Simulation and Estimation of Extreme Quantiles and Extreme Probabilities

Arnaud Guyader

Université Rennes 2, IRMAR & INRIA Rennes

ESSEC Working Group on Risk
October 21, 2013, Paris

Framework

- $X \sim \mu$ on $\mathbb{R}^d$ and $\Phi : \mathbb{R}^d \to \mathbb{R}$ is the score function.
- $Y = \Phi(X)$ is a random variable on $\mathbb{R}$.
- The real number q lies far out in the right-hand tail of the distribution of Y.

$\Rightarrow$ Goal: estimate $p := \mathbb{P}(\Phi(X) > q) = \mathbb{P}(Y > q) \approx 0$.

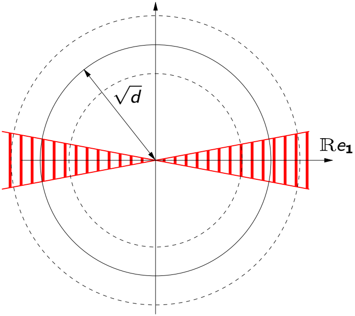
Remarks:

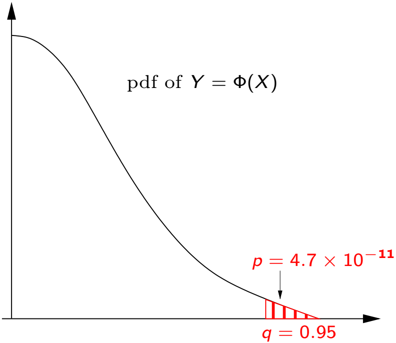
1. $p \approx 0 \Rightarrow$ Crude Monte Carlo is a disaster:

$$\frac{\mathrm{Var}(\hat{p}_{mc})}{p^2} = \frac{\mathrm{Var}(\#\{i : \ \Phi(X_i) > q\}/N)}{p^2} = \frac{1-p}{Np} \approx \frac{1}{Np}.$$

2. Assume that Φ acts as a black-box $\Rightarrow$ no importance sampling.

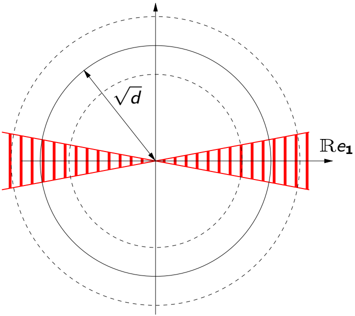
A Toy Example from Watermarking [Merhav & Sabbag, 2008]

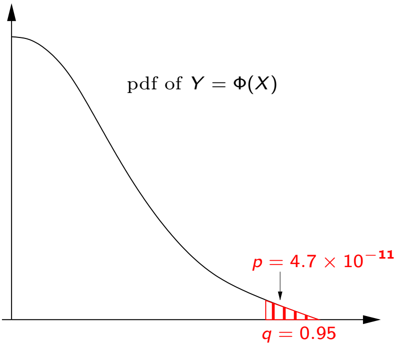




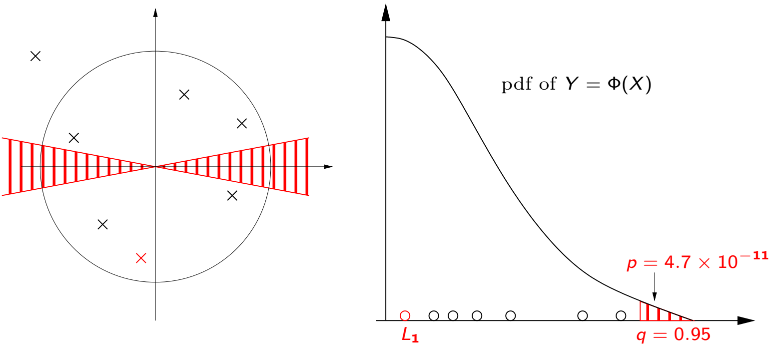
- Random vector $X \sim \mathcal{N}(0, I_d)$.
- Score function $\Phi(X) = \frac{|\langle X, e_1 \rangle|}{\|X\|} = \frac{|X_1|}{\sqrt{X_1^2 + \cdots + X_d^2}}$.
- Aim: Estimate $p := \mathbb{P}(\Phi(X) > q)$, double sheet hypercone.
- Example: if $d = 20$ and $q = 0.95$, direct numerical computation *via* the $\mathcal{F}^1_{19}$ Fisher law $\Rightarrow p = 4.704 \cdots \times 10^{-11}$.

A Selection/Mutation Algorithm

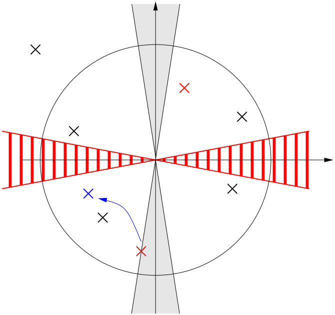


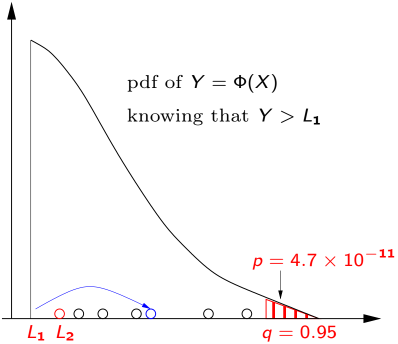


A Selection/Mutation Algorithm

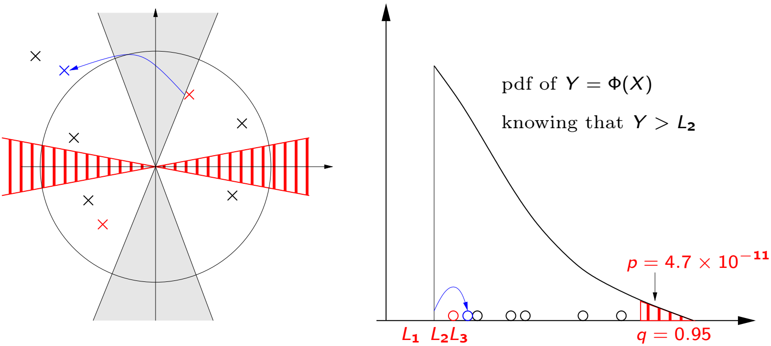


A Selection/Mutation Algorithm

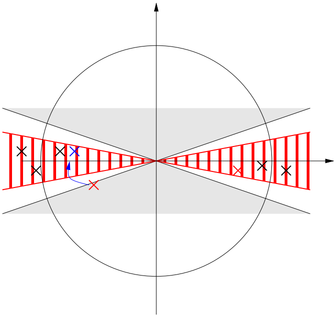


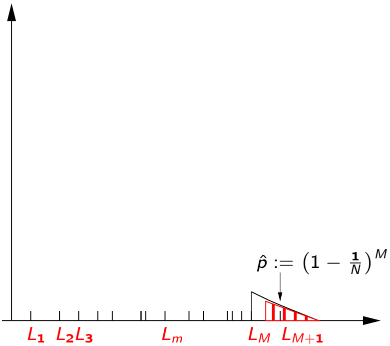


A Selection/Mutation Algorithm

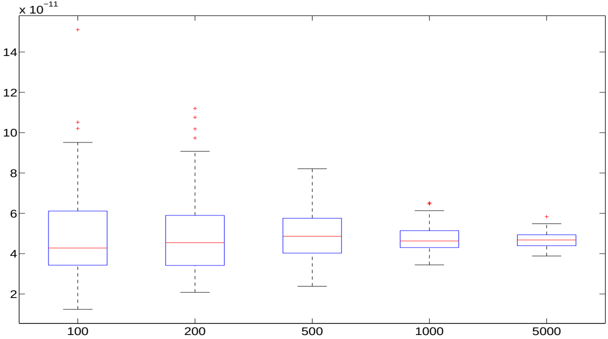


A Selection/Mutation Algorithm





Boxplots for $\hat{p}$



- Parameters: Boxplots obtained with 100 simulations for $N = 100$ to $N = 5,000$ particles.
- Recall (direct numerical computation): $p = 4.704 \cdots \times 10^{-11}$.

Distribution of the Levels

- Assumption $(\mathcal{H})$: the cdf F of $Y = \Phi(X)$ is continuous.
- Cumulative Hazard Function: $\Lambda(y) := -\log \mathbb{P}(Y > y)$.
- Remark: $(\mathcal{H}) \Rightarrow F(Y) \sim \mathcal{U}_{[0,1]}$ and $\Lambda(Y) \sim \mathcal{E}(1)$.

Theorem

The random variables $\Lambda(L_1), \Lambda(L_2), \Lambda(L_3), \dots$ are distributed as the successive arrival times of a Poisson process with rate N, that is,

$$\Lambda(L_m) \stackrel{\mathcal{L}}{=} \frac{1}{N} \sum_{j=1}^{m} E_j,$$

where $E_1, E_2, E_3, \dots$ are i.i.d. $\mathcal{E}(1)$.

Distributions of M and $\hat{p}$

Recall: by definition of the stopping rule

- the number of steps is random: $M := \max\{m : L_m \leq q\}$.
- the estimator is then defined as $\hat{p} := \left(1 - \frac{1}{N}\right)^M$.

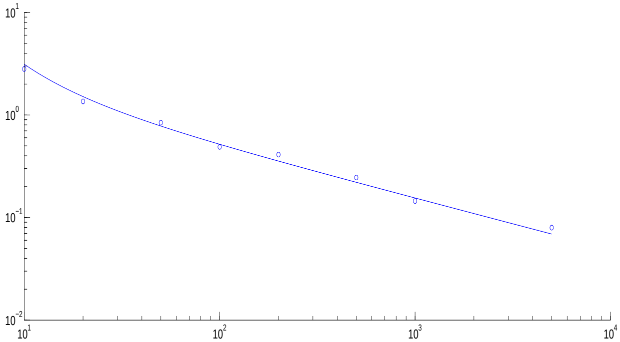
Corollary

1. *M is a Poisson random variable: $M \sim \mathcal{P}(-N \log p)$.*
2. *$\hat{p}$ is a discrete random variable with distribution*

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

It follows that $\hat{p}$ is an unbiased estimator of p.

Relative Standard Deviation

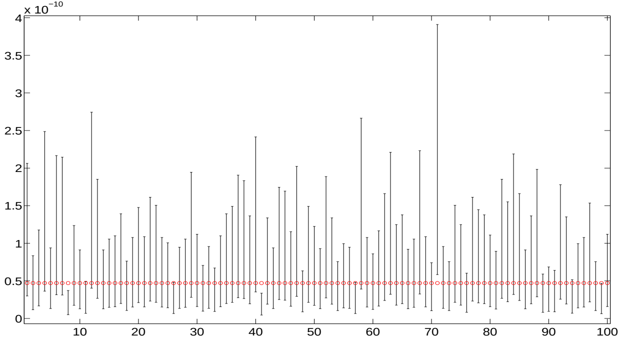


Recall: comparison with crude Monte Carlo

$$\frac{\sigma(\hat{p})}{p} = \sqrt{p^{-\frac{1}{N}} - 1} \approx \frac{\sqrt{-\log p}}{\sqrt{N}} \ll \frac{1}{\sqrt{Np}} \approx \frac{\sqrt{1-p}}{\sqrt{Np}} = \frac{\sigma(\hat{p}_{mc})}{p}$$

Log-log plot $\Rightarrow$ slope of the normalized SD $\approx -\frac{1}{2}$ for large N.

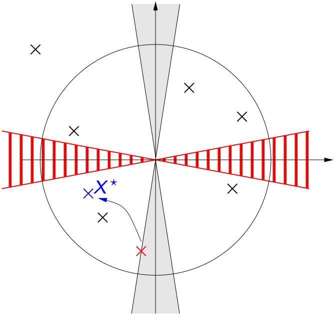
Confidence Intervals for $N = 100$ particles

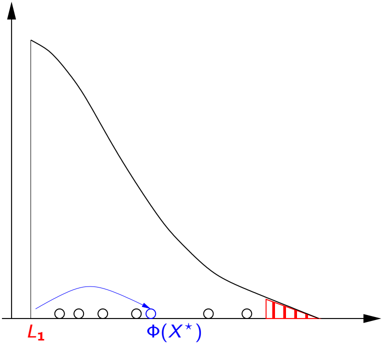


Recall: $M \sim \mathcal{P}(-N \log p) \approx \mathcal{N}(-N \log p, -N \log p)$, hence the following 95% confidence interval for p

$$\hat{p}\ \exp\left(-1.96\sqrt{\frac{-\log \hat{p}}{N}}\right) \leq p \leq \hat{p}\ \exp\left(+1.96\sqrt{\frac{-\log \hat{p}}{N}}\right)$$

Simulation of a New Particle





$\Rightarrow$ Goal: simulate $X^\star$ such that

(i) $X^\star \sim \mu_1 := \mathcal{L}(X|\Phi(X) > L_1)$,

(ii) $X^\star \perp (N-1)$ other particles.

A Possible Route

Ingredient: a μ-reversible transition kernel K.

- Example: if $X \sim \mathcal{N}(0, I_d)$, then $X' = \frac{X+\sigma W}{\sqrt{1+\sigma^2}}$ makes the job for any $\sigma > 0$ as long as $W \sim \mathcal{N}(0, I_d)$ and $W \perp X$.
- Remark: if no obvious K, then use Metropolis-Hastings.

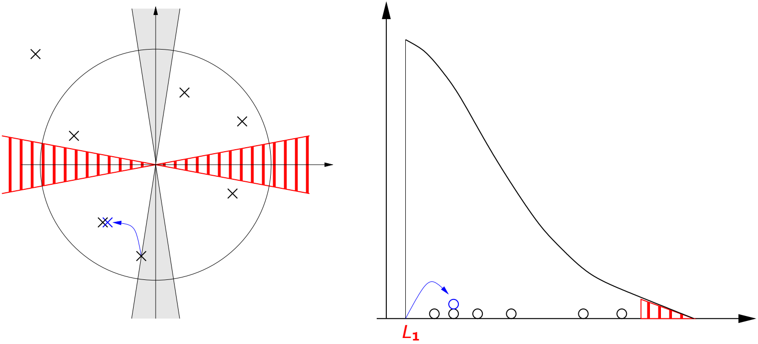
First Step: iterate the transition kernel K_1:

$$K_1(x, dx') := \begin{cases} K(x, dx') & \text{if } \Phi(x') > L_1 \\ \delta_x(dx') & \text{if } \Phi(x') \leq L_1 \end{cases}$$

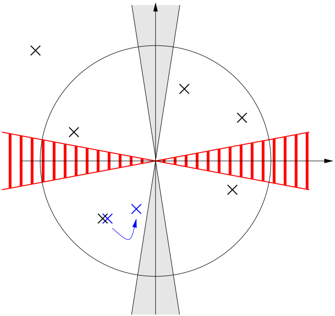
Remark: Importance of the tuning parameter σ

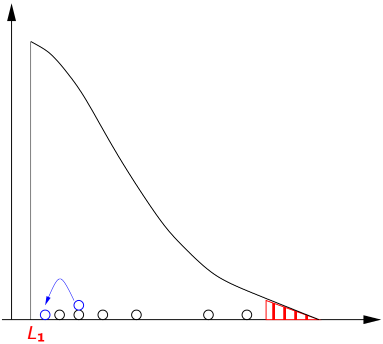
- σ too large $\Rightarrow$ most proposed transitions are refused.
- σ too small $\Rightarrow$ the particle moves slowly.

Illustration

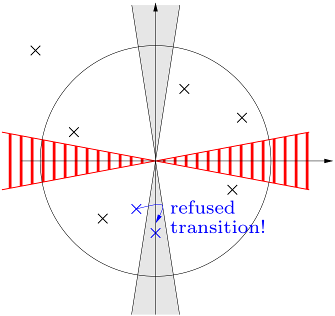


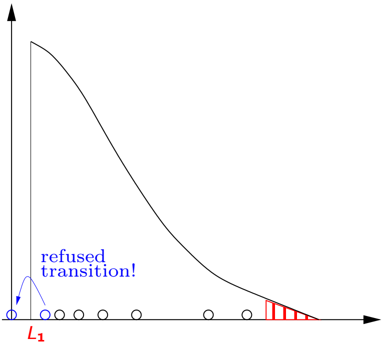
Illustration



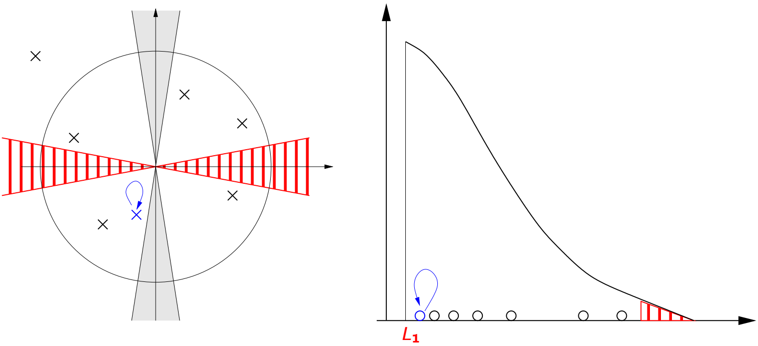


Illustration

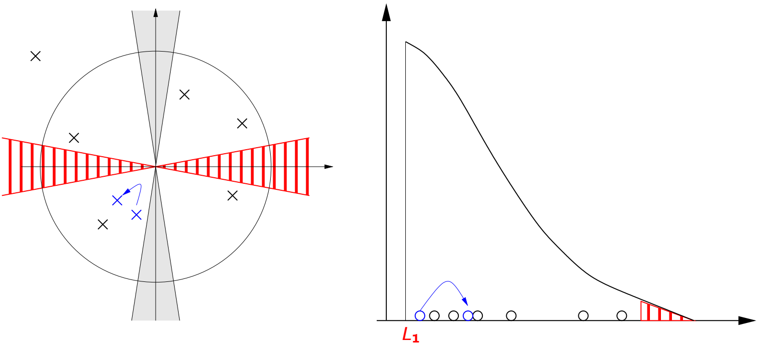




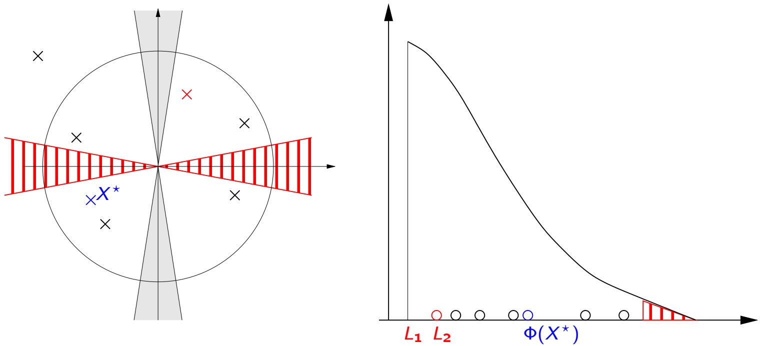
Illustration



Illustration



Illustration



A Possible Route

Ingredient: a μ-reversible transition kernel K.

- Example: if $X \sim \mathcal{N}(0, I_d)$, then $X' = \frac{X+\sigma W}{\sqrt{1+\sigma^2}}$ makes the job for any $\sigma > 0$ as long as $W \sim \mathcal{N}(0, I_d)$ and $W \perp X$.
- Remark: if no obvious K, then use Metropolis-Hastings.

First Step: iterate the transition kernel K_1:

$$K_1(x, dx') := \begin{cases} K(x, dx') & \text{if } \Phi(x') > L_1 \\ \delta_x(dx') & \text{if } \Phi(x') \leq L_1 \end{cases}$$

Remark: Importance of the tuning parameter σ

- σ too large $\Rightarrow$ most proposed transitions are refused.
- σ too small $\Rightarrow$ the particle moves slowly.

Efficiency [Hammersley & Handscomb, 1964]

The efficiency of an estimator $\hat{p}$ is defined as

$$E(\hat{p}) := \frac{1}{\mathrm{Var}(\hat{p}) \times C(\hat{p})}$$

- Crude Monte Carlo: $\mathrm{Var}(\hat{p}_{mc}) \times C(\hat{p}_{mc}) = p(1-p)$.
- Our estimator: $\mathrm{Var}(\hat{p}) \times C(\hat{p}) = k(p \log p)^2 \log N$.

$\Rightarrow$ $\hat{p}$ is more efficient than $\hat{p}_{mc}$ iff

$$k \log N < \frac{1-p}{p(\log p)^2}.$$

Ex: If $N = 200$ and $k = 10$, then $E(\hat{p}) > E(\hat{p}_{mc})$ iff $p < 10^{-4}$.

Quantile Estimation

Goal: given a (very low) probability p, estimate q such that $\mathbb{P}(\Phi(X) > q) = p$.

$\Rightarrow$ Algorithm: for $m := \left\lceil \frac{\log(p)}{\log(1-N^{-1})} \right\rceil$, simply set $\hat{q} = L_m$.

Proposition

If the cdf F of $Y = \Phi(X)$ is differentiable at point q, with density $f(q) \neq 0$, then

$$\sqrt{N}(\hat{q} - q) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{-p^2 \log p}{f(q)^2}\right).$$

Remark: for the crude Monte Carlo estimator, one has

$$\sqrt{N}(\hat{q}_{mc} - q) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{p(1-p)}{f(q)^2}\right).$$

Confidence Intervals

- The Poisson Trick: $\mathbb{P}(L_k \leq q < L_{k+1}) = \mathbb{P}(M = k)$, where

$$M \sim \mathcal{P}(m) \approx \mathcal{N}(m, m)$$

with $m := \left\lceil \frac{\log(p)}{\log(1-N^{-1})} \right\rceil$.

- Denote

$$\begin{cases} \max = \lceil m + 1.96\sqrt{m} \rceil \\ \min = \lfloor m - 1.96\sqrt{m} \rfloor \end{cases}$$

- 95% Confidence interval: $I = [L_{\min}, L_{\max}]$.
- Additional cost $\approx 1.96\sqrt{-N \log p}$ supplementary steps.

Portfolio Credit Risk

- Goal: Rare event estimation for portfolio credit risk (PCR).
- Principle: PCR is measured through probabilities of large losses, typically due to defaults of many obligors (sources of credit risk).
- Essential ingredient: The model of dependence between these sources of credit risk.
- Consequence: estimating PCR is challenging both because of
  - the rare event property of large losses,
  - the dependence between defaults.
- A possible route: for the Gaussian copula model, Glasserman *et al.* (2008) have proposed a relevant Importance Sampling technique.

Notations [Glasserman, Kang & Shahabuddin, 2008]

- n is the number of obligors to which the portfolio is exposed.
- T types of obligors, $t(k) \in \{1, \ldots, T\}$ is the type of obligor k.
- Y_k default indicator (=1 for default) for the k-th obligor.
- $p_k = \mathbb{P}(Y_k = 1)$ probability that the k-th obligor defaults.
- Dependence: multifactor Gaussian copula model with T types
  - $Y_k = \mathbb{1}\{V_k > q_{\mathcal{N}(0,1)}(1 - p_k)\}$ where
  - Latent variable $V_k = a'_{t(k)} Z + \sqrt{1 - \|a_{t(k)}\|^2}\ \varepsilon_k \sim \mathcal{N}(0, 1)$
  - Each factor-loading vector a_t is in the open ball $\mathcal{B}_d(0, 1)$.
  - $Z \sim \mathcal{N}(0, I_d)$ is the vector of systematic risks.
  - $\varepsilon_k \sim \mathcal{N}(0, 1)$ is the idiosyncratic risk of the k-th obligor.
- ℓ_k is the loss resulting from default of the k-th obligor.
- $L_n = \ell_1 Y_1 + \cdots + \ell_n Y_n$ is the total loss from defaults.

$\Rightarrow$ Objective of Glasserman *et al.*: estimate $\mathbb{P}(L_n > q) \approx 0$.

A Toy Example [Glasserman, Kang & Shahabuddin, 2008]

- Parameters: $n = 1,000$ obligors, $T = 25$ types of obligors, $d = 5$ factors, factor-loading vectors a_t have 60% of non-zero coefficients, marginal default probabilities $p_k \in [0.1\%, 5\%]$.
- Disclaimer: "*This example is not particularly realistic from a financial point of view (...). However we test this case as a challenging example to validate the efficiency of our method.*"
- Our goal: estimate q such that $\mathbb{P}(L_n > q) \leq 10^{-6}$.

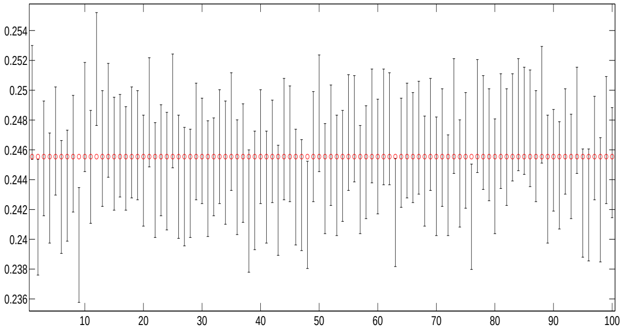
Principle: See the total loss L_n as a score function

$$L_n = \ell_1 Y_1 + \cdots + \ell_n Y_n = \Phi(Z, \varepsilon_1, \ldots, \varepsilon_n) = \Phi(X)$$

so that a particle $X \sim \mathcal{N}(0, I_{d+n})$ represents a set of (systematic + idiosyncratic) risks.

$\Rightarrow$ Direct application of the previous algorithm.

Confidence Intervals for $N = 1,000$ particles



Remark: $L_n = \Phi(X)$ is a purely discrete random variable, so that previous theoretical results should not apply here, but still...

To sum up

A new multilevel splitting method with a random number of levels.

- Pros
  - Non asymptotic theoretical results (law, moments, confidence intervals), see [GHM,2011] for details,
  - Extreme quantile estimation is new in multilevel splitting,
  - Works also in the context of rare events for stochastic processes, e.g. in molecular dynamics [CGLP,2011].
- Cons
  - Theoretical results are valid only under the "perfect resampling" assumption, and...
  - ... for continuous cdf, but *cf.* [CGRV,2011] for SAT problems,
  - Requires an efficient transition kernel at each resampling step,
  - Not parallelizable, see also [CDFG,2012] for a variant.

Remark: There is a connection with the Nested Sampling algorithm for the comparison of models in Bayesian Statistics [Skilling, 2004].

References

- A. Guyader, N. W. Hengartner and E. Matzner-Løber, Simulation and Estimation of Extreme Quantiles and Extreme Probabilities, *Applied Mathematics & Optimization*, 2011.
- F. Cérou, A. Guyader, T. Lelièvre and D. Pommier, A Multiple Replica Approach to Simulate Reactive Trajectories, *Journal of Chemical Physics*, 2011.
- F. Cérou, A. Guyader, R. Rubinstein and R. Vaisman, On the Use of Smoothing to Improve the Performance of the Splitting Method, *Stochastic Models*, 2011.
- F. Cérou, P. Del Moral, T. Furon and A. Guyader, Sequential Monte Carlo for Rare Event Estimation, *Statistics and Computing*, 2012.