Estimation of a threshold exceeding probability by using the convex order

JDS 2017

Lucie Bernard

Philippe Leduc - Florent Malrieu

LSTA - STMicrolectronics - LMPT

30 mai 2017

Framework

- Let $\mathbf{X} \in \mathbb{R}^d$ be a **random variable** and $\mathbf{X} \backsim \mathrm{P}_{\mathbf{X}}$.
- Let $g : \mathbb{R}^d \to \mathbb{R}$ be an **expensive to evaluate black-box function**.
- **Goal**: Estimate the **probability** p of $g(\mathbf{X})$ **exceeding a threshold** T,

$$p = \mathbb{P}(g(\mathbf{X}) \geqslant T) = \int_{\mathbb{R}^d} \mathbb{1}_{g(\mathbf{x}) \geqslant T} \mathrm{P}_{\mathbf{X}}(d\mathbf{x}),$$

from a **restricted number of observations** $\{g(\mathbf{x}_1), \ldots, g(\mathbf{x}_n)\}$ of the random variable $g(\mathbf{X})$.

Gaussian process regression

Basic ideas

- The function g is supposed to be a **realization of a Gaussian process** $\xi = \{\xi(\mathbf{x})\}_{\mathbf{x}\in\mathbb{R}^d}$.

- Given the observations $\{g(\mathbf{x}_1), \ldots, g(\mathbf{x}_n)\}$, the **posterior probability distribution** of $\xi(\mathbf{x})$ is still **Gaussian**, i.e.

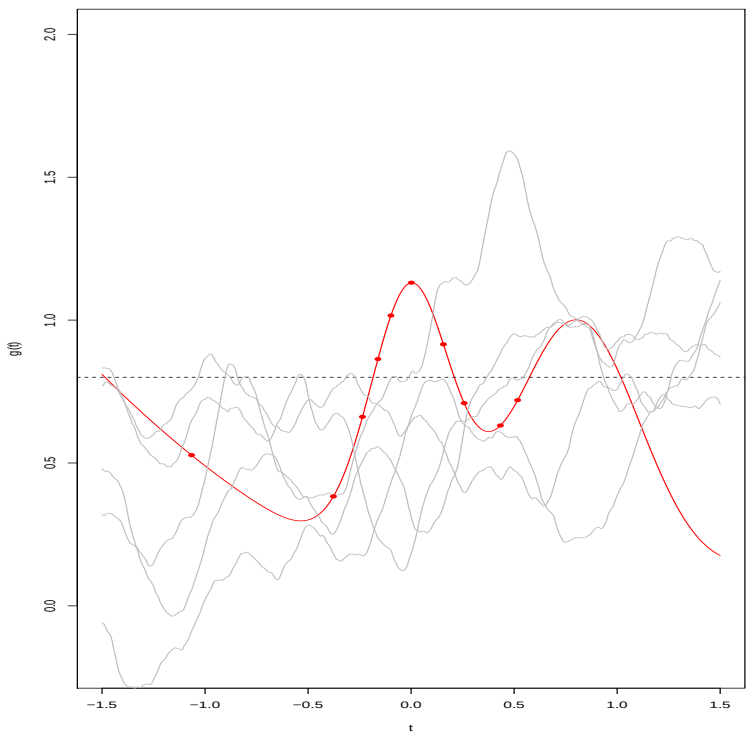
$$\forall\ \mathbf{x} \in \mathbb{R}^d, \quad [\xi(\mathbf{x}) \mid \xi(\mathbf{x}_1) = g(\mathbf{x}_1), \ldots, \xi(\mathbf{x}_n) = g(\mathbf{x}_n)] \sim \mathcal{N}(m_n(\mathbf{x}), \sigma_n^2(\mathbf{x})),$$

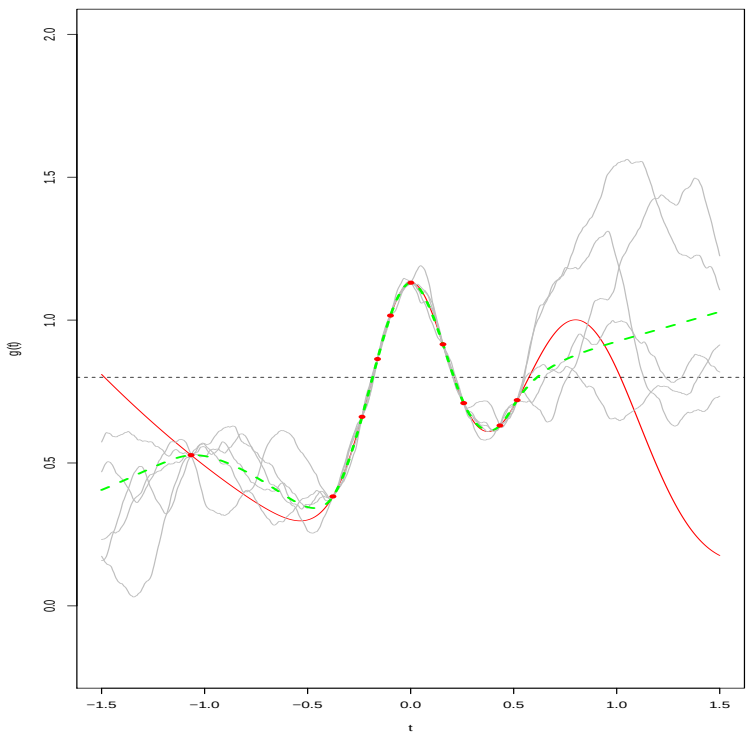
  where the posterior mean and variance $m_n(\mathbf{x})$ and $\sigma_n^2(\mathbf{x})$ can be easily computed.

- **Each realization** of the conditioned process is an **interpolation model** for g.

Example in dimension $d = 1$

- $T = 0.8$, $n = 10$, $\mathrm{P}_{\mathbf{X}} = \mathcal{N}(0, 0.4^2)$.
- The true function is $g(\mathbf{x}) = (0.4\mathbf{x} - 0.3)^2 + e^{-11.534|\mathbf{x}|^{1.95}} + e^{-5(\mathbf{x}-0.8)^2}$.





Left: Simulations of realizations of ξ with given unconditional mean and variance functions.

Right: Conditional simulations and mean function m_n .

A first estimation

As a result, the probability

$$p = \int_{\mathbb{R}^d} \mathbb{1}_{g(\mathbf{x}) \geqslant T} \mathrm{P}_{\mathbf{X}}(d\mathbf{x})$$

is a **realization of the random variable** $P_n \in [0, 1]$ defined by

$$P_n = \mathbb{P}\left(\xi_n(\mathbf{X}) \geqslant T \mid \xi_n\right) = \int_{\mathbb{X}} \mathbb{1}_{\xi_n(\mathbf{x}) \geqslant T} \mathrm{P}_{\mathbf{X}}(d\mathbf{x}),$$

where ξ_n is the process conditioned on the n available observations.

A first estimation

- A **natural estimator** $\widehat{p}_n$ **of** p is the **mean value** of P_n:

$$\widehat{p}_n = \mathbb{E}[P_n] = \int_{\mathbb{X}} \mathbb{E}\left[\mathbb{1}_{\xi_n(\mathbf{x}) \geqslant T}\right] \mathrm{P}_{\mathbf{X}}(d\mathbf{x}) = \mathbb{E}\left[s_n(\mathbf{X})\right],$$

  where the function s_n is easy to evaluate because

$$s_n(\mathbf{x}) = \mathbb{E}\left[\mathbb{1}_{\xi_n(\mathbf{x}) \geqslant T}\right] = \mathbb{P}\left(\xi_n(\mathbf{x}) \geqslant T\right) = \Phi\left(\frac{m_n(\mathbf{x}) - T}{\sigma_n(\mathbf{x})}\right).$$

- Finally, the estimator is

$$\widehat{p}^{MC}_{N,P_n} = \frac{1}{N} \sum_{i=1}^{N} s_n(\mathbf{X}_i),$$

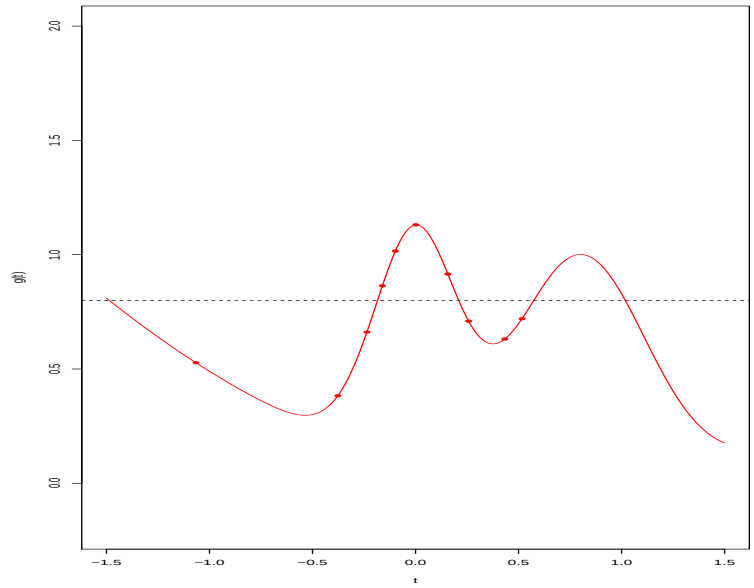
  where $\mathbf{X}_1, \ldots, \mathbf{X}_N \backsim \mathrm{P}_{\mathbf{X}}$.

Example in dimension $d = 1$

- We want to estimate the probability

$$p = \int_{\mathbb{R}} \mathbb{1}_{g(\mathbf{x}) \geqslant T} \mathrm{P}_{\mathbf{X}}(d\mathbf{x}),$$

where $\mathrm{P}_{\mathbf{X}} = \mathcal{N}(0, 0.4^2)$.



- We use a Monte Carlo sample of size $N = 10^4$,

$$\mathbf{X}_1, \ldots, \mathbf{X}_N \backsim \mathrm{P}_{\mathbf{X}}.$$

Estimator	Estimated value of p
$\widehat{p}_N^{MC} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{g(\mathbf{X}_i) \geqslant T}$	0,45
$\widehat{p}_{N,P_n}^{MC} = \frac{1}{N} \sum_{i=1}^{N} s_n(\mathbf{X}_i)$	0.42

Limits

- The distribution of P_n is unknown.
- **Simulating** P_n require realizations of a Gaussian process: it is **time consuming** and could lead to **numerical issues**.
- For example, the m-th moment of P_n is not easy to compute:

$$\mathbb{E}\left[P_n^m\right] = \int_{\mathbb{R}^d} \dots \int_{\mathbb{R}^d} \mathbb{P}\left(\bigcap_{i=1}^{m} \xi_n(\mathbf{x}_i) \geqslant T\right) \mathrm{P}_{\mathbf{X}}(d\mathbf{x}_1) \dots \mathrm{P}_{\mathbf{X}}(d\mathbf{x}_m).$$

$\hookrightarrow$ How to learn about the distribution of P_n ?

$\hookrightarrow$ For all $\alpha \in [0,1]$, what about the α-quantile $F_{P_n}^{-1}(\alpha)$ of P_n ?

An alternative estimation

- Let $U \in [0,1]$ be a real random variable.
- We introduce an **alternative random variable** $R_n = R_n(U)$ defined by

$$R_n = \mathbb{P}\left(s_n(\mathbf{X}) > U \mid U\right) = \int_{\mathbb{X}} \mathbb{1}_{s_n(\mathbf{x})>U} \, \mathrm{dP}_{\mathbf{X}}(\mathbf{x}).$$

Proposition

$$U \sim \mathcal{U}[0,1] \Leftrightarrow \mathbb{E}[P_n] = \mathbb{E}[R_n]$$

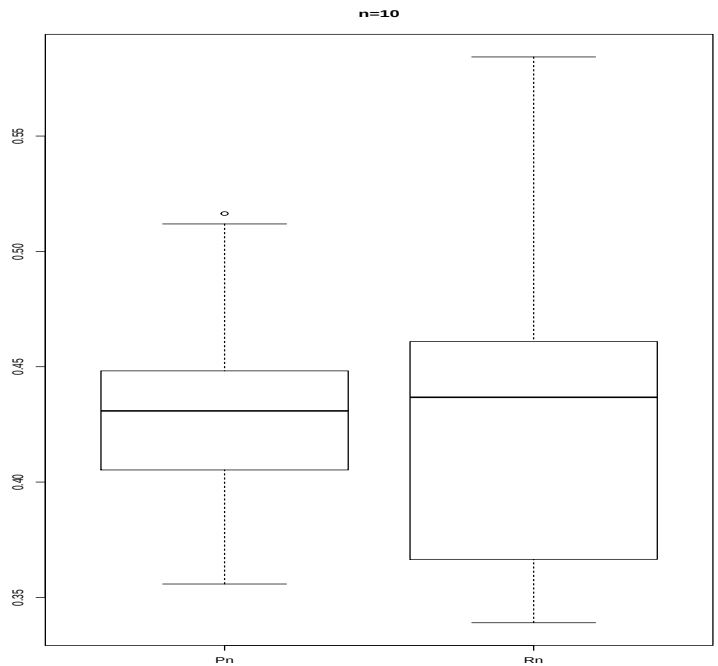
- The **mean value** of R_n is an **estimator** of p.

Example in dimension $d = 1$

- Necessary time to generate 10^3 realizations of P_n: 9.95 minutes.
- Necessary time to generate 10^3 realizations of R_n: 0.14 seconds.

- $M = 10^3$, $U_1, \ldots, U_M \sim \mathcal{U}[0,1]$.
- $N = 10^4$, $\mathbf{X}_1, \ldots, \mathbf{X}_N \backsim \mathrm{P}_{\mathbf{X}}$.
- For all $j = 1, \ldots, M$, $R_{n,j} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{s_n(\mathbf{X}_i) > U_j}$.

Estimator	Estimated value of p
$\widehat{p}_N^{MC} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{g(\mathbf{X}_i) \geqslant T}$	0,45
$\widehat{p}_{N,P_n}^{MC} = \frac{1}{N} \sum_{i=1}^{N} s_n(\mathbf{X}_i)$	0.42
$\widehat{p}_{N,R_n}^{MC} = \frac{1}{M} \sum_{j=1}^{M} R_{n,j}$	0.42



An alternative estimation

Proposition

$$U \sim \mathcal{U}[0,1] \Leftrightarrow P_n \leqslant_{cx} R_n.$$

- The inequality $P_n \leqslant_{cx} R_n$ means that for all **convex function** φ,

$$\mathbb{E}[\varphi(P_n)] \leqslant \mathbb{E}[\varphi(R_n)].$$

As a result,

- $\mathrm{Var}[P_n] \leqslant \mathrm{Var}[R_n]$.
- $\mathbb{E}[P_n^m] \leqslant \mathbb{E}[R_n^m]$, $m \geqslant 1$.
- $\dfrac{1}{1-\alpha}\displaystyle\int_\alpha^1 F_{P_n}^{-1}(t)\mathrm{d}t \leqslant \dfrac{1}{1-\alpha}\int_\alpha^1 F_{R_n}^{-1}(t)\mathrm{d}t$, $\alpha \in (0,1)$.

An alternative estimation

- Finally, by the **convex order** and **Markov inequality**, the α-quantile $F_{P_n}^{-1}(\alpha)$ of P_n satisfies

$$\max(0, \mathbb{E}[R_n] + \alpha - 1) \leqslant \frac{1}{\alpha}\int_0^\alpha F_{R_n}^{-1}(t)\mathrm{d}t \leqslant F_{P_n}^{-1}(\alpha) \leqslant \frac{1}{1-\alpha}\int_\alpha^1 F_{R_n}^{-1}(t)\mathrm{d}t \leqslant \min\left(1, \frac{\mathbb{E}[R_n]}{1-\alpha}\right).$$

- Explicit formulas for the **lower and upper bounds**

Proposition

For all $\alpha \in (0,1)$,

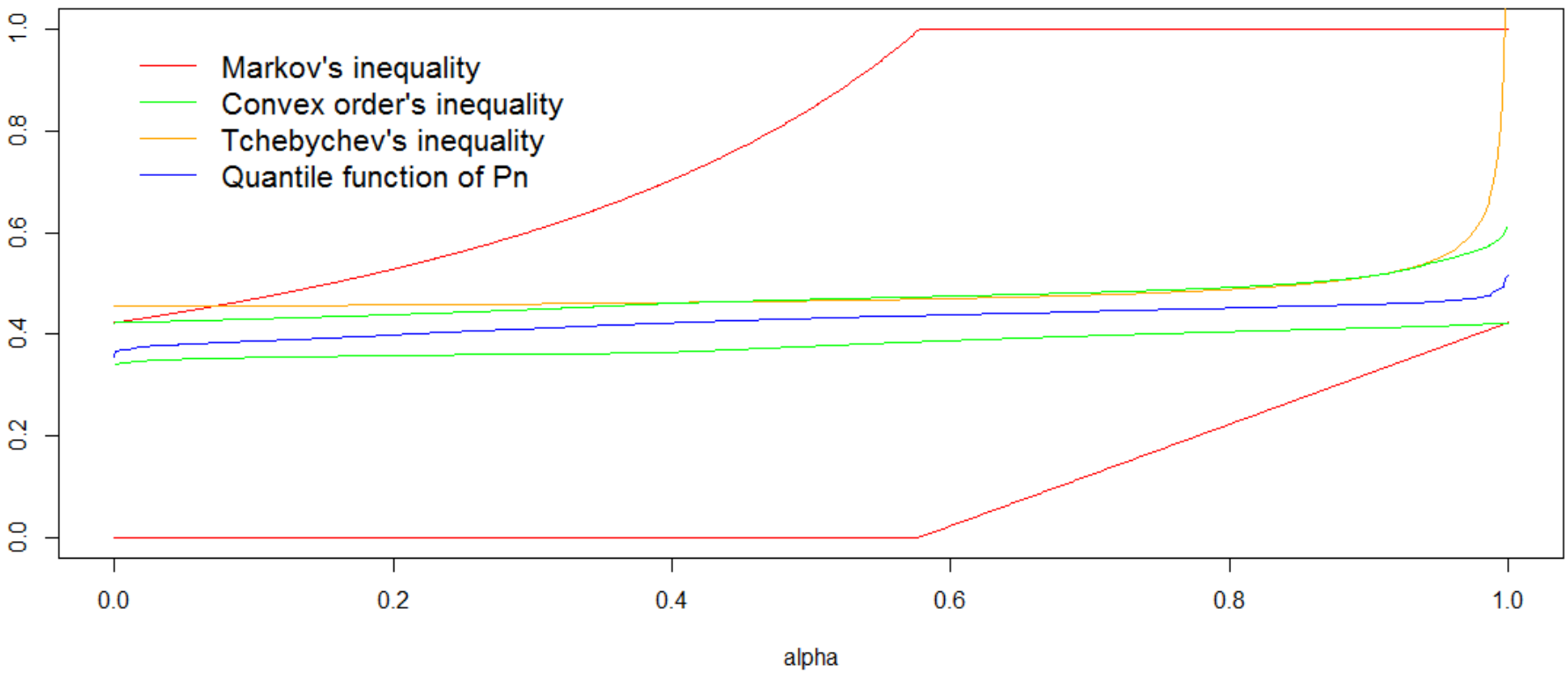
- $\frac{1}{\alpha}\int_0^\alpha F_{R_n}^{-1}(t)\mathrm{d}t = 1 - \int_{\mathbb{R}^d} \min\left(1, \frac{1-s_n(\mathbf{x})}{\alpha}\right) \mathrm{P}_{\mathbf{X}}(d\mathbf{x})$.
- $\frac{1}{1-\alpha}\int_\alpha^1 F_{R_n}^{-1}(t)\mathrm{d}t = \int_{\mathbb{R}^d} \min\left(1, \frac{s_n(\mathbf{x})}{1-\alpha}\right) \mathrm{P}_{\mathbf{X}}(d\mathbf{x})$.

Example in dimension $d = 1$, $n = 10$

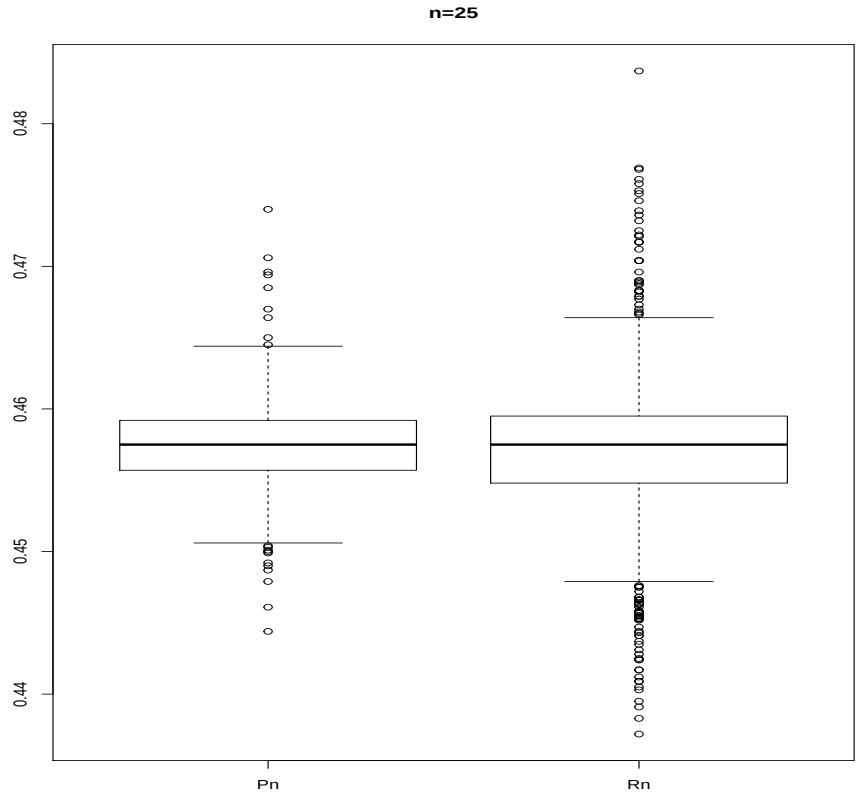
- The boundaries given by the convex order tell us that we have with probability $\geqslant 0,95$,

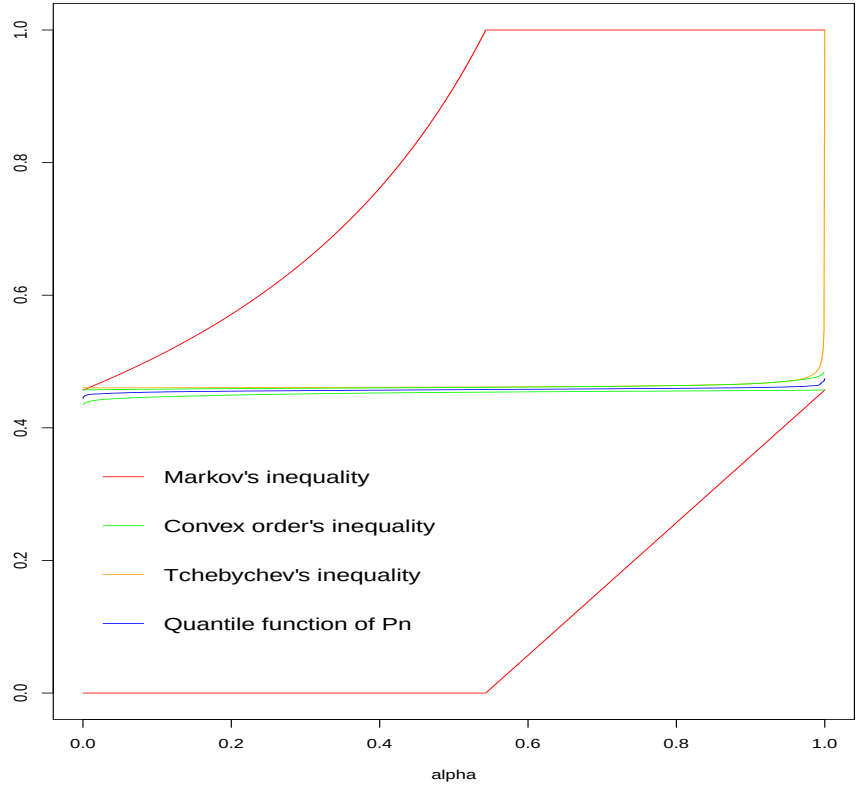
$$0,35 \leqslant P_n \leqslant 0,54.$$

- Tchebychev's inequality: $F_{P_n}^{-1}(\alpha) \leqslant \mathbb{E}[P_n] + \sqrt{\dfrac{\mathrm{Var}[P_n]}{1-\alpha}}$.



Exemple in dimension $d = 1$, $n = 25$





- The boundaries given by the convex order tell us that we have with probability $\geqslant 0,95$,

$$0,44 \leqslant P_n \leqslant 0,47.$$

- A sample Monte Carlo $\mathbf{X}_1, \ldots, \mathbf{X}_N \sim \mathrm{P}_{\mathbf{X}}$ of size $N = 10^6$ gives the estimate $\widehat{p}_N^{MC} = 0,448$.

Conclusion & Perspectives

- The distribution of the random variable P_n is unknown. We use an alternative random variable, which is easy to simulate and has the same mean value.

- We also prove that for all $m \in \mathbb{N}^*$,

$$P_n^m \leqslant_{cx} \int_{\mathbb{R}^d} \dots \int_{\mathbb{R}^d} \mathbb{1}_{\mathbb{P}\left(\bigcap_{i=1}^{m} \xi_n(\mathbf{x}_i) \geqslant T\right) > U} \mathrm{P}_{\mathbf{X}}(d\mathbf{x}_1) \dots \mathrm{P}_{\mathbf{X}}(d\mathbf{x}_m).$$

  Can we use this result to **refine** our credible interval ?

- One may use the **variance** of R_n in the context of **Stepwise Uncertainty Reduction strategies** and provide a criterion to find optimal locations for points of a design of experiments.

References

Y. Auffray, P. Barbillon, and J-M. Marin.

Bounding rare event probabilities in computer experiments.

Comput. Statist. Data Anal., 80 :153–166, 2014.

J. Oger, P. Leduc, and E Lesigne.

A random field model and decision support in industrial production.

J. SFdS, 156(3) :1–26, 2015.

Nicole Bäuerle and Alfred Müller.

Stochastic orders and risk measures : consistency and bounds.

Insurance Math. Econom., 38(1) :132–148, 2006.

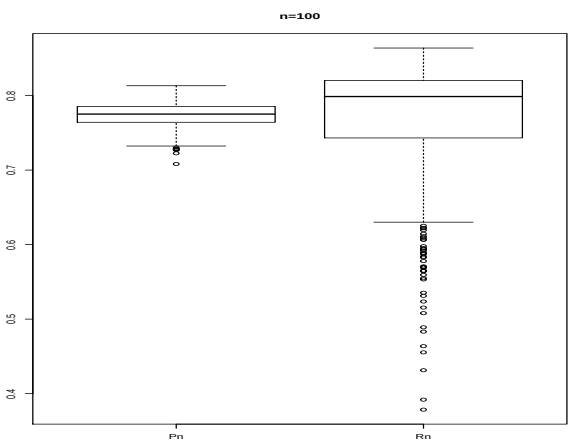
M. Shaked and J.G. Shanthikumar.

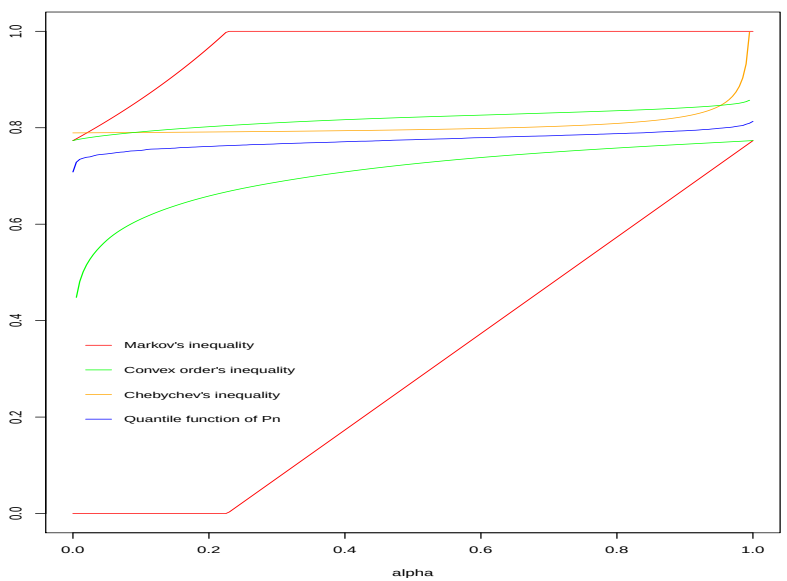
Stochastic Orders.

Springer Series in Statistics. Springer, New York, 2007.

Annexe : Example in dimension $d = 3$.

- $T = 1$, $n = 100$, $\mathrm{P}_{\mathbf{X}} = \mathcal{U}[-1, 1]$.
- $g(x_1, x_2, x_3) = 4.5 \sin(x_1) e^{-4x_2} - 0.6x_2^2 + 2.3(5 - x_3^2)$.
- A Monte Carlo sample of size $N = 10^4$ gives the estimate $\widehat{p}_N^{MC} = 0,8$.
- Necessary time to generate 10^3 realizations of $P_n \approx 20$ minutes.
- Necessary time to generate 10^3 realizations of $R_n \approx 0.32$ seconds.
- The mean value of the random variable P_n gives the estimates $\widehat{p}_{N,P_n}^{MC} = 0,77$.
- The boundaries given by the convex order tell us that we have with probability $\geqslant 0,95$, $0,57 \leqslant P_n \leqslant 0,86$.





Annexe : Example in dimension $d = 3$.

- $T = 1$, $n = 200$, $\mathrm{P}_{\mathbf{X}} = \mathcal{U}[-1, 1]$.
- $g(x_1, x_2, x_3) = 4.5 \sin(x_1) e^{-4x_2} - 0.6x_2^2 + 2.3(5 - x_3^2)$.
- A Monte Carlo sample of size $N = 10^4$ gives the estimate $\widehat{p}_N^{MC} = 0,8$.
- Necessary time to generate 10^3 realizations of $P_n \approx 20$ minutes.
- Necessary time to generate 10^3 realizations of $R_n \approx 0.32$ seconds.
- The mean value of the random variable P_n gives the estimates $\widehat{p}_{N,P_n}^{MC} = 0,79$.
- The boundaries given by the convex order tell us that we have with probability $\geqslant 0,95$, $0,68 \leqslant P_n \leqslant 0,82$.