Rare events: models and simulations

Josselin Garnier (Université Paris Diderot)

`http://www.proba.jussieu.fr/~garnier`

Cf: http://www.proba.jussieu.fr/~garnier/expo1.pdf

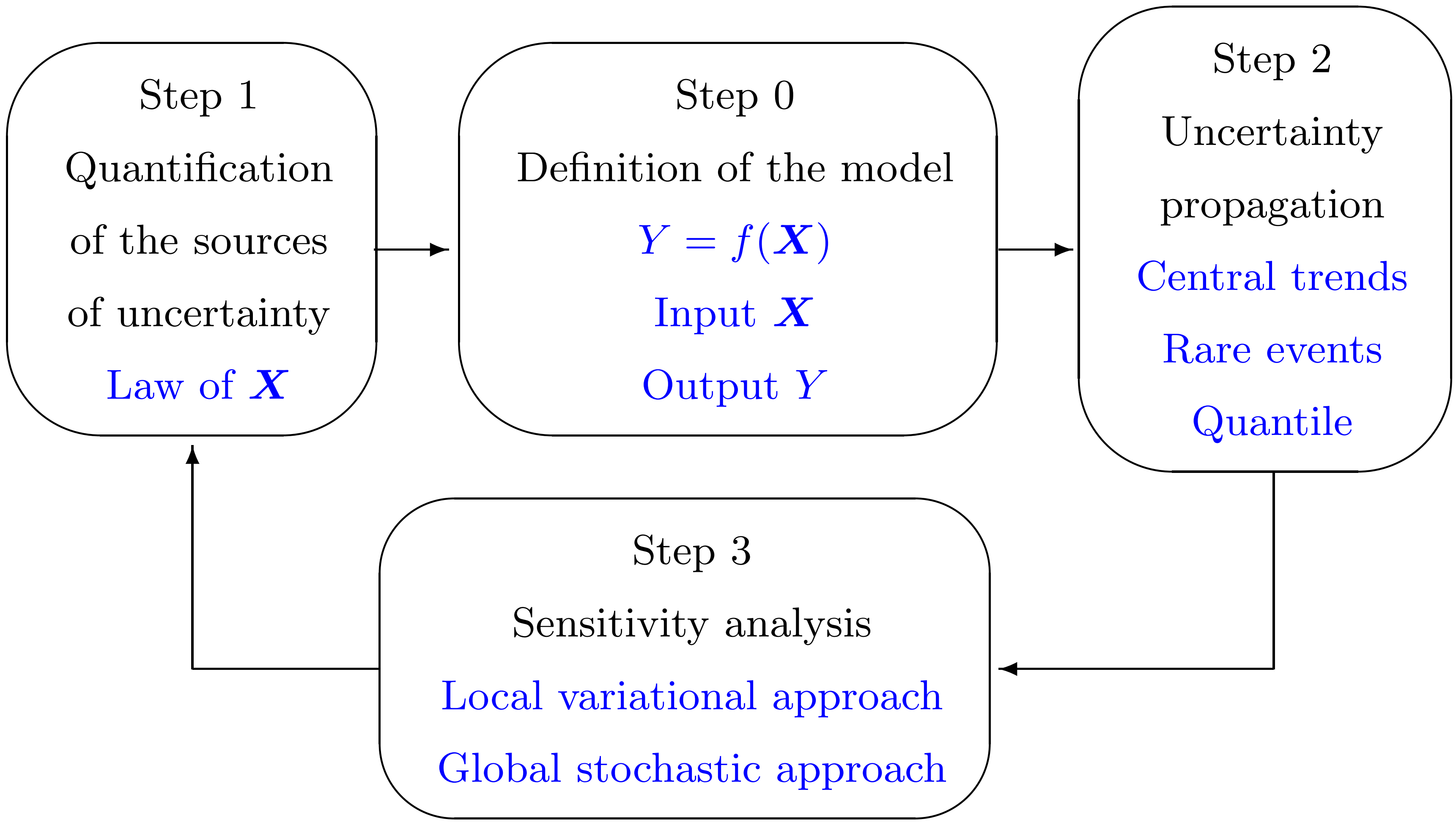
- Introduction to uncertainty quantification.

- Estimation of the probability of a rare event (such as the failure of a complex system).
- Standard methods (quadrature, Monte Carlo).
- Advanced Monte Carlo methods (variance reduction techniques).
- Interacting particle systems.

- Quantile estimation.

Uncertainty quantification

- General problem:
- How can we model the uncertainties in physical and numerical models ?
- How can we estimate (quantify) the variability of the output of a code or an experiment as a function of the variability of the input parameters ?
- How can we estimate quantify the sensitivity or the variability of the output of a code or an experiment with respect to one particular parameter ?


Step 1
Quantification
of the sources
of uncertainty
Law of $\boldsymbol{X}$
Step 0
Definition of the model
$Y = f(\boldsymbol{X})$
Input $\boldsymbol{X}$
Output Y
Step 2
Uncertainty
propagation
Central trends
Rare events
Quantile
Step 3
Sensitivity analysis
Local variational approach
Global stochastic approach

Uncertainty propagation

- Context: numerical code (black box) or experiment

$$Y = f(\boldsymbol{X})$$

with Y = output

$\boldsymbol{X}$ = random input parameters, with known distribution (with pdf $p(\boldsymbol{x})$)

$$\mathbb{P}(\boldsymbol{X} \in A) = \int_A p(\boldsymbol{x}) d\boldsymbol{x} \qquad \text{for any } A \in \mathcal{B}(\mathbb{R}^d)$$

f = deterministic function $\mathbb{R}^d \to \mathbb{R}$ (computationally expensive).

- Goal: estimation of a quantity of the form

$$\mathbb{E}[g(Y)]$$

with an "error bar" and the minimal number of simulations.

Examples (for a real-valued output Y):

- $g(y) = y \to$ mean of Y, i.e. $\mathbb{E}[Y]$
- $g(y) = y^2 \to$ variance of Y, i.e. $\text{Var}(Y) = \mathbb{E}[(Y - \mathbb{E}[Y])^2] = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2$
- $g(y) = \mathbf{1}_{[a,\infty)}(y) \to$ probability $\mathbb{P}(Y \geq a)$.

Analytic method

- The quantity to be estimated is a d-dimensional integral:

$$I = \mathbb{E}[g(Y)] = \mathbb{E}[h(\boldsymbol{X})] = \int_{\mathbb{R}^d} h(\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x}$$

where $p(\boldsymbol{x})$ is the pdf of $\boldsymbol{X}$ and $h(\boldsymbol{x}) = g(f(\boldsymbol{x}))$.

- In simple cases (when the pdf p and the function h have explicit expressions), one can sometimes evaluate the integral exactly (exceptional situation).

Quadrature method

• The quantity to be estimated is a d-dimensional integral:

$$I = \mathbb{E}[g(Y)] = \mathbb{E}[h(\boldsymbol{X})] = \int_{\mathbb{R}^d} h(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}$$

where $p(\boldsymbol{x})$ is the pdf of $\boldsymbol{X}$ and $h(\boldsymbol{x}) = g(f(\boldsymbol{x}))$.

• If $p(\boldsymbol{x}) = \prod_{i=1}^{d} p_0(x_i)$, then it is possible to apply Gaussian quadrature with a tensorized grid with n^d points:

$$\hat{I} = \sum_{j_1=1}^{n} \cdots \sum_{j_d=1}^{n} \rho_{j_1} \cdots \rho_{j_d} h(\xi_{j_1}, \ldots, \xi_{j_d})$$

with the weights $(\rho_j)_{j=1,\ldots,n}$ and the points $(\xi_j)_{j=1,\ldots,n}$ associated to the quadrature with weighting function p_0.

• There exist quadrature methods with sparse grids (cf Smolyak).

• Quadrature methods are efficient when:
- the function $\boldsymbol{x} \to h(\boldsymbol{x})$ is smooth (and not only f),
- the dimension d is “small” (even with sparse grids).
They require many calls.

Monte Carlo method

Principle: replace the statistical expectation $\mathbb{E}[g(Y)]$ by an empirical mean.

Monte Carlo method: “head and tail” model

A code gives a real-valued output $Y = f(\boldsymbol{X})$. For a given a we want to estimate

$$P = \mathbb{P}(f(\boldsymbol{X}) \geq a)$$

- Monte Carlo method:

1) n simulations are carried out with n independent realizations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ (with the distribution of $\boldsymbol{X}$).
2) let us define

$$Z_k = \mathbf{1}_{[a,\infty)}(f(\boldsymbol{X}_k)) = \begin{cases} 1 & \text{if } f(\boldsymbol{X}_k) \geq a \text{ (head)} \\ 0 & \text{if } f(\boldsymbol{X}_k) < a \text{ (tail)} \end{cases}$$

- Intuition: when n is large, the empirical proportion of “1”s is close to P

$$\frac{Z_1 + \cdots + Z_n}{n} \simeq P$$

Therefore the empirical proportion of “1”s can be used to estimate P.

- Empirical estimator of P:

$$\hat{P}_n := \frac{1}{n} \sum_{k=1}^{n} Z_k$$

• Empirical estimator of P:

$$\hat{P}_n := \frac{1}{n} \sum_{k=1}^{n} Z_k$$

• The estimator is unbiased:

$$\mathbb{E}\left[\hat{P}_n\right] = \mathbb{E}\left[\frac{1}{n}\sum_{k=1}^{n} Z_k\right] = \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}[Z_k] = \mathbb{E}[Z_1] = P$$

• The law of large numbers shows that the estimator is convergent:

$$\hat{P}_n = \frac{1}{n}\sum_{k=1}^{n} Z_k \stackrel{n\to\infty}{\longrightarrow} \mathbb{E}[Z_1] = P$$

because the variables Z_k are independent and identically distributed (i.i.d.).

• Result (law of large numbers): Let $(Z_n)_{n\in\mathbb{N}^*}$ be a sequence of i.i.d. random variables. If $\mathbb{E}[|Z_1|] < \infty$, then

$$\frac{1}{n}\sum_{k=1}^{n} Z_k \stackrel{n\to\infty}{\longrightarrow} m \text{ with probability 1, with } m = \mathbb{E}[Z_1]$$

"The empirical mean converges to the statistical mean".

Error analysis: we want to quantify the fluctuations of $\hat{P}_n$ around P.

- Variance calculation:

$$\begin{aligned}
\mathrm{Var}[\hat{P}_n] &= \mathbb{E}\left[(\hat{P}_n - P)^2\right] \qquad \text{(mean square error)} \\
&= \mathbb{E}\left[\left(\sum_{k=1}^n \left(\frac{Z_k - P}{n}\right)\right)^2\right] = \sum_{k=1}^n \mathbb{E}\left[\left(\frac{Z_k - P}{n}\right)^2\right] \\
&= \frac{1}{n^2}\sum_{k=1}^n \mathbb{E}\left[(Z_k - P)^2\right] = \frac{1}{n}\mathbb{E}\left[(Z_1 - P)^2\right] \\
&= \frac{1}{n}\left(\mathbb{E}\left[Z_1^2\right] - P^2\right) \\
&= \frac{1}{n}(P - P^2)
\end{aligned}$$

- The relative error is therefore:

$$\text{Error} = \frac{\sqrt{\mathrm{Var}(\hat{P}_n)}}{\mathbb{E}[\hat{P}_n]} = \frac{\sqrt{\mathrm{Var}(\hat{P}_n)}}{P} = \frac{1}{\sqrt{n}}\left(\frac{1}{P} - 1\right)^{1/2}$$

Confidence intervals

• *Question*: The estimator $\hat{P}_n$ gives an approximate value of P, all the better as n is larger. How to quantify the error ?

• *Answer*: We build a confidence interval at the level 0.95, i.e. an empirical interval $[\hat{a}_n, \hat{b}_n]$ such that

$$\mathbb{P}\Big(P \in [\hat{a}_n, \hat{b}_n]\Big) \geq 0.95$$

Construction based on the De Moivre theorem:

$$\mathbb{P}\left(\left|\hat{P}_n - P\right| < c\frac{\sqrt{P-P^2}}{\sqrt{n}}\right) \stackrel{n\to\infty}{\longrightarrow} \frac{2}{\sqrt{2\pi}}\int_0^c e^{-x^2/2}dx$$

The right member is 0.05 if $c = 1.96$. Therefore

$$\mathbb{P}\left(P \in \left[\hat{P}_n - 1.96\frac{\sqrt{P-P^2}}{\sqrt{n}}, \hat{P}_n + 1.96\frac{\sqrt{P-P^2}}{\sqrt{n}}\right]\right) \simeq 0.95$$

• Result (central limit theorem): Let $(Z_n)_{n\in\mathbb{N}^*}$ be a sequence of i.i.d. random variables. If $\mathbb{E}[Z_1^2] < \infty$, then

$$\sqrt{n}\Big(\frac{1}{n}\sum_{k=1}^n Z_k - m\Big) \stackrel{n\to\infty}{\longrightarrow} \mathcal{N}(0,\sigma^2) \text{ in distribution}$$

where $m = \mathbb{E}[Z_1]$ and $\sigma^2 = \mathrm{Var}(Z_1)$.
"For large n, the error $\frac{1}{n}\sum_{k=1}^n Z_k - m$ has Gaussian distribution $\mathcal{N}(0,\sigma^2/n)$."

$$\mathbb{P}\left(P \in \left[\hat{P}_n - 1.96\frac{\sqrt{P-P^2}}{\sqrt{n}}, \hat{P}_n + 1.96\frac{\sqrt{P-P^2}}{\sqrt{n}}\right]\right) \simeq 0.95$$

The unknown parameter P is still in the bounds of the interval ! Two solutions:
- $P \in [0,1]$, therefore $\sqrt{P-P^2} < 1/2$ and

$$\mathbb{P}\left(P \in \left[\hat{P}_n - 0.98\frac{1}{\sqrt{n}}, \hat{P}_n + 0.98\frac{1}{\sqrt{n}}\right]\right) \geq 0.95$$

- asymptotically, we can replace P in the bounds by $\hat{P}_n$ (OK if $nP > 10$ and $n(1-P) > 10$):

$$\mathbb{P}\left(P \in \left[\hat{P}_n - 1.96\frac{\sqrt{\hat{P}_n-\hat{P}_n^2}}{\sqrt{n}}, \hat{P}_n + 1.96\frac{\sqrt{\hat{P}_n-\hat{P}_n^2}}{\sqrt{n}}\right]\right) \simeq 0.95$$

Conclusion: There is no bounded interval of $\mathbb{R}$ that contains P with probability one. There are bounded intervals (called confidence intervals) that contain P with probability close to one (chosen by the user).

Monte Carlo estimation: black box model

- Black box model (numerical code)

$$Y = f(\boldsymbol{X})$$

We want to estimate $I = \mathbb{E}[g(Y)]$, for some function $g : \mathbb{R} \to \mathbb{R}$.

- Empirical estimator:

$$\widehat{I}_n = \frac{1}{n} \sum_{k=1}^{n} g(f(\boldsymbol{X}_k))$$

where $(\boldsymbol{X}_k)_{k=1,\ldots,n}$ is a n-sample of $\boldsymbol{X}$.

This is the empirical mean of a sequence of i.i.d. random variables.

- The estimator $\widehat{I}_n$ is unbiased: $\mathbb{E}[\widehat{I}_n] = I$.
- The law of large numbers gives the convergence of the estimator:

$$\widehat{I}_n \stackrel{n\to\infty}{\longrightarrow} I \quad \text{with probability 1}$$

- Error:

$$\mathrm{Var}(\widehat{I}_n) = \frac{1}{n}\mathrm{Var}(g(Y))$$

Proof: the variance of a sum of i.i.d. random variables if the sum of the variances.

- Asymptotic confidence interval:

$$\mathbb{P}\left(I \in \left[\widehat{I}_n - 1.96\frac{\hat{\sigma}_n}{\sqrt{n}}, \widehat{I}_n + 1.96\frac{\hat{\sigma}_n}{\sqrt{n}}\right]\right) \simeq 0.95$$

where

$$\hat{\sigma}_n = \left(\frac{1}{n}\sum_{k=1}^{n} g(f(\boldsymbol{X}_k))^2 - \widehat{I}_n^2\right)^{1/2}$$

- Advantages of the MC method:

1) no regularity condition for f, g (condition: $\mathbb{E}[g(f(\boldsymbol{X}))^2] < \infty$).

2) convergence rate $1/\sqrt{n}$ in any dimension.

3) can be applied for any quantity that can be expressed as an expectation.

- One needs to simulate samples of $\boldsymbol{X}$.

Simulation of random variables

• How do we generate random numbers with a computer ? There is nothing random in a computer !

• Strategy:
- find a pseudo random number generator that can generate a sequence of numbers that behave like independant copies of a random variable with uniform distribution over $(0, 1)$.
- use deterministic transforms to generate numbers with any prescribed distribution using only the uniform pseudo random number generator.

- Pseudo random number generator

A 32-bit multiplicative congruential generator:

$$x_{n+1} = a x_n \bmod b,$$

with $a = 7^5$, $b = 2^{31} - 1$, and some integer x_0.
This gives a sequence of integer numbers in $\{0, 1, ..., 2^{31} - 2\}$.
The sequence $u_n = x_n/(2^{31} - 1)$ gives a "quasi-real" number between 0 and 1.
Note: the sequence is periodic, with period $2^{31} - 1$.
This is the generator mcg16807 of matlab (used in early versions).
Today: matlab uses mt19937ar (the period is $2^{19937} - 1$).

- Inversion method.

A little bit of theory:

Result: Let X be a real random variable with the cumulative distribution function (cdf) $F(x)$:

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} p(y)dy$$

Let U be a random variable with the distribution $\mathcal{U}(0,1)$. If F is one-to-one, then X and $F^{-1}(U)$ have the same distribution.

Proof: Set $Y = F^{-1}(U)$.

$$\mathbb{P}(Y \leq x) = \mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x)\,,$$

which shows that the cdf of Y is F.

• Extension: Let F be a cdf. The generalized inverse of F is $F^{-1} : (0,1) \to \mathbb{R}$ defined by:

$$F^{-1}(u) := \inf A_u, \qquad \text{where } A_u := \{x \in \mathbb{R} \text{ such that } F(x) \geq u\}$$

The generalized inverse always exists because, for any $u \in (0,1)$:
(i) $\lim_{x\to+\infty} F(x) = 1$, therefore $A_u \neq \emptyset$.
(ii) $\lim_{x\to-\infty} F(x) = 0$, therefore A_u is bounded from below.

Result: Let X be a random variable with the cdf $F(x)$. Let U be a random variable with the distribution $\mathcal{U}(0,1)$. Then X and $F^{-1}(U)$ have the same distribution.

- Example. Let us write a simulator of an exponential random variable:

$$p(x) = e^{-x}\mathbf{1}_{[0,\infty)}(x)$$

Its cdf is

$$F(x) = \begin{cases} 0 & \text{if } x < 0, \\ \int_0^x e^{-y}dy = 1 - e^{-x} & \text{if } x \geq 0 \end{cases}$$

whose reciprocal is:

$$F^{-1}(u) = -\ln(1-u)\,.$$

Therefore if U is a uniform random variable on $[0,1]$, then the random variable $X := -\ln(1-U)$ obeys the exponential distribution.

Moreover, U and $1-U$ have the same distribution, therefore $-\ln(U)$ also obeys the exponential distribution.

• Simulation of Gaussian random variables.

The inversion method requires the knowledge of the reciprocal cdf F^{-1}.
We do not always have the explicit expression of this reciprocal function.
An important example is the Gaussian distribution such that
$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-s^2/2} ds$ for $x \in \mathbb{R}$.

Box-Muller algorithm:
Let U_1, U_2 two independent random variables with uniform distribution over $[0, 1]$.
If

$$X = (-2 \ln U_1)^{1/2} \cos(2\pi U_2)$$

$$Y = (-2 \ln U_1)^{1/2} \sin(2\pi U_2)$$

then the random variables X and Y are independent and distributed according to $\mathcal{N}(0, 1)$.

Proof: for any test function $f : \mathbb{R}^d \to \mathbb{R}$, write $\mathbb{E}[f(X, Y)]$ as a two-dimensional integral and use polar coordinates.

- Rejection method for uniform distributions.

- Goal: build a generator of a random variable with uniform distribution over $D \subset \mathbb{R}^d$.

- Preliminary: Find a rectangular domain B such that $D \subset B$.

- Algorithm:

Sample M_1, M_2, ... independent and identically distributed with the uniform distribution over B, until the first time T when $M_i \in D$.

- Result: M_T is a random variable with the uniform distribution over D.

- Warning: the distribution of M_T is not the distribution of M_1, because T is random !

The time T obeys a geometric distribution with mean $|B|/|D|$ ($\to$ it is better to look for the smallest domain B).

• Rejection method for continuous distributions.

• Goal: build a generator of a random variable with pdf $p(\boldsymbol{x})$.

• Preliminary: find a pdf $q(\boldsymbol{x})$ such that we know a generator of a random variable with pdf $q(\boldsymbol{x})$ and we know $C \geq 1$ such that $p(\boldsymbol{x}) \leq Cq(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^d$.

• Algorithm:
Sample $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, ... with pdf $q(\boldsymbol{x})$ and U_1, U_2, ... with distribution $\mathcal{U}(0,1)$ until the first time T when $U_k < p(\boldsymbol{X}_k)/[Cq(\boldsymbol{X}_k)]$.

• Result: $\boldsymbol{X}_T$ is a random variable with the pdf $p(\boldsymbol{x})$.

The time T obeys a geometric distribution with mean C.

Proof: for any Borel set A:

$$\begin{aligned}
\mathbb{P}(\boldsymbol{X}_T \in A) &= \sum_{k=1}^{\infty} \mathbb{P}(\boldsymbol{X}_k \in A, T = k) \\
&= \sum_{k=1}^{\infty} \mathbb{P}\Big(\boldsymbol{X}_k \in A, U_k < \frac{p(\boldsymbol{X}_k)}{Cq(\boldsymbol{X}_k)}, U_{k-1} \geq \frac{p(\boldsymbol{X}_{k-1})}{Cq(\boldsymbol{X}_{k-1})}, \dots, U_1 \geq \frac{p(\boldsymbol{X}_1)}{Cq(\boldsymbol{X}_1)}\Big) \\
&= \sum_{k=1}^{\infty} \mathbb{P}\Big(\boldsymbol{X}_1 \in A, U_1 < \frac{p(\boldsymbol{X}_1)}{Cq(\boldsymbol{X}_1)}\Big) \mathbb{P}\Big(U_1 \geq \frac{p(\boldsymbol{X}_1)}{Cq(\boldsymbol{X}_1)}\Big)^{k-1} \\
&= \frac{\mathbb{P}\big(\boldsymbol{X}_1 \in A, U_1 < \frac{p(\boldsymbol{X}_1)}{Cq(\boldsymbol{X}_1)}\big)}{\mathbb{P}\big(U_1 < \frac{p(\boldsymbol{X}_1)}{Cq(\boldsymbol{X}_1)}\big)} \\
&= \frac{\mathbb{E}_{\boldsymbol{X}}\big[\mathbf{1}_{\boldsymbol{X}_1 \in A} \mathbb{E}_U[\mathbf{1}_{U_1 < \frac{p(\boldsymbol{X}_1)}{Cq(\boldsymbol{X}_1)}}]\big]}{\mathbb{E}_{\boldsymbol{X}}\big[\mathbb{E}_U[\mathbf{1}_{U_1 < \frac{p(\boldsymbol{X}_1)}{Cq(\boldsymbol{X}_1)}}]\big]} \\
&= \frac{\mathbb{E}_{\boldsymbol{X}}\big[\mathbf{1}_{\boldsymbol{X}_1 \in A} \frac{p(\boldsymbol{X}_1)}{Cq(\boldsymbol{X}_1)}\big]}{\mathbb{E}_{\boldsymbol{X}}\big[\frac{p(\boldsymbol{X}_1)}{Cq(\boldsymbol{X}_1)}\big]} \\
&= \frac{\int \mathbf{1}_{\boldsymbol{x} \in A} \frac{p(\boldsymbol{x})}{Cq(\boldsymbol{x})} q(\boldsymbol{x}) d\boldsymbol{x}}{\int \frac{p(\boldsymbol{x})}{Cq(\boldsymbol{x})} q(\boldsymbol{x}) d\boldsymbol{x}} \\
&= \int \mathbf{1}_{\boldsymbol{x} \in A} p(\boldsymbol{x}) d\boldsymbol{x}
\end{aligned}$$

Estimation of the probability of a rare event

- We look for an estimator for

$$P = \mathbb{P}(f(\boldsymbol{X}) \geq a)$$

where a is large (so that $P \ll 1$).

- Possible by Monte Carlo:

$$\hat{P}_n = \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{f(\boldsymbol{X}_k) \geq a}$$

where $(\boldsymbol{X}_k)_{k=1,\ldots,n}$ is a n-sample of $\boldsymbol{X}$.

Relative error:

$$\frac{\mathbb{E}[(\hat{P}_n - P)^2]^{1/2}}{P} = \frac{1}{\sqrt{n}} \frac{\mathrm{Var}(\mathbf{1}_{f(\boldsymbol{X}) \geq a})^{1/2}}{P} = \frac{1}{\sqrt{n}} \frac{\sqrt{1-P}}{\sqrt{P}} \overset{P \ll 1}{\simeq} \frac{1}{\sqrt{nP}}$$

$\hookrightarrow$ We need $nP > 1$ so that the relative error is smaller than 1 (not surprising) !

$\hookrightarrow$ We need variance reduction techniques.

Uncertainty propagation by metamodels

The complex code/experiment f is replaced by a metamodel (reduced model) $f_{\rm r}$ and one of the previous techniques is applied with $f_{\rm r}$ (analytic, quadrature, Monte Carlo).

$\rightarrow$ It is possible to call many times the metamodel.

$\rightarrow$ The choice of the metamodel is critical.

$\rightarrow$ The error control is not simple.

Taylor expansions

- We approximate the output $Y = f(\boldsymbol{X})$ by a Taylor series expansion $Y_{\rm r} = f_{\rm r}(\boldsymbol{X})$.

- Example:

\- We want to estimate $\mathbb{E}[Y]$ and $\text{Var}(Y)$ for $Y = f(\boldsymbol{X})$ with X_i uncorrelated, $\mathbb{E}[X_i] = \mu_i$ and $\text{Var}(X_i) = \sigma_i^2$ known, σ_i^2 small.
\- We approximate $Y = f(\boldsymbol{X})$ by $Y_{\rm r} = f_{\rm r}(\boldsymbol{X}) = f(\boldsymbol{\mu}) + \nabla f(\boldsymbol{\mu}) \cdot (\boldsymbol{X} - \boldsymbol{\mu})$. We find:

$$\mathbb{E}[Y] \simeq \mathbb{E}[Y_{\rm r}] = f(\boldsymbol{\mu}), \qquad \text{Var}(Y) \simeq \text{Var}(Y_{\rm r}) = \sum_{i=1}^{d} \partial_{x_i} f(\boldsymbol{\mu})^2 \sigma_i^2$$

We just need to compute $f(\boldsymbol{\mu})$ and $\nabla f(\boldsymbol{\mu})$ (evaluation of the gradient by finite differences, about $d+1$ calls to f).

- Rapid, analytic, allows to evaluate approximately central trends of the output (mean, variance).

- Suitable for small variations of the input parameters and a smooth model (that can be linearized).

- Local approach. In general, no error control.

Reliability method for estimation of the probability of a rare event:

$$P = \mathbb{P}(f(\boldsymbol{X}) \geq a) = \mathbb{P}(\boldsymbol{X} \in \mathcal{F}) = \int_{\mathcal{F}} p(\boldsymbol{x}) d\boldsymbol{x}, \qquad \mathcal{F} = \{\boldsymbol{x} \in \mathbb{R}^d,\ f(\boldsymbol{x}) \geq a\}$$

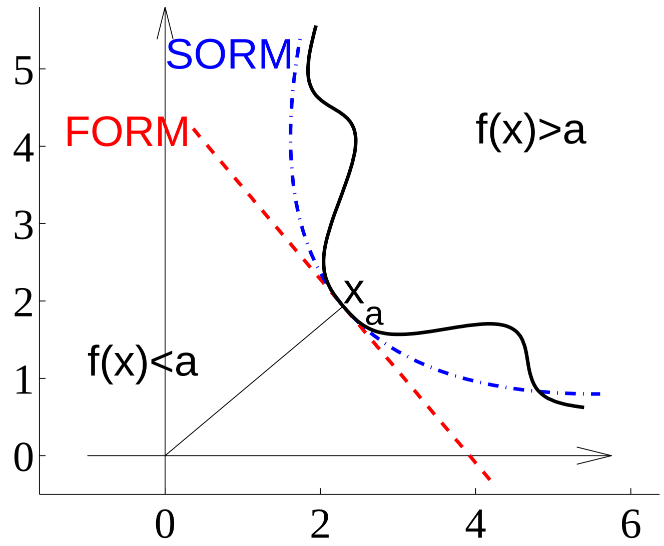
- The FORM-SORM method is analytic but approximate, without control error.
- The X_i are assumed to be independent and with Gaussian distribution with mean zero and variance one (or we use an isoprobabilist transform to deal with this situation): $p(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}} \exp(-\frac{|\boldsymbol{x}|^2}{2})$.
- One gets by optimization (with contraint) the "design point" $\boldsymbol{x}_a$ (the most probable failure point), i.e. $\boldsymbol{x}_a = \text{argmin}\{\|\boldsymbol{x}\|^2 \text{ s.t. } f(\boldsymbol{x}) \geq a\}$.
- One approximates the failure surface $\{\boldsymbol{x} \in \mathbb{R}^d, f(\boldsymbol{x}) = a\}$ by a smooth surface $\hat{\mathcal{F}}$ that allows for an analytic calculation $\hat{P} = \int_{\hat{\mathcal{F}}} p(\boldsymbol{x}) d\boldsymbol{x}$:

- a hyperplane for FORM

(and then $\hat{P} = \frac{1}{2}\text{erfc}\big(\frac{|\boldsymbol{x}_a|}{\sqrt{2}}\big)$),

- a quadratic form for SORM

(and then $\hat{P}$ = Breitung's formula).



Cf: O. Ditlevsen et H.O. Madsen,

Structural reliability methods, Wiley, 1996.

Variance reduction techniques

Goal: reduce the variance of the Monte Carlo estimator:

$$\mathbb{E}\big[(\hat{I}_n - I)^2\big] = \frac{1}{n}\mathrm{Var}(h(\boldsymbol{X}))$$

where $h(\boldsymbol{x}) = g(f(\boldsymbol{x}))$, $I = \mathbb{E}[h(\boldsymbol{X})]$, $\widehat{I}_n = \frac{1}{n}\sum_{k=1}^{n} h(\boldsymbol{X}_k)$.

• The methods

- Importance sampling
- Control variates
- Antithetic variables
- Stratification

reduce the constant without changing $1/n$, stay close to the Monte Carlo method (parallelizable).

• The methods

- Quasi-Monte Carlo

aim at changing $1/n$.

• The methods

- Interacting particle systems (genetic algorithms, subset sampling,...)

are more different from Monte Carlo (sequential).

Importance sampling

- The goal is to estimate $I = \mathbb{E}[h(\boldsymbol{X})]$ for $\boldsymbol{X}$ a random vector and $h(\boldsymbol{x}) = g(f(\boldsymbol{x}))$ a deterministic function.

- Observation: the representation of I as an expectation is not unique. If $\boldsymbol{X}$ has the pdf $p(\boldsymbol{x})$:

$$I = \mathbb{E}_p[h(\boldsymbol{X})] = \int h(\boldsymbol{x})p(\boldsymbol{x})dx = \int \frac{h(\boldsymbol{x})p(\boldsymbol{x})}{q(\boldsymbol{x})} q(\boldsymbol{x})d\boldsymbol{x} = \mathbb{E}_q\Big[\frac{h(\boldsymbol{X})p(\boldsymbol{X})}{q(\boldsymbol{X})}\Big]$$

The choice of the pdf q depends on the user.

- Idea: when we know that $h(\boldsymbol{X})$ is sensitive to certain values of $\boldsymbol{X}$, instead of sampling $\boldsymbol{X}_k$ with the original pdf $p(\boldsymbol{x})$ of $\boldsymbol{X}$, a biased pdf $q(\boldsymbol{x})$ is used that makes more likely the "important" realizations.

- Using the representation

$$I = \mathbb{E}_p[h(\boldsymbol{X})] = \mathbb{E}_q\left[h(\boldsymbol{X})\frac{p(\boldsymbol{X})}{q(\boldsymbol{X})}\right]$$

we can propose the estimator:

$$\hat{I}_n = \frac{1}{n}\sum_{k=1}^{n} h(\boldsymbol{X}_k)\frac{p(\boldsymbol{X}_k)}{q(\boldsymbol{X}_k)}$$

where $(\boldsymbol{X}_k)_{k=1,\dots,n}$ is a n-sample with the distribution with pdf q.

- The estimator is unbiased:

$$\begin{aligned}
\mathbb{E}_q[\widehat{I}_n] &= \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}_q\Big[h(\boldsymbol{X}_k)\frac{p}{q}(\boldsymbol{X}_k)\Big] = \mathbb{E}_q\Big[h(\boldsymbol{X})\frac{p}{q}(\boldsymbol{X})\Big] \\
&= \int h(\boldsymbol{x})\frac{p}{q}(\boldsymbol{x})q(\boldsymbol{x})d\boldsymbol{x} = \int h(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} = \mathbb{E}_p\Big[h(\boldsymbol{X})\Big] = I
\end{aligned}$$

- The estimator is convergent:

$$\hat{I}_n = \frac{1}{n}\sum_{k=1}^{n} h(\boldsymbol{X}_k)\frac{p(\boldsymbol{X}_k)}{q(\boldsymbol{X}_k)} \stackrel{n\to\infty}{\longrightarrow} \mathbb{E}_q\Big[h(\boldsymbol{X})\frac{p(\boldsymbol{X})}{q(\boldsymbol{X})}\Big] = \mathbb{E}_p\big[h(\boldsymbol{X})\big] = I$$

- The variance of the estimator is:

$$\mathrm{Var}(\hat{I}_n) = \frac{1}{n}\mathrm{Var}_q\Big(h(\boldsymbol{X})\frac{p(\boldsymbol{X})}{q(\boldsymbol{X})}\Big) = \frac{1}{n}\left(\mathbb{E}_p\left[h(\boldsymbol{X})^2\frac{p(\boldsymbol{X})}{q(\boldsymbol{X})}\right] - \mathbb{E}_p\left[h(\boldsymbol{X})\right]^2\right)$$

By a judicious choice of q the variance can be dramatically reduced.

- Critical points: it is necessary to know the likelihood ratio $\frac{p(\boldsymbol{x})}{q(\boldsymbol{x})}$ and to know how to simulate $\boldsymbol{X}$ with the pdf q.

• Optimal importance sampling.

The best importance distribution is the one that minimizes the variance $\mathrm{Var}(\hat{I}_n)$.

It is the solution of the minimization problem: find the pdf $q(\boldsymbol{x})$ minimizing

$$\mathbb{E}_p\left[h(\boldsymbol{X})^2\frac{p(\boldsymbol{X})}{q(\boldsymbol{X})}\right] = \int h(\boldsymbol{x})^2\frac{p^2(\boldsymbol{x})}{q(\boldsymbol{x})}d\boldsymbol{x}$$

Solution (when h is nonnegative-valued):

$$q_{\mathrm{opt}}(\boldsymbol{x}) = \frac{h(\boldsymbol{x})p(\boldsymbol{x})}{\int h(\boldsymbol{x}')p(\boldsymbol{x}')d\boldsymbol{x}'}$$

We then find

$$\mathrm{Var}(\hat{I}_n) = \frac{1}{n}\left(\mathbb{E}_p\left[h(\boldsymbol{X})^2\frac{p(\boldsymbol{X})}{q_{\mathrm{opt}}(\boldsymbol{X})}\right] - \mathbb{E}_p\left[h(\boldsymbol{X})\right]^2\right) = 0 \;!$$

Pratically: the denominator of $q_{\mathrm{opt}}(\boldsymbol{x})$ is the desired quantity $\mathbb{E}[h(\boldsymbol{X})]$, which is unknown. Therefore the optimal importance distribution is unknown (principle for an adaptive method).

- Example: We want to estimate

$$I = \mathbb{E}[h(X)]$$

with $X \sim \mathcal{N}(0,1)$ and $h(x) = \mathbf{1}_{[4,\infty)}(x)$.

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathbf{1}_{[4,\infty)}(x) e^{-\frac{x^2}{2}} dx = \frac{1}{2}\mathrm{erfc}\Big(\frac{4}{\sqrt{2}}\Big) \simeq 3.17\,10^{-5}$$

Monte Carlo: With a sample $(X_k)_{k=1,\dots,n}$ with the original distribution $\mathcal{N}(0,1)$.

$$\hat{I}_n = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k \geq 4}, \quad X_k \sim \mathcal{N}(0,1)$$

We have $\mathrm{Var}(\hat{I}_n) = \frac{1}{n} 3.17\,10^{-5}$.

Importance sampling: With a sample $(X_k)_{k=1,\dots,n}$ with the distribution $\mathcal{N}(4,1)$.

$$\hat{I}_n = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k \geq 4} \frac{e^{-\frac{X_k^2}{2}}}{e^{-\frac{(X_k-4)^2}{2}}} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k \geq 4} e^{-4X_k+8}, \quad X_k \sim \mathcal{N}(4,1)$$

We have $\mathrm{Var}(\hat{I}_n) = \frac{1}{n} 5.53\,10^{-8}$.

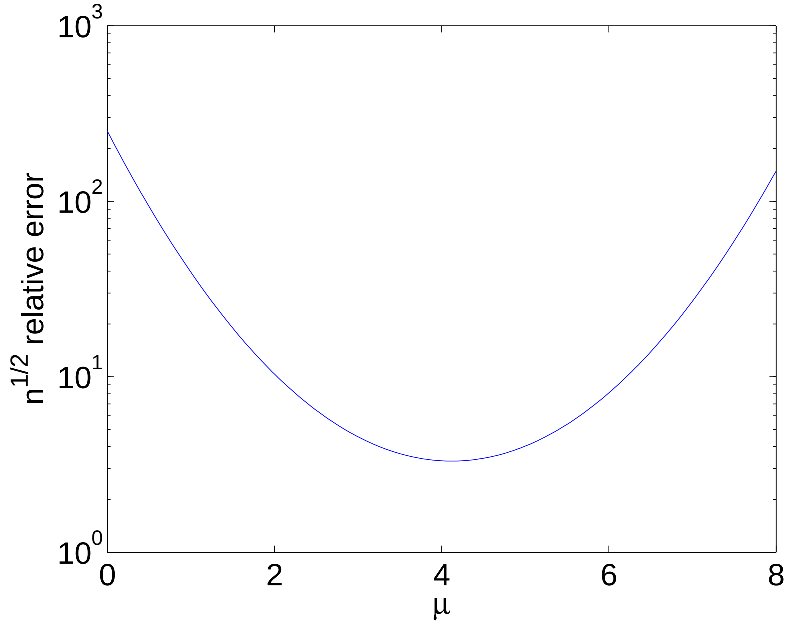
The IS method needs 1000 times less simulations to reach the same precision as MC !

Warning: we should not bias too much.

Importance sampling: With a sample $(X_k)_{k=1,\dots,n}$ with the distribution $\mathcal{N}(\mu,1)$.

$$\hat{I}_n = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k \geq 4} \frac{e^{-\frac{X_k^2}{2}}}{e^{-\frac{(X_k-\mu)^2}{2}}} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k \geq 4} e^{-\mu X_k + \frac{\mu^2}{2}}, \qquad X_k \sim \mathcal{N}(\mu,1)$$

We have $\mathrm{Var}(\hat{I}_n) = \frac{1}{n}\frac{e^{\mu^2}}{2}\mathrm{erfc}\big(\frac{4+\mu}{\sqrt{2}}\big) - \frac{1}{n}I^2$, which gives the normalized relative error $\sqrt{n}\mathbb{E}[(\hat{I}_n - I)^2]^{1/2}/I$:



If the bias is too large, the fluctuations of the likelihood ratios become large.

- Example: we want to estimate

$$I = \mathbb{E}[h(X)]$$

with $X \sim \mathcal{N}(0,1)$ and $h(x) = \exp(x)$.

$$I = \frac{1}{\sqrt{2\pi}} \int e^x e^{-\frac{x^2}{2}} dx = e^{\frac{1}{2}}$$

The large values of X are important.

Importance sampling: With a sample $(X_k)_{k=1,\dots,n}$ with the distribution $\mathcal{N}(\mu,1)$, $\mu > 0$.

$$\hat{I}_n = \frac{1}{n} \sum_{k=1}^{n} h(X_k) \frac{e^{-\frac{[X_k]^2}{2}}}{e^{-\frac{[X_k-\mu]^2}{2}}} = \frac{1}{N} \sum_{k=1}^{n} h(X_k) e^{-\mu X_k + \frac{\mu^2}{2}}$$

$$\mathrm{Var}(\hat{I}_n) = \frac{1}{n} \left(e^{\mu^2 - 2\mu + 2} - e^1 \right)$$

Monte Carlo $\mu = 0$: $\mathrm{Var}(\hat{I}_n) = \frac{1}{n} \left(e^2 - e^1 \right)$

Optimal importance sampling $\mu = 1$: $\mathrm{Var}(\hat{I}_n) = 0$.

Control variates

- The goal is to estimate $I = \mathbb{E}[h(\boldsymbol{X})]$ for $\boldsymbol{X}$ a random vector and $h(\boldsymbol{x}) = g(f(\boldsymbol{x}))$ a deterministic function.

- Assume that we have a reduced model $f_{\rm r}(\boldsymbol{X})$.

- Importance sampling method: first we evaluate (we approximate) the optimal density $q_{\rm opt,r}(\boldsymbol{x}) = \frac{g(f_{\rm r}(\boldsymbol{x}))p(\boldsymbol{x})}{I_{\rm r}}$, with $I_{\rm r} = \int g(f_{\rm r}(\boldsymbol{x}))p(\boldsymbol{x})d\boldsymbol{x}$, then we use it as a biased density for estimating I (dangerous, use conservative version).

- Control variates method:

We denore $h(\boldsymbol{x}) = g(f(\boldsymbol{x}))$, $h_{\rm r}(\boldsymbol{x}) = g(f_{\rm r}(\boldsymbol{x}))$.
We assume that we know $I_{\rm r} = \mathbb{E}[h_{\rm r}(\boldsymbol{X})]$.

By considering the representation

$$I = \mathbb{E}[h(\boldsymbol{X})] = I_{\rm r} + \mathbb{E}[h(\boldsymbol{X}) - h_{\rm r}(\boldsymbol{X})]$$

we propose the estimator:

$$\hat{I}_n = I_{\rm r} + \frac{1}{n}\sum_{k=1}^{n} h(\boldsymbol{X}_k) - h_{\rm r}(\boldsymbol{X}_k),$$

where $(\boldsymbol{X}_k)_{k=1,\ldots,n}$ is a n-sample (with the pdf p).

- Estimator:

$$\hat{I}_n = I_{\rm r} + \frac{1}{n}\sum_{k=1}^n h(\boldsymbol{X}_k) - h_{\rm r}(\boldsymbol{X}_k)$$

- The estimator is unbiased:

$$\begin{aligned}\mathbb{E}\big[\widehat{I}_n\big] &= I_{\rm r} + \frac{1}{n}\sum_{k=1}^n \mathbb{E}\big[h(\boldsymbol{X}_k) - h_{\rm r}(\boldsymbol{X}_k)\big] = I_{\rm r} + \mathbb{E}[h(\boldsymbol{X})] - \mathbb{E}[h_{\rm r}(\boldsymbol{X})] \\ &= I_{\rm r} + \mathbb{E}[h(\boldsymbol{X})] - I_{\rm r} = I\end{aligned}$$

- The estimator is convergent:

$$\widehat{I}_n \stackrel{n\to\infty}{\longrightarrow} I_{\rm r} + \mathbb{E}\big[h(\boldsymbol{X}) - h_{\rm r}(\boldsymbol{X})\big] = I$$

- The variance of the estimator is:

$$\mathrm{Var}(\widehat{I}_n) = \frac{1}{n}\mathrm{Var}\big[h(\boldsymbol{X}) - h_{\rm r}(\boldsymbol{X})\big]$$

$\hookrightarrow$ The use of a reduced model can reduce the variance.

- Example: we want to estimate

$$I = \mathbb{E}[h(X)]$$

with $X \sim \mathcal{U}(0,1)$, $h(x) = \exp(x)$.
Result: $I = e - 1 \simeq 1.72$.

Monte Carlo.

$$\hat{I}_n = \frac{1}{n} \sum_{k=1}^{n} \exp[X_k]$$

Variance of the MC estimator$= \frac{1}{n}(2e-1) \simeq \frac{1}{n} 4.44$.

Control variates. Reduced model: $h_{\rm r}(x) = 1 + x$ (here $I_{\rm r} = \frac{3}{2}$). CV estimator:

$$\hat{I}_n = I_{\rm r} + \frac{1}{n} \sum_{k=1}^{n} \{\exp[X_k] - 1 - X_k\}$$

Variance of the CV estimator $= \frac{1}{n}(3e - \frac{e^2}{2} - \frac{53}{12}) \simeq \frac{1}{n} 0.044$.

The CV method needs 100 times less simulations to reach the same precision as MC !

- Application: estimation of

$$I = \mathbb{E}[g(f(\boldsymbol{X}))]$$

We have a reduced model $f_{\rm r}$ of the full code f. The ratio between the computational cost of one call of f and one call of $f_{\rm r}$ is $q > 1$.

Estimator

$$\hat{I}_n = \frac{1}{n_{\rm r}} \sum_{k=1}^{n_{\rm r}} h_{\rm r}(\tilde{\boldsymbol{X}}_k) + \frac{1}{n} \sum_{k=1}^{n} h(\boldsymbol{X}_k) - h_{\rm r}(\boldsymbol{X}_k)$$

with $n_{\rm r} > n$, $h(\boldsymbol{x}) = g(f(\boldsymbol{x}))$, $h_{\rm r}(\boldsymbol{x}) = g(f_{\rm r}(\boldsymbol{x}))$.
Allocation between calls to the full code and calls to the reduced model can be optimized with the contraint $n_{\rm r}/q + n(1 + 1/q) = n_{\rm tot}$.
Classical trade-off between approximation error and estimation error.

Used when $f(\boldsymbol{X})$ is the solution of an ODE or PDE with fine grid, while $f_{\rm r}(\boldsymbol{X})$ is the solution obtained with a coarse grid (MultiLevel Monte Carlo).

• Not very useful for the estimation of the probability of a rare event.

Example: we want to estimate

$$I = \mathbb{P}(f(X) \geq 2.7)$$

with $X \sim \mathcal{U}(0,1)$, $f(x) = \exp(x)$.
Result: $I = 1 - \ln(2.7) \simeq 6.7\,10^{-3}$.

The reduced model $f_{\rm r}(x) = 1 + x$ is here useless.

The reduced model should be good in the important region.

Stratification

Principle: The sample $(\boldsymbol{X}_k)_{k=1,\ldots,n}$ is enforced to obey theoretical distributions in some "strata".
Method used in polls (representative sample).

Here: we want to estimate $\mathbb{E}[g(f(\boldsymbol{X}))]$, $\boldsymbol{X}$ with values in D.

• Two ingredients:
i) A partition of the state space D: $D = \bigcup_{i=1}^m D_i$. We know $p_i = \mathbb{P}(\boldsymbol{X} \in D_i)$.
ii) Total probability formula:

$$I = \mathbb{E}[g(f(\boldsymbol{X}))] = \sum_{i=1}^m \underbrace{\mathbb{E}[g(f(\boldsymbol{X}))|\boldsymbol{X} \in D_i]}_{J^{(i)}} \underbrace{\mathbb{P}(\boldsymbol{X} \in D_i)}_{p_i}$$

• Estimation:
1) For all $i = 1, \ldots, m$, I_i is estimated by Monte Carlo with n_i simulations:

$$\widehat{J}^{(i)}_{n_i} = \frac{1}{n_i} \sum_{j=1}^{n_i} g(f(\boldsymbol{X}^{(i)}_j)), \qquad \boldsymbol{X}^{(i)}_j \sim \mathcal{L}(\boldsymbol{X}|\boldsymbol{X} \in D_i)$$

2) The estimator is

$$\widehat{I}_n = \sum_{i=1}^m \widehat{J}^{(i)}_{n_i} p_i$$

$$\widehat{I}_n = \sum_{i=1}^{m} p_i \widehat{J}^{(i)}_{n_i}, \qquad \widehat{J}^{(i)}_{n_i} = \frac{1}{n_i} \sum_{j=1}^{n_i} g(f(\boldsymbol{X}^{(i)}_j)), \quad \boldsymbol{X}^{(i)}_j \sim \mathcal{L}(\boldsymbol{X} | \boldsymbol{X} \in D_i)$$

The total number of simulations is $n = \sum_{i=1}^{m} n_i$.

- The estimator is unbiased, convergent and its variance is

$$\text{Var}\big(\widehat{g_n}\big)_S = \sum_{i=1}^{m} p_i^2 \text{Var}\big(\widehat{J}^{(i)}_{n_i}\big) = \sum_{i=1}^{m} p_i^2 \frac{\sigma_i^2}{n_i}, \text{ with } \sigma_i^2 = \text{Var}\big(g(f(\boldsymbol{X}))|\boldsymbol{X} \in D_i\big)$$

The user is free to choose the allocations n_i (with the constraint $\sum_{i=1}^{m} n_i = n$).

- **Proportional stratification**: $n_i = p_i n$.

$$\widehat{I}_n = \sum_{i=1}^{m} \frac{p_i}{n_i} \sum_{j=1}^{n_i} g(f(\boldsymbol{X}^{(i)}_j)) = \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n_i} g(f(\boldsymbol{X}^{(i)}_j)), \qquad \boldsymbol{X}^{(i)}_j \sim \mathcal{L}(\boldsymbol{X} | \boldsymbol{X} \in D_i)$$

Then

$$\text{Var}\big(\widehat{I}_n\big)_{SP} = \frac{1}{n} \sum_{i=1}^{m} p_i \sigma_i^2$$

We have:

$$\text{Var}\big(\widehat{I}_n\big)_{MC} = \frac{1}{n} \text{Var}\big(g(f(\boldsymbol{X}))\big) \geq \frac{1}{n} \sum_{i=1}^{m} p_i \sigma_i^2 = \text{Var}\big(\widehat{I}_n\big)_{SP}$$

Proof: We have

$$\begin{aligned}
\mathbb{E}[h(\boldsymbol{X})]^2 &= \Big(\sum_{i=1}^{m} p_i \mathbb{E}[h(\boldsymbol{X})|\boldsymbol{X} \in D_i] \Big)^2 \\
&\leq \Big(\sum_{i=1}^{m} p_i \Big) \Big(\sum_{i=1}^{m} p_i \mathbb{E}[h(\boldsymbol{X})|\boldsymbol{X} \in D_i]^2 \Big) \\
&= \sum_{i=1}^{m} p_i \mathbb{E}[h(\boldsymbol{X})|\boldsymbol{X} \in D_i]^2
\end{aligned}$$

Therefore

$$\begin{aligned}
\sum_{i=1}^{m} p_i \sigma_i^2 &= \sum_{i=1}^{m} p_i \Big(\mathbb{E}[h(\boldsymbol{X})^2|\boldsymbol{X} \in D_i] - \mathbb{E}[h(\boldsymbol{X})|\boldsymbol{X} \in D_i]^2 \Big) \\
&= \mathbb{E}[h(\boldsymbol{X})^2] - \sum_{i=1}^{m} p_i \mathbb{E}[h(\boldsymbol{X})|\boldsymbol{X} \in D_i]^2 \\
&\leq \mathbb{E}[h(\boldsymbol{X})^2] - \mathbb{E}[h(\boldsymbol{X})]^2 = \mathrm{Var}(h(\boldsymbol{X}))
\end{aligned}$$

However, the proportional allocation is not optimal !

- The optimal allocation is the one that minimizes the variance $\mathrm{Var}(\widehat{I}_n)_S = \sum_{i=1}^m p_i^2 \frac{\sigma_i^2}{n_i}$.

It is the solution of the minimization problem: find $(n_i)_{i=1,\dots,m}$ minimizing

$$\sum_{i=1}^m p_i^2 \frac{\sigma_i^2}{n_i} \text{ with the constraint } \sum_{i=1}^m n_i = n$$

Solution (optimal allocation, obtained with Lagrange multiplier method):

$$n_i = n \frac{p_i \sigma_i}{\sum_{l=1}^m p_l \sigma_l}$$

The minimal variance is

$$\mathrm{Var}\big(\widehat{I}_n\big)_{SO} = \frac{1}{n} \left(\sum_{i=1}^m p_i \sigma_i \right)^2 ,$$

We have:

$$\mathrm{Var}\big(\widehat{I}_n\big)_{SO} \leq \mathrm{Var}\big(\widehat{I}_n\big)_{SP} \leq \mathrm{Var}\big(\widehat{I}_n\big)_{MC}$$

Practically: the σ_i's are unknown, therefore the optimal allocation is unknown (principle of an adaptive method).

- Example: we want to estimate

$$\mathbb{E}[g(f(X))]$$

with $X \sim \mathcal{U}(-1,1)$, $f(x) = \exp(x)$ and $g(y) = y$.

Result: $\mathbb{E}[f(X)] = \sinh(1) \simeq 1.18$.

Monte Carlo. With a sample X_1, ..., X_n with the distribution $\mathcal{U}(-1,1)$

$$\widehat{I}_n = \frac{1}{n}\sum_{k=1}^{n} \exp[X_k]$$

Variance of the estimator $= \frac{1}{n}(\frac{1}{2} - \frac{e^{-2}}{2}) \simeq \frac{1}{n}0.43$.

Proportional stratification. With a sample
- X_1, ..., $X_{n/2}$ with the distribution $\mathcal{U}(-1,0)$,
- $X_{n/2+1}$, ..., X_n with the distribution $\mathcal{U}(0,1)$.

$$\widehat{I}_n = \frac{1}{n}\sum_{k=1}^{n/2} \exp[X_k] + \frac{1}{n}\sum_{k=n/2+1}^{n} \exp[X_k] = \frac{1}{n}\sum_{k=1}^{n} \exp[X_k]$$

Variance of the PS estimator $\simeq \frac{1}{n}0.14$.
The PS method needs 3 times less simulations to reach the same precision as MC.

Nonproportional stratification. With a sample
- X_1, ..., $X_{n/4}$ with the distribution $\mathcal{U}(-1,0)$,
- $X_{n/4+1}$, ..., X_n with the distribution $\mathcal{U}(0,1)$.

$$\widehat{I}_n = \frac{2}{n}\sum_{k=1}^{n/4} \exp[X_k] + \frac{1}{2n}\sum_{k=n/4+1}^{n} \exp[X_k]$$

Variance of the estimator $\simeq \frac{1}{n}0.048$.
The stratification method needs 9 times less simulations to reach the same precision as MC.

Antithetic variables

- We want to compute

$$I = \int_{[0,1]^d} h(\boldsymbol{x}) d\boldsymbol{x}$$

Monte Carlo with a n-sample $(\boldsymbol{X}_1, \dots, \boldsymbol{X}_n)$ with the distribution $\mathcal{U}([0,1]^d)$:

$$\hat{I}_n = \frac{1}{n} \sum_{k=1}^{n} h(\boldsymbol{X}_k)$$

$$\mathbb{E}\big[(\hat{I}_n - I)^2\big] = \frac{1}{n} \mathrm{Var}(h(\boldsymbol{X})) = \frac{1}{n} \Big(\int_{[0,1]^d} h^2(\boldsymbol{x}) d\boldsymbol{x} - I^2 \Big)$$

- We consider the representations

$$I = \int_{[0,1]^d} h(1-\boldsymbol{x}) d\boldsymbol{x} \text{ and } I = \int_{[0,1]^d} \frac{h(\boldsymbol{x}) + h(1-\boldsymbol{x})}{2} d\boldsymbol{x}$$

Monte Carlo with a $n/2$-sample $(\boldsymbol{X}_1, \dots, \boldsymbol{X}_{n/2})$ with the distribution $\mathcal{U}([0,1]^d)$:

$$\tilde{I}_n = \frac{1}{n} \sum_{k=1}^{n/2} h(\boldsymbol{X}_k) + h(1-\boldsymbol{X}_k)$$

- Monte Carlo estimator with the sample $(\tilde{\boldsymbol{X}}_1,\ldots,\tilde{\boldsymbol{X}}_n) := (\boldsymbol{X}_1,\ldots,\boldsymbol{X}_{n/2}, 1-\boldsymbol{X}_1,\ldots,1-\boldsymbol{X}_{n/2})$ that is not i.i.d.:

$$\tilde{I}_n = \frac{1}{n}\sum_{k=1}^{n} h(\tilde{\boldsymbol{X}}_k)$$

The function f is called n times.

- Error:

$$\begin{aligned}
\mathbb{E}\big[(\tilde{I}_n - I)^2\big] &= \frac{1}{n}\Big(\mathrm{Var}(h(\boldsymbol{X})) + \mathrm{Cov}(h(\boldsymbol{X}), h(1-\boldsymbol{X}))\Big) \\
&= \frac{1}{n}\Big(\int_{[0,1]^d} h^2(\boldsymbol{x}) + h(\boldsymbol{x})h(1-\boldsymbol{x})d\boldsymbol{x} - 2I^2\Big)
\end{aligned}$$

The variance is reduced if $\mathrm{Cov}(h(\boldsymbol{X}), h(1-\boldsymbol{X})) < 0$. Sufficient condition: h is monotoneous.
Proof: If $\boldsymbol{X}, \boldsymbol{X}'$ i.i.d.

$$\begin{aligned}
[h(\boldsymbol{X}) - h(\boldsymbol{X}')][-h(1-\boldsymbol{X}) + h(1-\boldsymbol{X}')] &\geq 0 \text{ a.s.} \\
-2\mathbb{E}\big[h(\boldsymbol{X})h(1-\boldsymbol{X})\big] + 2\mathbb{E}\big[h(\boldsymbol{X})\big]^2 &\geq 0
\end{aligned}$$

- Example:

$$I = \int_0^1 \frac{1}{1+x} dx$$

Result: $I = \ln 2$.

Monte Carlo:

$$\hat{I}_n = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{1+X_k}$$

$\mathrm{Var}(\hat{I}_n) = \frac{1}{n}\big(\int_0^1 (1+x)^{-2} dx - \ln 2^2\big) = \frac{1}{n}\big(\frac{1}{2} - \ln 2^2\big) \simeq \frac{1}{n} 1.95\, 10^{-2}$.

Antithetic variables:

$$\tilde{I}_n = \frac{1}{n} \sum_{k=1}^{n/2} \frac{1}{1+X_k} + \frac{1}{2-X_k}$$

$\mathrm{Var}(\tilde{I}_n) = \frac{2}{n}\Big(\int_0^1 \big(\frac{1}{2(1+x)} + \frac{1}{2(2-x)}\big)^2 dx - \ln 2^2\Big) \simeq \frac{1}{n} 1.2\, 10^{-3}$.

The AV method requires 15 times less simulations than MC.

- More generally: one needs to find a pair $(\boldsymbol{X}, \tilde{\boldsymbol{X}})$ such that $h(\boldsymbol{X})$ and $h(\tilde{\boldsymbol{X}})$ have the same distribution and $\mathrm{Cov}(h(\boldsymbol{X}), h(\tilde{\boldsymbol{X}})) < 0$.

- Monte Carlo with an i.i.d. sample $((\boldsymbol{X}_1, \tilde{\boldsymbol{X}}_1), \ldots, (\boldsymbol{X}_{n/2}, \tilde{\boldsymbol{X}}_{n/2}))$:

$$\tilde{I}_n = \frac{1}{n} \sum_{k=1}^{n/2} h(\boldsymbol{X}_k) + h(\tilde{\boldsymbol{X}}_k)$$

$$\mathbb{E}\big[(\tilde{I}_n - I)^2\big] = \frac{1}{n}\Big(\mathrm{Var}(h(\boldsymbol{X})) + \mathrm{Cov}(h(\boldsymbol{X}), h(\tilde{\boldsymbol{X}}))\Big)$$

- Recent application: computation of effective tensors in stochastic homogenization (the effective tensor is the expectation of a functional of the solution of an elliptic PDE with random coefficients; antithetic pairs of the realizations of the composite medium are sampled; gain by a factor 3; in fact, better results with control variates; cf C. Le Bris, F. Legoll).

- Not very useful for the estimation of probabilities of rare events.

Low-discrepancy sequences (quasi Monte Carlo)

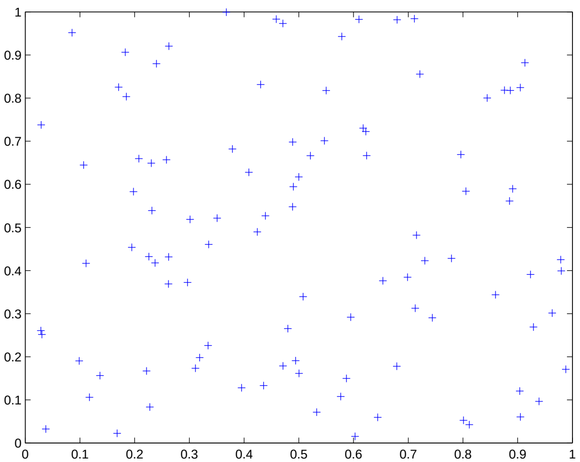
• The sample $(\boldsymbol{X}_k)_{k=1,\ldots,n}$ is selected so as to fill the (random) gaps that appear in a MC sample (for a uniform sampling of an hypercube).

• This technique
- can reduce the variance if h has good properties (bounded variation in the sense of Hardy and Krause); the asymptotic variance can be of the form $C_d(\log n)^{s(d)}/n^2$,
- can be applied in low-moderate dimension,
- can be viewed as a compromise between quadrature and MC.

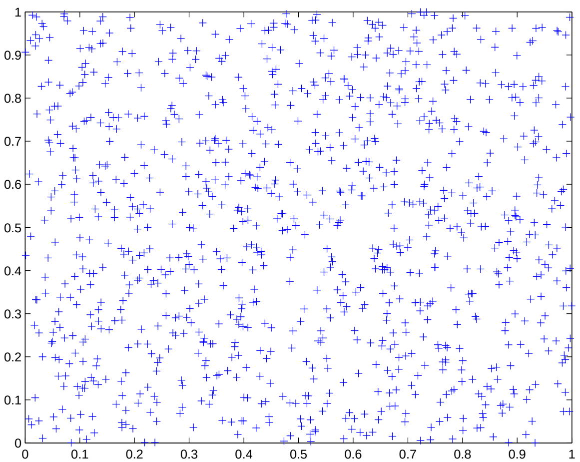
• A few properties:
- the error estimate is deterministic, but often not precise (Koksma-Hlawka inequality),
- the independence property is lost ($\rightarrow$ it is not easy to add points),
- the method is not adapted for the estimation of the probability of a rare event.

Cf lecture by G. Pagès.

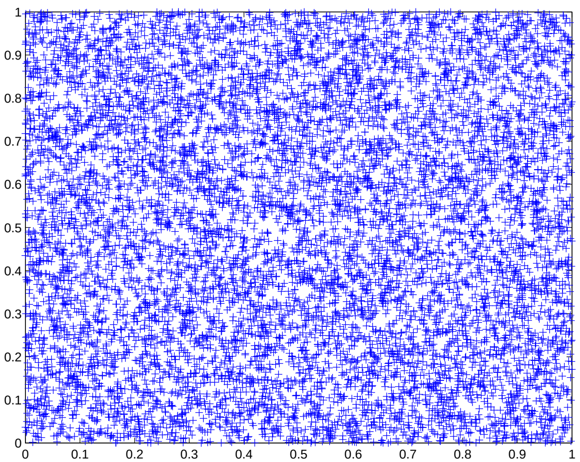
Example: Monte Carlo sample.



$n = 100$

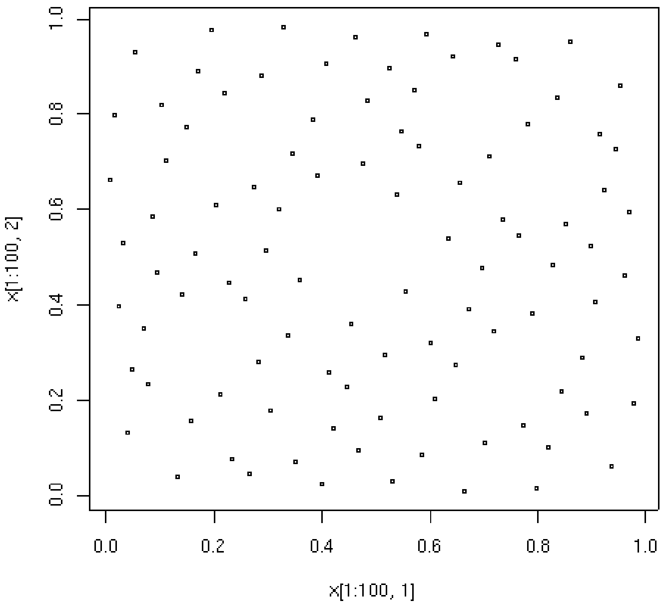


$n = 1000$

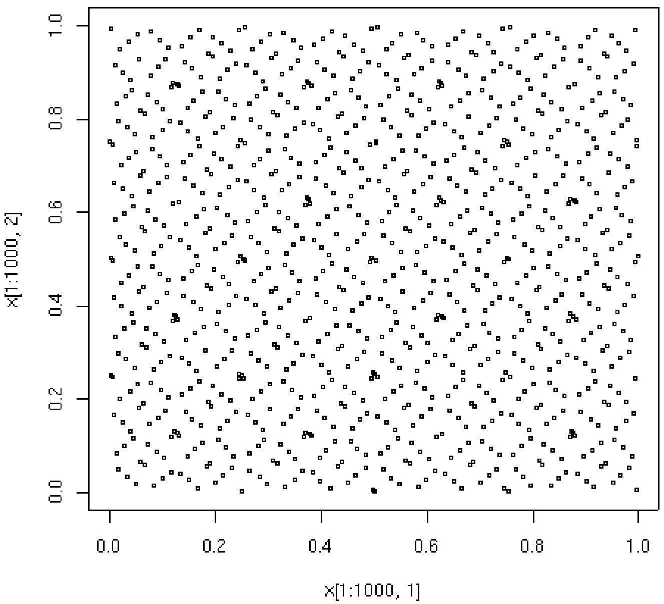


$n = 10000$

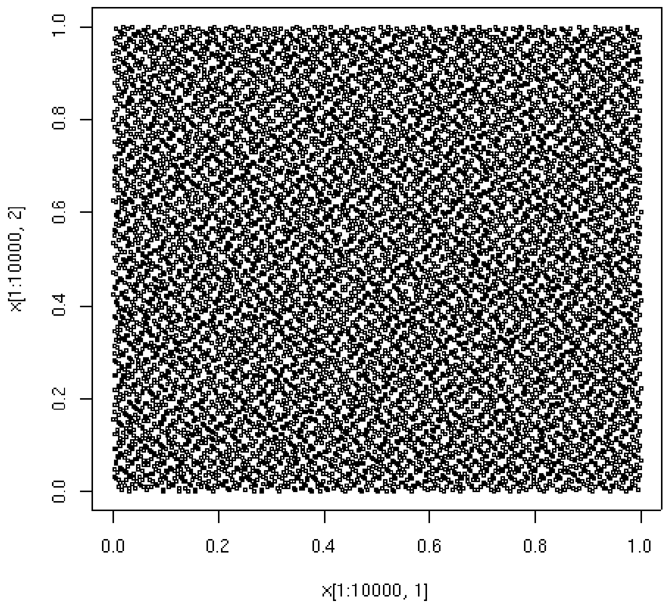
Example: Sobol sequence in dimension 2.



$n = 100$



$n = 1000$



$n = 10000$