the i.i.d n random variables $\Pi_\tau^1, ..., \Pi_\tau^n$ to estimate $\pi(\tau)$.

$$\mathbb{P}\left(\left|\frac{\widehat{\Pi_n(\tau)} - \pi(\tau)}{\pi(\tau)}\right| > \epsilon\right) < \delta \tag{6}$$

- Chebyshev's inequality gives us

$$\mathbb{P}\left(\left|\frac{\widehat{\Pi_n(\tau)} - \pi(\tau)}{\pi(\tau)}\right| > \epsilon\right) \leq \frac{Var(\frac{\widehat{\Pi_n(\tau)}}{\pi(\tau)})}{\epsilon^2} = \frac{Var(\Pi_\tau)}{n\pi(\tau)^2\epsilon^2} \tag{7}$$

- Strongly efficient : There is a $K \in \mathbb{N}$ such that

$$\frac{Var(\Pi_\tau)}{\pi(\tau)^2} \leq K \Rightarrow \mathbb{P}\left(\left|\frac{\widehat{\Pi_n(\tau)} - \pi(\tau)}{\pi(\tau)}\right| > \epsilon\right) \leq \frac{K}{n\epsilon^2} \Rightarrow n \approx \mathcal{O}(1)$$

- CMC estimator :

$$\frac{Var(\Pi_\tau)}{n\pi(\tau)^2\epsilon^2} = \frac{1-\pi(\tau)}{n\pi(\tau)\epsilon^2} \approx \frac{1}{n\pi(\tau)\epsilon^2} \Rightarrow n \approx \mathcal{O}(\frac{1}{\pi(\tau)})$$

Importance sampling

$$\pi = \int \mathbb{1}_A(x)\, f(x)\mathrm{d}x = \int \mathbb{1}_A(x) \frac{f(x)}{q(x)} q(x)\, \mathrm{d}x = \mathbb{E}_q \left[\mathbb{1}_A(X) \frac{f(X)}{q(X)} \right] \tag{8}$$

Importance sampling

$$\pi = \int \mathbb{1}_A(x)\, f(x)\mathrm{d}x = \int \mathbb{1}_A(x) \frac{f(x)}{q(x)} q(x)\, \mathrm{d}x = \mathbb{E}_q \left[\mathbb{1}_A(X) \frac{f(X)}{q(X)} \right] \quad (8)$$

- q must be such that $q(x) > 0$ for all x where $\mathbb{1}_A(x) f(x) \neq 0$

$$\Pi_n^{IS} = \frac{1}{n} \sum_{i=1}^{n} \frac{f(X_i)}{q(X_i)} \mathbb{1}_A(X_i) \quad (9)$$

where this time $X_i \overset{i.i.d}{\sim} q$.

Importance sampling

$$\pi = \int \mathbb{1}_A(x)\, f(x) \mathrm{d}x = \int \mathbb{1}_A(x) \frac{f(x)}{q(x)} q(x)\, \mathrm{d}x = \mathbb{E}_q \left[\mathbb{1}_A(X) \frac{f(X)}{q(X)} \right] \tag{8}$$

- q must be such that $q(x) > 0$ for all x where $\mathbb{1}_A(x) f(x) \neq 0$

$$\Pi_n^{IS} = \frac{1}{n} \sum_{i=1}^{n} \frac{f(X_i)}{q(X_i)} \mathbb{1}_A(X_i) \tag{9}$$

where this time $X_i \overset{i.i.d}{\sim} q$.

- $$Var(\hat{\Pi_n^{IS}}) = \frac{1}{n} \int \frac{(\mathbb{1}_A(x)\, f(x) - \pi q(x))^2}{q(x)}\, \mathrm{d}x \tag{10}$$

$$q^*(x) = \frac{\mathbb{1}_A(x) f(x)}{\pi} \tag{11}$$

Importance sampling

$$\pi = \int \mathbb{1}_A(x)\, f(x)\mathrm{d}x = \int \mathbb{1}_A(x) \frac{f(x)}{q(x)} q(x)\, \mathrm{d}x = \mathbb{E}_q \left[\mathbb{1}_A(X) \frac{f(X)}{q(X)} \right] \quad (8)$$

- q must be such that $q(x) > 0$ for all x where $\mathbb{1}_A(x) f(x) \neq 0$

$$\Pi_n^{IS} = \frac{1}{n} \sum_{i=1}^{n} \frac{f(X_i)}{q(X_i)} \mathbb{1}_A(X_i) \quad (9)$$

  where this time $X_i \overset{i.i.d}{\sim} q$.

- $$Var(\hat{\Pi_n^{IS}}) = \frac{1}{n} \int \frac{(\mathbb{1}_A(x)\, f(x) - \pi q(x))^2}{q(x)} \, \mathrm{d}x \quad (10)$$

$$q^*(x) = \frac{\mathbb{1}_A(x) f(x)}{\pi} \quad (11)$$

- We note $\mathbb{P}$ the probability associated to the pdf f and $\mathbb{P}^*$ the probability associated to the pdf q^*, we have $\mathbb{P}^*(.) = \mathbb{P}(.|A)$ and $\mathbb{P}^*(dx) = \frac{\mathbb{1}_A(x)}{\mathbb{P}(A)} \mathbb{P}(dx)$

Estimating the probability of the union of rare events

- X a d-dimensional random vector and J fixed integer
- fixed unit vector $\omega_j \in \mathbb{R}^d$ and $\tau_j \in \mathbb{R}$ for $j \in \{1, ..., J\}$
- Rare events : $H_j = \left\{x : x^\top \omega_j \geqslant \tau_j\right\}$
- $H = \cup_{j=1}^{J} H_j$

Estimating the probability of the union of rare events

- X a d-dimensional random vector and J fixed integer
- fixed unit vector $\omega_j \in \mathbb{R}^d$ and $\tau_j \in \mathbb{R}$ for $j \in \{1, \ldots, J\}$
- Rare events : $H_j = \left\{x : x^\top \omega_j \geqslant \tau_j\right\}$
- $H = \cup_{j=1}^{J} H_j$

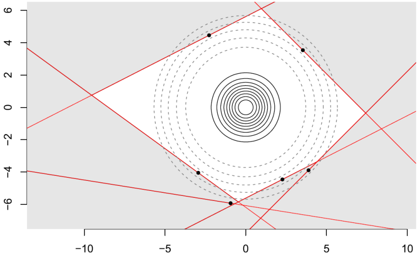


Figure: Graphical representation of a standard bivariate Gaussian distribution with J = 6 half-spaces which are the rare events

Distribution restriction : $X \sim \mathcal{N}(\eta, \Sigma)$

Distribution restriction : $X \sim \mathcal{N}(\eta, \Sigma)$
If we have $X \sim \mathcal{N}(0, I)$ then :

- $\omega_j^\top X \sim \mathcal{N}(0, \omega_j^\top \mathbf{I} \omega_j) = \mathcal{N}(0, \omega_j^\top \omega_j) = \mathcal{N}(0, 1)$

Distribution restriction : $X \sim \mathcal{N}(\eta, \Sigma)$
If we have $X \sim \mathcal{N}(0, I)$ then :

- $\omega_j^\top X \sim \mathcal{N}(0, \omega_j^\top \mathbf{I} \omega_j) = \mathcal{N}(0, \omega_j^\top \omega_j) = \mathcal{N}(0, 1)$
- $P_j = \mathrm{P}\left(X \in H_j\right) = P(\omega_j^\top X \geq \tau_j) = P(\omega_j^\top X \leq -\tau_j) = \Phi(-\tau_j)$

Distribution restriction : $X \sim \mathcal{N}(\eta, \Sigma)$
If we have $X \sim \mathcal{N}(0, I)$ then :

- $\omega_j^\top X \sim \mathcal{N}(0, \omega_j^\top \mathbf{I} \omega_j) = \mathcal{N}(0, \omega_j^\top \omega_j) = \mathcal{N}(0, 1)$
- $P_j = \mathrm{P}\left(X \in H_j\right) = P(\omega_j^\top X \geq \tau_j) = P(\omega_j^\top X \leq -\tau_j) = \Phi(-\tau_j)$
- $\max_{1 \leqslant j \leqslant J} P_j =: \underline{\mu} \leqslant \mu \leqslant \bar{\mu} := \sum_{j=1}^{J} P_j$

Distribution restriction : $X \sim \mathcal{N}(\eta, \Sigma)$
If we have $X \sim \mathcal{N}(0, I)$ then :

- $\omega_j^\top X \sim \mathcal{N}(0, \omega_j^\top \mathbf{I} \omega_j) = \mathcal{N}(0, \omega_j^\top \omega_j) = \mathcal{N}(0, 1)$
- $P_j = \mathrm{P}\left(X \in H_j\right) = P(\omega_j^\top X \geq \tau_j) = P(\omega_j^\top X \leq -\tau_j) = \Phi(-\tau_j)$
- $\max_{1 \leqslant j \leqslant J} P_j =: \underline{\mu} \leqslant \mu \leqslant \bar{\mu} := \sum_{j=1}^J P_j$

Some notations :

- $H_j(x) = \mathbb{1}_{H_j}(x)$
- $S(x) = \sum_{j=1}^J H_j(x)$ counts the number of rare events
- $T_s = \Pr(S = s)$ gives the distribution of S.

The mixture components are the conditional distributions $q_j = (X \mid \omega_j^\top X \geqslant \tau_j)$. We have $q_j(x) = p(x)H_j(x)/P_j$. And the mixture distribution resulting is :

$$q_\alpha^* = \sum_{j=1}^{J} \alpha_j^* q_j \tag{12}$$

with $\alpha_j^* \equiv \frac{P_j}{\sum P_j}$

The mixture components are the conditional distributions $q_j = (X \mid \omega_j^\top X \geqslant \tau_j)$. We have $q_j(x) = p(x)H_j(x)/P_j$. And the mixture distribution resulting is :

$$q_\alpha^* = \sum_{j=1}^{J} \alpha_j^* q_j \tag{12}$$

with $\alpha_j^* \equiv \frac{P_j}{\sum P_j}$

Theorem

If $1 \leq J < \infty$ *and* $\min_j P_j > 0$ *and* $n \geq 1$, *then*

$$\hat{\mu}_{\alpha^*} = \frac{\bar{\mu}}{n} \sum_{i=1}^{n} \frac{1}{S(X_i)} \tag{13}$$

where $X_i \sim q_\alpha^*$, *satisfies* $\mathbb{P}(\frac{\bar{\mu}}{J} \leq \hat{\mu}_{\alpha^*} \leq \bar{\mu}) = 1$,

$$\mathbb{E}\left(\hat{\mu}_{\alpha^*}\right) = \mu \tag{14}$$

and

$$\mathrm{Var}\left(\hat{\mu}_{\alpha^*}\right) \leq \frac{1}{n}\left(\bar{\mu} \sum_{s=1}^{J} \frac{T_s}{s} - \mu^2\right) \leqslant \frac{\mu(\bar{\mu} - \mu)}{n} \tag{15}$$

- $Y \sim \mathcal{N}(\eta, \Sigma)$ and the half-spaces are defined by $\gamma_j^\top Y \geq \kappa_j$.

- $Y \sim \mathcal{N}(\eta, \Sigma)$ and the half-spaces are defined by $\gamma_j^\top Y \geq \kappa_j$.
- $X = \Sigma^{-\frac{1}{2}}(Y - \eta) \sim \mathcal{N}(0, \Sigma^{-\frac{1}{2}}\Sigma\Sigma^{-\frac{1}{2}}) = \mathcal{N}(0, I)$ and $Y = \eta + \Sigma^{\frac{1}{2}}X$.

- $Y \sim \mathcal{N}(\eta, \Sigma)$ and the half-spaces are defined by $\gamma_j^\top Y \geq \kappa_j$.
- $X = \Sigma^{-\frac{1}{2}}(Y - \eta) \sim \mathcal{N}(0, \Sigma^{-\frac{1}{2}}\Sigma\Sigma^{-\frac{1}{2}}) = \mathcal{N}(0, I)$ and $Y = \eta + \Sigma^{\frac{1}{2}}X$.
- We take $\omega_j^\top = \frac{\gamma_j^\top \Sigma^{\frac{1}{2}}}{\sqrt{\gamma_j^\top \Sigma \gamma_j}}$ and $\tau_j = \frac{\kappa_j - \gamma_j^\top \eta}{\sqrt{\gamma_j^\top \Sigma \gamma_j}}$.

$$\begin{aligned} &\gamma_j^\top y \geq \kappa_j \\ \Leftrightarrow\, &\omega_j^\top x \geq \tau_j \end{aligned} \tag{16}$$

- $Y \sim \mathcal{N}(\eta, \Sigma)$ and the half-spaces are defined by $\gamma_j^\top Y \geq \kappa_j$.
- $X = \Sigma^{-\frac{1}{2}}(Y - \eta) \sim \mathcal{N}(0, \Sigma^{-\frac{1}{2}}\Sigma\Sigma^{-\frac{1}{2}}) = \mathcal{N}(0, I)$ and $Y = \eta + \Sigma^{\frac{1}{2}}X$.
- We take $\omega_j^\top = \frac{\gamma_j^\top \Sigma^{\frac{1}{2}}}{\sqrt{\gamma_j^\top \Sigma \gamma_j}}$ and $\tau_j = \frac{\kappa_j - \gamma_j^\top \eta}{\sqrt{\gamma_j^\top \Sigma \gamma_j}}$.

$$\begin{aligned} \gamma_j^\top y &\geq \kappa_j \\ \Leftrightarrow \omega_j^\top x &\geq \tau_j \end{aligned} \tag{16}$$

Sampling method : To sample $X_i = x_i$ from the q_α^* distribution, we have to be able to sample $X \sim \mathcal{N}(0, I)$ conditionally on $X^\top \omega \geqslant \tau$ for a unit vector ω and scalar τ

- $Y \sim \mathcal{N}(\eta, \Sigma)$ and the half-spaces are defined by $\gamma_j^\top Y \geq \kappa_j$.
- $X = \Sigma^{-\frac{1}{2}}(Y - \eta) \sim \mathcal{N}(0, \Sigma^{-\frac{1}{2}}\Sigma\Sigma^{-\frac{1}{2}}) = \mathcal{N}(0, I)$ and $Y = \eta + \Sigma^{\frac{1}{2}}X$.
- We take $\omega_j^\top = \frac{\gamma_j^\top \Sigma^{\frac{1}{2}}}{\sqrt{\gamma_j^\top \Sigma \gamma_j}}$ and $\tau_j = \frac{\kappa_j - \gamma_j^\top \eta}{\sqrt{\gamma_j^\top \Sigma \gamma_j}}$.

$$\begin{aligned} \gamma_j^\top y &\geq \kappa_j \\ \Leftrightarrow \omega_j^\top x &\geq \tau_j \end{aligned} \tag{16}$$

Sampling method : To sample $X_i = x_i$ from the q_α^* distribution, we have to be able to sample $X \sim \mathcal{N}(0, I)$ conditionally on $X^\top \omega \geqslant \tau$ for a unit vector ω and scalar τ
1) Sample $Z \sim \mathcal{N}(0, I)$.
2) Sample $U \sim \mathbb{U}(0, 1)$.

- $Y \sim \mathcal{N}(\eta, \Sigma)$ and the half-spaces are defined by $\gamma_j^\top Y \geq \kappa_j$.
- $X = \Sigma^{-\frac{1}{2}}(Y - \eta) \sim \mathcal{N}(0, \Sigma^{-\frac{1}{2}}\Sigma\Sigma^{-\frac{1}{2}}) = \mathcal{N}(0, I)$ and $Y = \eta + \Sigma^{\frac{1}{2}} X$.
- We take $\omega_j^\top = \frac{\gamma_j^\top \Sigma^{\frac{1}{2}}}{\sqrt{\gamma_j^\top \Sigma \gamma_j}}$ and $\tau_j = \frac{\kappa_j - \gamma_j^\top \eta}{\sqrt{\gamma_j^\top \Sigma \gamma_j}}$.

$$\begin{aligned} \gamma_j^\top y &\geq \kappa_j \\ \Leftrightarrow \omega_j^\top x &\geq \tau_j \end{aligned} \tag{16}$$

Sampling method : To sample $X_i = x_i$ from the q_α^* distribution, we have to be able to sample $X \sim \mathcal{N}(0, I)$ conditionally on $X^\top \omega \geqslant \tau$ for a unit vector ω and scalar τ

1) Sample $Z \sim \mathcal{N}(0, I)$.
2) Sample $U \sim \mathbb{U}(0, 1)$.
3) Let $Y = \Phi^{-1}(\Phi(\tau) + U(1 - \Phi(\tau)))$.
4) Deliver $X = \omega Y + \left(I - \omega\omega^\top\right) Z$.

- $Y \sim \mathcal{N}(\eta, \Sigma)$ and the half-spaces are defined by $\gamma_j^\top Y \geq \kappa_j$.
- $X = \Sigma^{-\frac{1}{2}}(Y - \eta) \sim \mathcal{N}(0, \Sigma^{-\frac{1}{2}}\Sigma\Sigma^{-\frac{1}{2}}) = \mathcal{N}(0, I)$ and $Y = \eta + \Sigma^{\frac{1}{2}}X$.
- We take $\omega_j^\top = \frac{\gamma_j^\top \Sigma^{\frac{1}{2}}}{\sqrt{\gamma_j^\top \Sigma \gamma_j}}$ and $\tau_j = \frac{\kappa_j - \gamma_j^\top \eta}{\sqrt{\gamma_j^\top \Sigma \gamma_j}}$.

$$\begin{aligned} &\gamma_j^\top y \geq \kappa_j \\ \Leftrightarrow\ &\omega_j^\top x \geq \tau_j \end{aligned} \tag{16}$$

Sampling method : To sample $X_i = x_i$ from the q_α^* distribution, we have to be able to sample $X \sim \mathcal{N}(0, I)$ conditionally on $X^\top \omega \geqslant \tau$ for a unit vector ω and scalar τ

1) Sample $Z \sim \mathcal{N}(0, I)$.
2) Sample $U \sim \mathbb{U}(0, 1)$.
3) Let $Y = \Phi^{-1}(\Phi(\tau) + U(1 - \Phi(\tau)))$.
4) Deliver $X = \omega Y + \left(I - \omega\omega^\top\right) Z$.

$$Z = \omega\omega^\top Z + (I - \omega\omega^\top)Z \tag{17}$$

- $Y \sim \mathcal{N}(\eta, \Sigma)$ and the half-spaces are defined by $\gamma_j^\top Y \geq \kappa_j$.
- $X = \Sigma^{-\frac{1}{2}}(Y - \eta) \sim \mathcal{N}(0, \Sigma^{-\frac{1}{2}} \Sigma \Sigma^{-\frac{1}{2}}) = \mathcal{N}(0, I)$ and $Y = \eta + \Sigma^{\frac{1}{2}} X$.
- We take $\omega_j^\top = \frac{\gamma_j^\top \Sigma^{\frac{1}{2}}}{\sqrt{\gamma_j^\top \Sigma \gamma_j}}$ and $\tau_j = \frac{\kappa_j - \gamma_j^\top \eta}{\sqrt{\gamma_j^\top \Sigma \gamma_j}}$.

$$\begin{gathered} \gamma_j^\top y \geq \kappa_j \\ \Leftrightarrow \omega_j^\top x \geq \tau_j \end{gathered} \tag{16}$$

Sampling method : To sample $X_i = x_i$ from the q_α^* distribution, we have to be able to sample $X \sim \mathcal{N}(0, I)$ conditionally on $X^\top \omega \geqslant \tau$ for a unit vector ω and scalar τ

1) Sample $Z \sim \mathcal{N}(0, I)$.
2) Sample $U \sim \mathbb{U}(0, 1)$.
3) Let $Y = \Phi^{-1}(\Phi(\tau) + U(1 - \Phi(\tau)))$.
4) Deliver $X = \omega Y + \left(I - \omega\omega^\top\right) Z$.

$$Z = \omega\omega^\top Z + (I - \omega\omega^\top)Z \tag{17}$$

In practice, $Y = \Phi^{-1}(U\Phi(-\tau))$ and we deliver $X = -X$. We get better numerical stability in sampling $Z \sim \mathcal{N}(0, I)$ conditionally on $X^\top \omega \leq -\tau$

Efficiency of the ALOE algorithm

Corollary

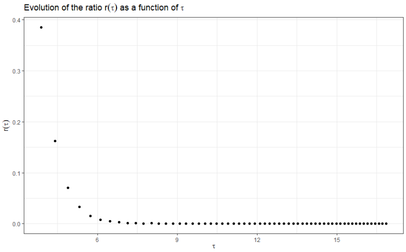
Let $\hat{\mu}_{\alpha^*} = \frac{\bar{\mu}}{n}\sum_{i=1}^{n}\frac{1}{S(X_i)}$ *where* $X_i \sim q_\alpha^*$ *with* q_α^* *defined in (12).*
Then $Var(\hat{\mu}_{\alpha^*}) \leq \frac{\bar{\mu}^2}{4n}$. *If* $\underline{\mu} \geq \frac{\bar{\mu}}{2}$ *then also* $Var(\hat{\mu}_{\alpha^*}) \leq \frac{\underline{\mu}(\bar{\mu}-\underline{\mu})}{n}$. *Similarly,*

$$\frac{Var(\hat{\mu}_{\alpha^*})}{\mu^2} \leq \frac{1}{n}\min\left\{\frac{\bar{\mu}}{\underline{\mu}} - 1, J-1\right\} \leq \frac{J-1}{n} \tag{18}$$

This corollary gives us the property of strong efficiency for (13) since it implies in particular for J fixed :

$$\lim_{\mu\to 0}\frac{Var(\hat{\mu}_{\alpha^*})}{\mu^2} < \infty \tag{19}$$

$$\tau_j = \tau \quad P_j = \Phi(-\tau) \quad \bar{\mu} = \sum P_j = J\Phi(-\tau) \quad r(\tau) = \frac{Var(\hat{\mu}_{\alpha^*})}{\mu^2}$$



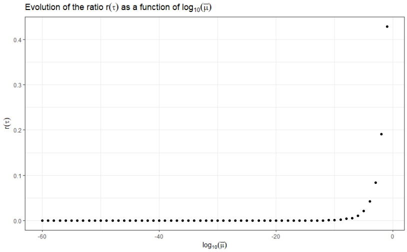


Figure: Results of the computation of the ratios $r(\tau)$ where the first and the second moments have been estimated from 1000 samples

Union of rare events where boundary lines are defined by the general equation $h(x) = \tau$

- New rare events : $H_j = \{x : h_j(x) \geqslant \tau_j\}$

Union of rare events where boundary lines are defined by the general equation $h(x) = \tau$

- New rare events : $H_j = \{x : h_j(x) \geqslant \tau_j\}$
- Mutli-level splitting method : We want to estimate $p = \mathbb{P}(h(X) \geq \tau)$

$$-\infty = L_0 < L_1 < \ldots < L_m = \tau$$

$$p = \mathbb{P}(h(X) \geq \tau) = \prod_{k=1}^{m} \mathbb{P}(h(X) > L_k | h(X) > L_{k-1}) \tag{20}$$

Union of rare events where boundary lines are defined by the general equation $h(x) = \tau$

- New rare events : $H_j = \{x : h_j(x) \geqslant \tau_j\}$
- Mutli-level splitting method : We want to estimate $p = \mathbb{P}(h(X) \geq \tau)$

$$-\infty = L_0 < L_1 < ... < L_m = \tau$$

$$p = \mathbb{P}(h(X) \geq \tau) = \prod_{k=1}^{m} \mathbb{P}(h(X) > L_k | h(X) > L_{k-1}) \tag{20}$$

- i.i.d. sample $(X_1, ..., X_N)$ from the distribution π, $X_1^1 = X_1, ..., X_N^1 = X_N$. For each k,

$$L_k = \min(h(X_1^k), ..., h(X_N^k)) \tag{21}$$

$$X_i^{k+1} = \begin{cases} X_i^k & \text{if } h(X_i^k) > L_k \\ X^* \sim (X|h(X) > L_k) & \text{if } h(X_i^k) = L_k \end{cases} \tag{22}$$

A new i.i.d sample $(X_1^{k+1}, ..., X_N^{k+1})$ from the distribution π_k : $(X|h(X) > L_k)$

Union of rare events where boundary lines are defined by the general equation $h(x) = \tau$

- New rare events : $H_j = \{x : h_j(x) \geqslant \tau_j\}$
- Mutli-level splitting method : We want to estimate $p = \mathbb{P}(h(X) \geq \tau)$

$$-\infty = L_0 < L_1 < ... < L_m = \tau$$

$$p = \mathbb{P}(h(X) \geq \tau) = \prod_{k=1}^{m} \mathbb{P}(h(X) > L_k | h(X) > L_{k-1}) \tag{20}$$

- i.i.d. sample $(X_1, ..., X_N)$ from the distribution π, $X_1^1 = X_1, ..., X_N^1 = X_N$. For each k,

$$L_k = \min(h(X_1^k), ..., h(X_N^k)) \tag{21}$$

$$X_i^{k+1} = \begin{cases} X_i^k & \text{if } h(X_i^k) > L_k \\ X^* \sim (X | h(X) > L_k) & \text{if } h(X_i^k) = L_k \end{cases} \tag{22}$$

  A new i.i.d sample $(X_1^{k+1}, ..., X_N^{k+1})$ from the distribution π_k : $(X | h(X) > L_k)$
- Computation cost : $\mathcal{O}(-N \log(N) \log(p))$

Properties

$$\hat{p} = \left(1 - \frac{1}{N}\right)^M \text{ with } M = \max\{k : L_k \leq \tau\} \tag{23}$$

Properties

$$\hat{p} = \left(1 - \frac{1}{N}\right)^M \text{ with } M = \max\{k : L_k \leq \tau\} \tag{23}$$

Proposition

The estimator $\hat{p}$ is a discrete random variable taking values in

$$\left\{1, \left(1 - \frac{1}{N}\right), \left(1 - \frac{1}{N}\right)^2, \ldots\right\}$$

with probability

$$\mathbb{P}\left(\hat{p} = \left(1 - \frac{1}{N}\right)^k\right) = \frac{p^N(-N\log(p))^k}{k!}, \quad k = 0, 1, 2, \ldots \tag{24}$$

It follows that

$$\mathbb{E}[\hat{p}] = p \quad \textit{and} \quad Var(\hat{p}) = p^2(p^{-\frac{1}{N}} - 1) \tag{25}$$

Properties

$$\hat{p} = \left(1 - \frac{1}{N}\right)^M \text{ with } M = \max\{k : L_k \leq \tau\} \qquad (23)$$

Proposition

The estimator $\hat{p}$ is a discrete random variable taking values in

$$\left\{1, \left(1 - \frac{1}{N}\right), \left(1 - \frac{1}{N}\right)^2, \ldots\right\}$$

with probability

$$\mathbb{P}\left(\hat{p} = \left(1 - \frac{1}{N}\right)^k\right) = \frac{p^N(-N\log(p))^k}{k!}, \quad k = 0, 1, 2, \ldots \qquad (24)$$

It follows that

$$\mathbb{E}[\hat{p}] = p \quad \textit{and} \quad Var(\hat{p}) = p^2(p^{-\frac{1}{N}} - 1) \qquad (25)$$

$$CV^2_{CMC} = \frac{1}{Np} \quad \text{versus} \quad CV^2 = \frac{Var(\hat{p})}{p^2} = (p^{-\frac{1}{N}} - 1) \approx \frac{-\log(p)}{N} \qquad (26)$$

New estimator ALOE

- $$\hat{q}_{\alpha^*|\hat{P}_1,\ldots,\hat{P}_J}(x) \approx \sum_{j=1}^{J} \frac{\hat{P}_j}{\sum_k \hat{P}_k} q_j(x) = \sum_{j=1}^{J} \frac{\hat{P}_j}{\sum_k \hat{P}_k} \frac{p(x)H_j(x)}{P_j} \qquad (27)$$

New estimator ALOE

- $$\hat{q}_{\alpha^*|\hat{P}_1,\dots,\hat{P}_J}(x) \approx \sum_{j=1}^{J} \frac{\hat{P}_j}{\sum_k \hat{P}_k} q_j(x) = \sum_{j=1}^{J} \frac{\hat{P}_j}{\sum_k \hat{P}_k} \frac{p(x)H_j(x)}{P_j} \tag{27}$$
- $$\hat{\mu}_{\alpha^*} = \frac{1}{n} \sum_{i=1}^{n} \frac{p(x_i)H_{1:J}(x_i)}{\sum_{j=1}^{J} \frac{\hat{P}_j}{\sum_k \hat{P}_k} \frac{p(x_i)H_j(x_i)}{P_j}} = \frac{\sum_k \hat{P}_k}{n} \sum_{i=1}^{n} \frac{1}{\sum_{j=1}^{J} \frac{\hat{P}_j}{P_j} H_j(x_i)} \tag{28}$$

New estimator ALOE

- $$\hat{q}_{\alpha^*|\hat{P}_1,\dots,\hat{P}_J}(x) \approx \sum_{j=1}^{J} \frac{\hat{P}_j}{\sum_k \hat{P}_k} q_j(x) = \sum_{j=1}^{J} \frac{\hat{P}_j}{\sum_k \hat{P}_k} \frac{p(x)H_j(x)}{P_j} \tag{27}$$

- $$\hat{\mu}_{\alpha^*} = \frac{1}{n} \sum_{i=1}^{n} \frac{p(x_i)H_{1:J}(x_i)}{\sum_{j=1}^{J} \frac{\hat{P}_j}{\sum_k \hat{P}_k} \frac{p(x_i)H_j(x_i)}{P_j}} = \frac{\sum_k \hat{P}_k}{n} \sum_{i=1}^{n} \frac{1}{\sum_{j=1}^{J} \frac{\hat{P}_j}{P_j} H_j(x_i)} \tag{28}$$

Proposition

Let $\hat{\mu}_{\alpha^*}$ given by (28). Then

$$\mathbb{E}[\hat{\mu}_{\alpha^*}] \approx \mu \tag{29}$$

and

$$Var(\hat{\mu}_{\alpha^*}) \approx \frac{1}{n}\mathbb{E}\left[\sum_k \hat{P}_k \int_H \frac{1}{\sum_{j=1}^{J} \frac{\hat{P}_j}{P_j} H_j(x)} p(x)\,\mathrm{d}x\right] - \frac{\mu^2}{n} \tag{30}$$

- If there is a ϵ such that $\frac{\hat{P}_j}{P_j} \geq \epsilon$ for all $j \Rightarrow Var(\hat{\mu}_{\alpha^*}) \leq \frac{\mu}{n}\left(\frac{\bar{\mu}}{\epsilon} - \mu\right)$

- If there is a ϵ such that $\frac{\hat{P}_j}{P_j} \geq \epsilon$ for all $j \Rightarrow Var(\hat{\mu}_{\alpha^*}) \leq \frac{\mu}{n}\left(\frac{\bar{\mu}}{\epsilon} - \mu\right)$
- $$\hat{P}_j = (1 - \frac{1}{N})^{M_j} \text{ with } M_j = \min(\{k : L_k \leq \tau_j\}, I^j_{\max}) \tag{31}$$

$$\frac{\hat{P}_j}{P_j} \geq \frac{(1 - \frac{1}{N})^{I^j_{max}}}{P_j} \geq \min_j \frac{(1 - \frac{1}{N})^{I^j_{max}}}{P_j} = \epsilon \tag{32}$$

Direct estimation of μ only with the multilevel splitting method

$$\Psi(x) = \max_j \frac{h_j(x)}{\tau_j} \Rightarrow \mathbb{P}(\Psi(X) \geq 1) = \mu \tag{33}$$

Direct estimation of μ only with the multilevel splitting method

$$\Psi(x) = \max_j \frac{h_j(x)}{\tau_j} \Rightarrow \mathbb{P}(\Psi(X) \geq 1) = \mu \tag{33}$$

- N : number of particles used to compute each P_j
- kN : number of particles used in the only run with Ψ

Direct estimation of μ only with the multilevel splitting method

$$\Psi(x) = \max_j \frac{h_j(x)}{\tau_j} \Rightarrow \mathbb{P}(\Psi(X) \geq 1) = \mu \tag{33}$$

- N : number of particles used to compute each P_j
- kN : number of particles used in the only run with Ψ

Variance of the multilevel splitting method : $\mu^2(\mu^{-\frac{1}{kN}} - 1)$

Direct estimation of μ only with the multilevel splitting method

$$\Psi(x) = \max_j \frac{h_j(x)}{\tau_j} \Rightarrow \mathbb{P}(\Psi(X) \geq 1) = \mu \tag{33}$$

- N : number of particles used to compute each P_j
- kN : number of particles used in the only run with Ψ

Variance of the multilevel splitting method : $\mu^2(\mu^{-\frac{1}{kN}} - 1)$

- Variance ratio :

$$\frac{\frac{\mu}{n}\left(\frac{\bar{\mu}}{\epsilon} - \mu\right)}{\mu^2(\mu^{-\frac{1}{kN}} - 1)} = \frac{1}{n}\frac{\frac{1}{\epsilon}\frac{\bar{\mu}}{\mu} - 1}{\mu^{-\frac{1}{kN}} - 1} \underset{N \gg 1}{\approx} \frac{1}{n}\frac{\frac{1}{\epsilon}\frac{\bar{\mu}}{\mu} - 1}{\frac{-\log(\mu)}{kN}} = \frac{kN}{n}\frac{\frac{1}{\epsilon}\frac{\bar{\mu}}{\mu} - 1}{-\log(\mu)} \tag{34}$$

Direct estimation of μ only with the multilevel splitting method

$$\Psi(x) = \max_j \frac{h_j(x)}{\tau_j} \Rightarrow \mathbb{P}(\Psi(X) \geq 1) = \mu \tag{33}$$

- N : number of particles used to compute each P_j
- kN : number of particles used in the only run with Ψ

Variance of the multilevel splitting method : $\mu^2(\mu^{-\frac{1}{kN}} - 1)$

- Variance ratio :

$$\frac{\frac{\mu}{n}\left(\frac{\bar{\mu}}{\epsilon} - \mu\right)}{\mu^2(\mu^{-\frac{1}{kN}} - 1)} = \frac{1}{n}\frac{\frac{1}{\epsilon}\frac{\bar{\mu}}{\mu} - 1}{\mu^{-\frac{1}{kN}} - 1} \underset{N \gg 1}{\approx} \frac{1}{n}\frac{\frac{1}{\epsilon}\frac{\bar{\mu}}{\mu} - 1}{\frac{-\log(\mu)}{kN}} = \frac{kN}{n}\frac{\frac{1}{\epsilon}\frac{\bar{\mu}}{\mu} - 1}{-\log(\mu)} \tag{34}$$

- Computation cost :

$$\frac{-N\log(N)\log(P)J}{-kN\log(kN)log(\mu)J} = \frac{\log(N)\log(P)}{k\log(kN)\log(JP)} = \frac{1}{k}\frac{1}{1 + \frac{\log(k)}{\log(N)}}\frac{1}{1 + \frac{\log(J)}{\log(P)}} \tag{35}$$

Numerical examples - Circumscribed Polygon

$\mathcal{P}(J,\tau) \subset \mathbb{R}^2$ a regular polygon of $J \geq 3$ sides circumscribed around the circle of radius $\tau > 0$

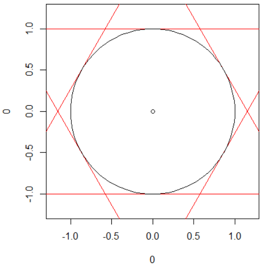


Figure: Circumscribed polygon for $J = 6$

Numerical examples - Circumscribed Polygon

$\mathcal{P}(J, \tau) \subset \mathbb{R}^2$ a regular polygon of $J \geq 3$ sides circumscribed around the circle of radius $\tau > 0$

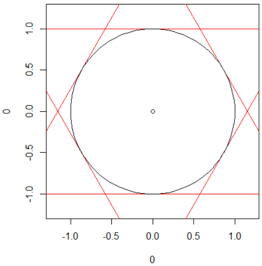
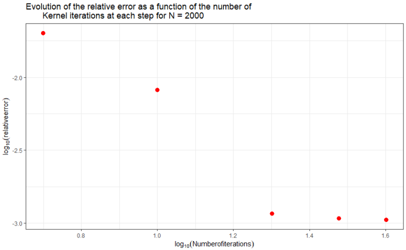
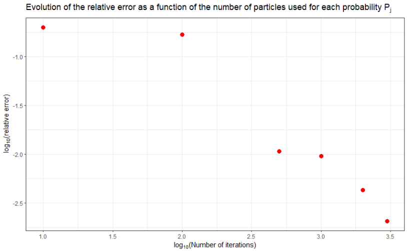


Figure: Circumscribed polygon for $J = 6$

- $1 \geq \frac{\mathbb{P}(X \in H)}{\exp(\frac{-\tau^2}{2})} \geq 1 - \frac{\pi^2 \tau^2}{6J^2}$
- For $J = 360, \tau = 6$, we have $\mu \approx 1.52 \times 10^{-8}$



(a) Relative error on the estimate $\hat{\mu}$ of the ALOE algorithm for different numbers of kernel iterations at each step in the multilevel splitting part of the algorithm



(b) Relative error on the estimate $\hat{\mu}$ of the ALOE algorithm in logarithmic scale for different numbers N of particles

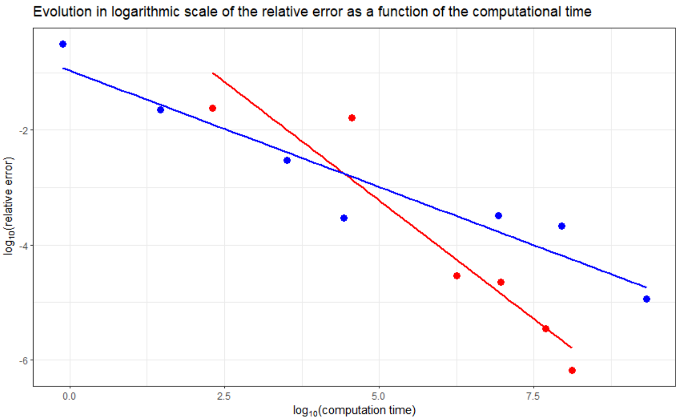
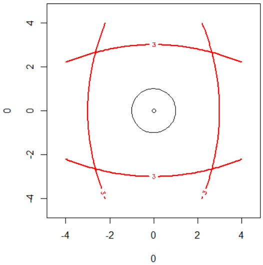


Figure: Relative error as a function of the computational time. The red dots are obtained from the new ALOE estimator and the blue ones from the more intuitive one. The lines are the linear regressions performed on these sets of points

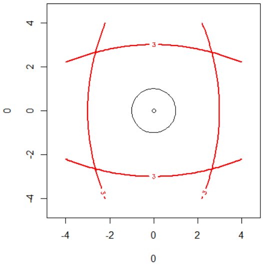
Numerical examples - Rare events with quadratic curves boundaries

- $H_j = \{(x,y) : c(x\sin(\theta_j) + y\cos(\theta_j))^2 + (x\cos(\theta_j) - y\sin(\theta_j)) \geq \tau_j\}$ where c a constant and $0 \leq \theta_j \leq 2\pi$



Numerical examples - Rare events with quadratic curves boundaries

- $H_j = \{(x, y) : c(x\sin(\theta_j) + y\cos(\theta_j))^2 + (x\cos(\theta_j) - y\sin(\theta_j)) \geq \tau_j\}$ where c a constant and $0 \leq \theta_j \leq 2\pi$



We run our numerical example with :

- θ_j goes from 0 to 2π by step of $\frac{\pi}{8}$, $J = 16$
- $\tau_j = \tau = 4$

For $N = 3000$, the new ALOE estimator gives $\hat{\mu} = 2.854815 \times 10^{-7}$

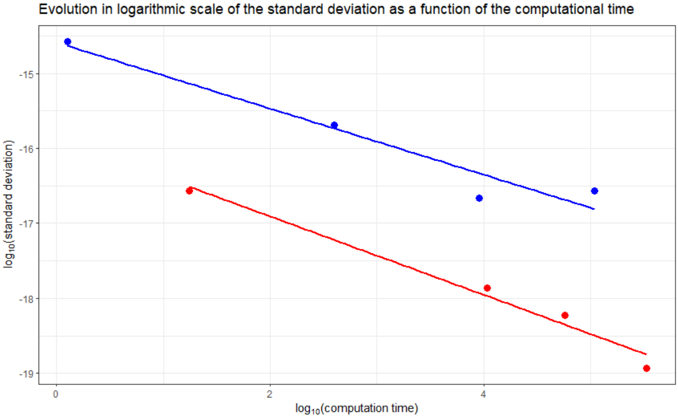


Figure: Standard deviation of $\hat{\mu}$ as a function of the computation time. The red dots are obtained from the new ALOE estimator and the blue ones from the more intuitive one. The lines are the linear regressions performed on these sets of points