Estimation of rare events probabilities in computer experiments

Yves Auffray
Dassault Aviation &
Département de Mathématiques,
Université Paris-Sud, France

Pierre Barbillon
INRIA Saclay, projet SELECT,
Département de Mathématiques,
Université Paris-Sud, France

Jean-Michel Marin [*][†]
Institut de Mathématiques et Modélisation de Montpellier
Université Montpellier 2

Abstract

We are interested in estimating probabilities of rare events in the context of computer experiments. These rare events depend on the output of a physical model with random input variables. Since the model is only known through an expensive black box function, standard efficient Monte Carlo estimates of rare events probabilities can not be used. We then propose two strategies to deal with this difficulty: a Bayesian estimate and an importance sampling method. Both proposals rely on Kriging metamodeling and are able to achieve sharp upper confidence bounds on the rare events probabilities. The variability due to the Kriging metamodeling step is properly taking into account.

The proposed methodologies are applied to a toy example and a real case study which consists of finding an upper bound of the probability that the trajectory of an airborne load collides the aircraft that has released it.

Keywords: computer experiments, rare events, Kriging, importance sampling, Bayesian estimates, risk assessment with fighter aircraft.

---

[*]Corresponding author: place Eugène Bataillon, Case Courrier 051, 34095 Montpellier cedex 5
[†]JEAN-MICHEL.MARIN@UNIV-MONTP2.FR

1 Introduction

Rare events are a major concern in reliability of complex systems (Heidelberg, 1995; Shahabuddin, 1995). We focus here on rare events depending on computer experiments. A computer experiment (Welch et al., 1992; Koehler and Owen, 1996) consists of an evaluation of a black box function which describes a physical model,

$$y = f(\mathbf{x}), \tag{1.1}$$

where $y \in \mathbb{R}$ and $\mathbf{x} \in E$ where E is a compact subset of $\mathbb{R}$. The code which computes f is expensive since the model is complex. We assume that no more than N calls to f are possible. The input $\mathbf{x}$ are measured with a lack of precision and some variables are uncontrollabe. Both sources of uncertainties are modeled by a random distribution on E. Let $\mathbf{X}$ be the random variable. Our goal is to estimate the probability:

$$\pi_\rho = \mathbb{P}(f(\mathbf{X}) < \rho)) = \mathbb{P}(\mathbf{X} \in R_\rho) = \mathbb{P}_{\mathbf{X}}(R_\rho),$$

where R_ρ is a subset of E defined by $R_\rho = \{\mathbf{x} : f(\mathbf{x}) < \rho\}$ and $\rho \in \mathbb{R}$ is a given threshold.

A crude Monte Carlo scheme leads to the following estimator of π_ρ:

$$\hat{\pi}_{\rho,N} = \frac{\Gamma(f, \mathbf{X}_{1:N}, \rho)}{N}, \tag{1.2}$$

where $\Gamma(f, \mathbf{X}_{1:N}, \rho)$ is defined by

$$\Gamma(f, \mathbf{X}_{1:N}, \rho) = \sum_{i=1}^{N} \mathbb{I}_{]-\infty,\rho[}(f(\mathbf{X}_i)), \tag{1.3}$$

and $\mathbf{X}_{1:N} = (\mathbf{X}_1, \ldots, \mathbf{X}_N)$ is a N-sample of random variables with the same distribution than $\mathbf{X}$. Its expectation and its variance are:

$$\mathbb{E}(\hat{\pi}_{\rho,N}) = \mathbb{P}(\mathbf{X} \in R_\rho) = \pi_\rho, \quad \mathbb{V}(\hat{\pi}_{\rho,N}) = \frac{1}{N}\pi_\rho(1 - \pi_\rho).$$

Hence, its relative error is $\dfrac{(\mathbb{V}(\hat{\pi}_{\rho,N}))^{1/2}}{\mathbb{E}(\hat{\pi}_{\rho,N})} \approx (\pi_\rho N)^{-1/2}$ when $\pi_\rho << \frac{1}{N}$. Therefore, the relative error can be very large. Furthermore, since $\Gamma(f, \mathbf{X}_{1:N}, \rho)$ follows a binomial distribution with parameters N and π_ρ, an exact confidence upper bound on π_ρ:

$$\mathbb{P}(\pi_\rho \leq b(\Gamma(f, \mathbf{X}_{1:N}, \rho), N, \alpha)) \geq 1 - \alpha,$$

is available.

Indeed, let T be a random variable which follows a binomial distribution with parameters N and p. For any real number $\alpha \in [0, 1]$, we can easily shown that the upper confidence bound b on p:

$$\mathbb{P}_T(p \leq b(T, N, \alpha)) \geq 1 - \alpha$$

is such that:

$$\begin{cases} b = 1 & \text{if } T = N \\ b \text{ is the solution of equation } \sum_{k=0}^{T} \binom{N}{k} b^k (1-b)^{N-k} = \alpha & \text{otherwise} \end{cases} \quad . \tag{1.4}$$

This upper bound is not in closed form but easily computable.

In the case where $\Gamma(f, \mathbf{X}_{1:N}, \rho) = 0$ which happens with probability $(1 - \pi_\rho)^N$, the $(1 - \alpha)$-confidence interval is $[0, 1 - (\alpha)^{1/N}]$. As an example, if the realization of $\Gamma(f, \mathbf{X}_{1:N}, \rho)$ is equal to 0, an upper confidence upper bound at level 0.9, $\pi_\rho \leq 10^{-5}$ can be warranted only if more than 230,000 calls to f were performed.

When the purpose is to assess the reliability of a system under the constrain of a limited number of calls to f, there is a need for a sharper upper bound on π_ρ. Several way to improve the precision of estimation have been proposed in the literature.

Since Monte Carlo estimation works better for frequent event, the first idea is to change the crude scheme in such a manner that the event becomes less rare. It is what importance sampling and splitting methods schemes try to achieve.

For example L'Ecuyer et al. (2007) showed that randomized quasi-Monte Carlo can be used jointly with splitting and/or importance sampling. By analysing a rare event as a cascade of intermediate less rare events, Del Moral and Garnier (2005) developped a genealogical particle system approach to explore the space of inputs E. Cérou and Guyader (2007a,b) proposed an adaptive multilevel splitting also based on particle systems. An adaptive directional sampling method is presented by Munoz Zuniga et al. (2010) to accelerate the Monte Carlo simulation method. These methods can still need too many calls to f and the importance distribution is hard to set for an importance sampling method.

A general approach in computer experiments is to make use of a metamodel which is a fast computing function which approximates f. It has to be built on the basis of data $\{f(\mathbf{x}_1), \cdots, f(\mathbf{x}_n)\}$ which are evaluations of f at points of a well chosen design $D_n = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$. The bet is that these n evaluations will allow to build more accurate estimators and bounds on the probability of the target event.

Kriging is such a metamodeling tool, one can see Santner et al. (2003) and more recently Li and Sudjianto (2005); Joseph (2006); Bingham et al. (2006). The function f is seen as a realization of a Gaussian process which is a Bayesian prior.

The related posterior distribution is computed conditionally to the data. It is still a Gaussian process whose mean can be used as a prediction of f everywhere on E and the variance as a pointwise measure of the accuracy of the prediction.

By using this mean and this variance, Oakley (2004) has developed a sequential method to estime quantiles and Vazquez and Bect (2009) a sequential method to estimate the probability of a rare event. Cannamela et al. (2008) have proposed some sampling strategies based only on a reduced model which is a coarse approximation of f (no information about the accuracy of prediction are given), to estimate quantiles.

3

Two approaches are investigated in that paper. Both rely on the hypothesis that f is a realization of a Gaussian process F independent of $\mathbf{X}$. As a consequence, π_ρ is a realization of the random variable:

$$\Pi_\rho = \mathbb{E}(\mathbb{1}_{]-\infty,\rho[}(F(\mathbf{X}))|F)\,.$$

The first approach consists of focusing on the posterior distribution of Π_ρ which depends on the posterior distribution of f given its computed evaluations. We show that a Bayesian estimator of Π_ρ can be computed and a credible bound is reachable by simulating Gaussian processes to obtain realizations of Π_ρ.

The other approach is an importance sampling method whose the importance distribution is based on the metamodel.

The paper is organized as follows: Section 2 describes the posterior distribution of the Gaussian process and how to obtain an estimator and a credible interval on Π_ρ. Section 3 presents the importance sampling method and the confidence upper bound which is provided with a high probability. Finally in Section 4, these methods are used on a toy example to ensure that they perform well and a solution to a real aeronautical case study about the risk that the trajectory of an airborne load collides the aircraft that has released it, is proposed.

2 Bayesian estimator and credible interval

The first step for Kriging metamodeling is to choose a design $D_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ of numerical experiments (one can see Morris and Mitchell (1995); Koehler and Owen (1996) and more recently Fang et al. (2006); Mease and Bingham (2006); Dette and Pepelyshev (2010)). Let $y_{D_n} = (y_1 = f(\mathbf{x}_1), \ldots, y_n = f(\mathbf{x}_n))$ be the evaluations of f on D_n.

Let us start from a statistical model consisting of Gaussian processes $F_{\boldsymbol{\beta},\sigma,\boldsymbol{\theta}}$ whose the expressions are given by: for $\mathbf{x} \in E$,

$$F_{\boldsymbol{\beta},\sigma,\boldsymbol{\theta}}(\mathbf{x}) = \sum_{k=1}^{L} \beta_j h_j(\mathbf{x}) + \zeta(\mathbf{x}) = H(\mathbf{x})^T \boldsymbol{\beta} + \zeta(\mathbf{x})\,, \tag{2.1}$$

where

- $h_1, \ldots, h_L$ are regression functions, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_L)$ is a vector a parameters,

- ζ is a centered Gaussian process with covariance

$$\mathrm{Cov}(\zeta(\mathbf{x}), \zeta(\mathbf{x}')) = \sigma^2 K_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}')\,,$$

  where $K_{\boldsymbol{\theta}}$ is a correlation function depending on some parameters $\boldsymbol{\theta}$ (for details about kernels, see Koehler and Owen, 1996).

The maximum likelihood estimates $\hat{\boldsymbol{\beta}}, \hat{\sigma}, \hat{\boldsymbol{\theta}}$ of $\boldsymbol{\beta}, \sigma, \boldsymbol{\theta}$ are computed on the basis of the observations. Then, the Bayesion prior on f is chosen to be $F = F_{\hat{\boldsymbol{\beta}},\hat{\sigma},\hat{\boldsymbol{\theta}}}$ and the process F is assumed independent of $\mathbf{X}$. We denote F^{D_n} the process F conditionally to $F(\mathbf{x}_1) = y_1, \ldots, F(\mathbf{x}_n) = y_n$, in short $Y_{D_n} = y_{D_n}$.

The process F^{D_n} is still a Gaussian process (see Santner et al., 2003) with

- mean: $\forall \mathbf{x}$,
$$m_{D_n}(\mathbf{x}) = H(\mathbf{x})^T \hat{\boldsymbol{\beta}} + \Sigma_{\mathbf{x}D_n}^T \Sigma_{D_n D_n}^{-1} (y_{D_n} - H_{D_n} \hat{\boldsymbol{\beta}}), \tag{2.2}$$

- covariance: $\forall \mathbf{x}, \mathbf{x}'$,
$$K_{D_n}(\mathbf{x}, \mathbf{x}') = \hat{\sigma}^2 (K_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, \mathbf{x}') - \Sigma_{\mathbf{x}D_n}^T \Sigma_{D_n D_n}^{-1} \Sigma_{\mathbf{x}'D_n}), \tag{2.3}$$

where
$$(\Sigma_{D_n D_n})_{1 \le i,j \le n} = K_{\hat{\boldsymbol{\theta}}}(\mathbf{x}_i, \mathbf{x}_j) \text{ and } \Sigma_{\mathbf{x}D_n} = (K_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, \mathbf{x}_i))_{1 \le i \le n}^T.$$

In this approach the conditioning to the data regard the parameters as fixed although they are estimated.

The Bayesian prior distribution $\mathbb{P}_F$ on f leads to a Bayesian prior distribution on Π_ρ. Our goal is to use the distribution of the posterior process F^{D_n} conditionally to the observation of Y_{D_n}, to learn about the posterior distribution of Π_ρ. The random variable $\Pi_\rho^{D_n}$ is defined by:
$$\Pi_\rho^{D_n} = \mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{X}))|F^{D_n}). \tag{2.4}$$

Its distribution is the posterior distribution of Π_ρ conditionally to $Y_{D_n} = y_{D_n}$, as the following useful lemma states.

Lemma 2.1. *For all measurable function $g : \mathbb{R} \mapsto \mathbb{R}$,*
$$\mathbb{E}(g(\Pi_\rho^{D_n})) = \mathbb{E}(g(\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{X}))|F^{D_n}))).$$

Proof

$$\begin{aligned}
\mathbb{E}(g(\Pi_\rho^{D_n})) &= \mathbb{E}(g(\Pi_\rho)|Y_{D_n} = y_{D_n}) \\
&= \mathbb{E}(g(\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F(\mathbf{X}))|F))|Y_{D_n} = y_{D_n}) \\
&= \int_{\mathbb{R}^E} g(\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F(\mathbf{X}))|F = \varphi))\mathbb{P}_{F|Y_{D_n}=y_{D_n}}(d\varphi) \\
&= \int_{\mathbb{R}^E} g(\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F(\mathbf{X}))|F = \varphi))\mathbb{P}_{F^{D_n}}(d\varphi).
\end{aligned}$$

Since $\mathbf{X}$ and F are independent,
$$\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F(\mathbf{X}))|F = \varphi) = \mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(\varphi(\mathbf{X}))).$$

Hence,
$$\begin{aligned}
\mathbb{E}(g(\Pi_\rho^{D_n})) &= \int_{\mathbb{R}^E} g(\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(\varphi(\mathbf{X}))))\mathbb{P}_{F^{D_n}}(d\varphi) \\
&= \int_{\mathbb{R}^E} g(\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{X}))|F^{D_n} = \varphi))\mathbb{P}_{F^{D_n}}(d\varphi) \\
&= \mathbb{E}(g(\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{X}))|F^{D_n}))).
\end{aligned}$$

$\square$

The mean and the variance of $\Pi_\rho^{D_n}$ are, then, given by:

Proposition 2.1.

$$\mathbb{E}(\Pi_\rho^{D_n}) = \int_E \mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{x})))\mathbb{P}_\mathbf{X}(d\mathbf{x}) = \mathbb{E}\left(\Phi\left(\frac{\rho - m_{D_n}(\mathbf{X})}{\sqrt{K_{D_n}(\mathbf{X},\mathbf{X})}}\right)\right), \qquad (2.5)$$

where Φ is the cumulative distribution function of a centered reduced Gaussian random variable.

$$\mathbb{V}(\Pi_\rho^{D_n}) = \int_{E \times E} \mathbb{C}\text{ov}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{x})), \mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{x}'))\mathbb{P}_\mathbf{X} \times \mathbb{P}_\mathbf{X}(d\mathbf{x},d\mathbf{x}'). \qquad (2.6)$$

Proof

From Lemma 2.1, it comes

$$\begin{aligned}
\mathbb{E}(\Pi_\rho^{D_n}) &= \mathbb{E}(\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{X}))|F^{D_n})) \\
&= \mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{X}))) = \int_E \mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{x})))\mathbb{P}_\mathbf{X}(d\mathbf{x}).
\end{aligned}$$

Since $F^{D_n}(\mathbf{x})$ follows Gaussian distribution with mean $m_{D_n}(\mathbf{x})$ and variance $K_{D_n}(\mathbf{x},\mathbf{x})$,

$$\mathbb{E}(\Pi_\rho^{D_n}) = \mathbb{E}\left(\Phi\left(\frac{\rho - m_{D_n}(\mathbf{X})}{\sqrt{K_{D_n}(\mathbf{X},\mathbf{X})}}\right)\right).$$

Then, $\mathbb{E}((\Pi_\rho^{D_n})^2)$ is computed by using again Lemma 2.1 and the independence of $\mathbf{X}$ and F^{D_n}:

$$\begin{aligned}
\mathbb{E}((\Pi_\rho^{D_n})^2) &= \mathbb{E}\left(\left[\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{X}))|F^{D_n})\right]^2\right) \\
&= \int_{\mathbb{R}^E} \left[\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(\varphi(\mathbf{X})))\right]^2 \mathbb{P}_{F^{D_n}}(d\varphi) \\
&= \int_{\mathbb{R}^E} \int_E \mathbb{I}_{]-\infty,\rho[}(\varphi(\mathbf{x}))\mathbb{P}_\mathbf{X}(d\mathbf{x}) \int_E \mathbb{I}_{]-\infty,\rho[}(\varphi(\mathbf{x}))\mathbb{P}_\mathbf{X}(d\mathbf{x})\mathbb{P}_{F^{D_n}}(d\varphi) \\
&= \int_{E^2} \mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{x}))\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{x}')))\mathbb{P}_\mathbf{X} \times \mathbb{P}_\mathbf{X}(d\mathbf{x},d\mathbf{x}').
\end{aligned}$$

As, it also holds:

$$\mathbb{E}(\Pi_\rho^{D_n})^2 = \int_{E^2} \mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{x})))\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{x}')))\mathbb{P}_\mathbf{X} \times \mathbb{P}_\mathbf{X}(d\mathbf{x},d\mathbf{x}'),$$

$$\begin{aligned}
\mathbb{V}(\Pi_\rho^{D_n}) &= \mathbb{E}((\Pi_\rho^{D_n})^2) - \mathbb{E}(\Pi_\rho^{D_n})^2 \\
&= \int_{E^2} \mathbb{C}\text{ov}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{x})), \mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{x}'))\mathbb{P}_\mathbf{X} \times \mathbb{P}_\mathbf{X}(d\mathbf{x},d\mathbf{x}').
\end{aligned}$$

$\square$

A numerical Monte Carlo integration can be used to compute the posterior mean and variance since they do not need more calls to f. However, the computation time requested by a massive Monte Carlo integration, especially for $\mathbb{V}(\Pi_\rho^{D_n})$, can be very long as it is noticed in the examples.

The mean and the variance of $\Pi_\rho^{D_n}$ can be used to obtain credible bounds. As a consequence of Markov inequality, it holds, for any $\alpha \in [0, 1]$,

$$\mathbb{P}\left(\Pi_\rho^{D_n} \leq \frac{\mathbb{E}(\Pi_\rho^{D_n})}{\alpha}\right) \geq 1 - \alpha. \tag{2.7}$$

Likewise, Chebychev inequality gives, for any $\alpha \in [0, 1]$,

$$\mathbb{P}\left(\Pi_\rho^{D_n} \leq \mathbb{E}(\Pi_\rho^{D_n}) + \sqrt{\frac{\mathbb{V}(\Pi_\rho^{D_n})}{\alpha}}\right) \geq 1 - \alpha. \tag{2.8}$$

The quantiles of $\Pi_\rho^{D_n}$ are exactly the upper bounds that are sought. They can be reached through massive simulation of $\Pi_\rho^{D_n}$. For example, the following algorithm provides realizations of $\Pi_\rho^{D_n}$. It relies on a discretization of the Gaussian process to be simulated.

Algorithm 2.1.

1. **Simulate a realization of a Gaussian process:** *A realization of the vector of points $\tilde{y} = (y_{\tilde{\mathbf{x}}_i})_{1 \leq i \leq \tilde{n}}$ is drawned according to the distribution F^{D_n} of the Gaussian process.*

2. **Reconstruction of the realization:** *By a Kriging method, the points $\tilde{y} \cup y_{D_n}$ are interpolated. This interpolation is considered as a realization of F^{D_n} on E.*

3. **Numerical integration:** *The realization π_ρ corresponding to the realization of the Gaussian process is hence computed using a massive Monte Carlo integration with respect to the distribution of $\mathbf{X}$.*

Using a lot of iterations, it is possible to obtain an approximation of the cumulative distribution function of $\Pi_\rho^{D_n}$ which gives estimates of quantiles. Thus, a credible interval on Π_ρ is constructed. A constant $a \in [0, 1]$ is found such that:

$$\mathbb{P}(\Pi_\rho^{D_n} < a) \geq 1 - \alpha.$$

This approach can suffer of an error due to the spatial discretization needed at step 1 of the algorithm. In the next Section, an alternative approach based on importance sampling is proposed.

3 Importance sampling

As it was explained in Section 1, the major drawback of the crude Monte Carlo scheme is the high level of uncertainty when it is used for estimating the probability of a rare event.

Importance sampling is a way to tackle this problem. The basic idea is to change the distribution to make the target event more frequent. We aim at sampling according to the importance distribution:

$$\mathbb{P}_{\mathbf{Z}} : A \subset E \mapsto \mathbb{P}_{\mathbf{X}}(A|\hat{R}_\rho),$$

where $\hat{R}_\rho \subset E$ is to be designed close to $R_\rho = \{\mathbf{x} \in E : f(\mathbf{x}) < \rho\}$. Thanks to n calls to the metamodel, a set $\hat{R}_\rho$ can be chosen as follows:

$$\hat{R}_\rho = \hat{R}_{\rho,\kappa} = \left\{\mathbf{x} : m_{D_n}(\mathbf{x}) < \rho + \kappa\sqrt{K_{D_n}(\mathbf{x},\mathbf{x})}\right\}, \tag{3.1}$$

where κ is fixed such that "$\{\mathbf{x} : F(\mathbf{x}) < \rho\} \subset \hat{R}_{\rho,\kappa}$ with a good confidence level". In other words, if $\mathbf{x}$ is such that $f(\mathbf{x}) < \rho$, we want $\mathbf{x}$ to be in $\hat{R}_{\rho,\kappa}$. We recall that the posterior mean $m_{D_n}(\mathbf{x})$ is an approximation of $f(\mathbf{x})$ and $\kappa\sqrt{K_{D_n}(\mathbf{x},\mathbf{x})}$ has been added to take into account the uncertainty of the approximation.

A set of m points, $\mathbf{Z}_{1:m} = (\mathbf{Z}_1,\dots,\mathbf{Z}_m)$, is drawned to be an i.i.d. sample following the importance distribution. The corresponding importance sampling estimator of π_ρ is

$$\frac{\mathbb{P}_{\mathbf{X}}(\hat{R}_\rho)}{m}\Gamma(f,\mathbf{Z}_{1:m}) = \frac{\mathbb{P}_{\mathbf{X}}(\hat{R}_\rho)}{m}\sum_{k=1}^{m}\mathbb{I}_{]-\infty,\rho[}(f(\mathbf{Z}_k)). \tag{3.2}$$

The probability $\mathbb{P}_{\mathbf{X}}(\hat{R}_\rho)$ is computable by a Monte Carlo integration since it does not depend on f; yet, m more calls to f are necessary to compute $\mathbb{I}_{]-\infty,\rho[}(f(\mathbf{Z}_k))$. This estimator is only unbiased provided that $R_\rho \subset \hat{R}_\rho$. Nevertheless, it is an unbiased estimator of $\mathbb{E}_{\mathbf{X}}(\mathbb{I}_{]-\infty,\rho[}(f(\mathbf{X}))\mathbb{I}_{\hat{R}_\rho}(\mathbf{X}))$. Since $\Gamma(f,\mathbf{Z}_{1:m})$ follows a binomial distribution $\mathcal{B}\left(m, \frac{\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(f(\mathbf{X}))\mathbb{I}_{\hat{R}_\rho}(\mathbf{X}))}{\mathbb{P}_{\mathbf{X}}(\hat{R}_\rho)}\right)$, for any $\alpha \in]0;1[$, the following confidence upper bound holds:

$$\mathbb{P}\left(\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(f(\mathbf{X}))\mathbb{I}_{\hat{R}_\rho}(\mathbf{X})) \leq b(\Gamma(f,\mathbf{Z}_{1:m},\rho),m,\alpha)\mathbb{P}_{\mathbf{X}}(\hat{R}_\rho)\right) > 1-\alpha, \tag{3.3}$$

by using the bound (1.4). This is an upper bound on π_ρ only if the estimator (3.2) is unbiased i.e. only if $R_\rho \subset \hat{R}_\rho$. As it is noticed in the decomposition:

$$\pi_\rho = \mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(f(\mathbf{X}))) = \mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(f(\mathbf{X}))\mathbb{I}_{\hat{R}_\rho}(\mathbf{X})) + \mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(f(\mathbf{X}))(1 - \mathbb{I}_{\hat{R}_\rho}(\mathbf{X}))),$$

the second term on the right-hand side which is the opposite of the bias has to be controlled. That is why the random variable

$$\Pi_\rho^{D_n} = \mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{X}))|F^{D_n}),$$

whose a realisation is π_ρ, is considered.
Similarly to the previous decomposition, it holds

$$\Pi_\rho^{D_n} = \mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{X}))\mathbb{I}_{\hat{R}_\rho}(\mathbf{X})|F^{D_n}) + \mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{X}))(1 - \mathbb{I}_{\hat{R}_\rho}(\mathbf{X}))|F^{D_n}). \tag{3.4}$$

A bound on $\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{X}))\mathbb{I}_{\hat{R}_\rho}(\mathbf{X})|F^{D_n})$ comes from (3.3).

Proposition 3.1. *For $\alpha \in]0,1[$, it holds*

$$\mathbb{P}\Big(\Big(\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{X}))\mathbb{I}_{\hat{R}_\rho}(\mathbf{X})|F^{D_n}) \leq \mathbf{b}\ \mathbb{P}_{\mathbf{X}}(\hat{R}_\rho)\Big) \geq 1 - \alpha\,, \tag{3.5}$$

where $\mathbf{b}$ *stands for* $b(\Gamma(F^{D_n}, \mathbf{Z}_{1:m}, \rho), m, \alpha)$.

Proof

Let φ be any realisation of F^{D_n}.
As in (3.3), we have

$$\mathbb{P}\left(\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(\varphi(\mathbf{X}))\mathbb{I}_{\hat{R}_\rho}(\mathbf{X})) \leq b(\Gamma(\varphi, \mathbf{Z}_{1:m}, \rho), m, \alpha)\mathbb{P}_{\mathbf{X}}(\hat{R}_\rho)\right) \geq 1 - \alpha\,.$$

Thus, since this result holds for any realisation of F^{D_n},

$$\mathbb{P}\left(\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{X}))\mathbb{I}_{\hat{R}_\rho}(\mathbf{X})|F^{D_n}) \leq b(\Gamma(F^{D_n}, \mathbf{Z}_{1:m}, \rho), m, \alpha)\mathbb{P}_{\mathbf{X}}(\hat{R}_\rho)\right) \geq 1 - \alpha\,.$$

□

The next proposition states an upper bound for the second term in (3.4).

Proposition 3.2. *For $\beta \in]0,1[$, it holds*

$$\mathbb{P}\left(\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{X}))(1 - \mathbb{I}_{\hat{R}_\rho}(\mathbf{X}))|F^{D_n}) \leq \frac{\mathbf{c}}{\beta}\right) \geq 1 - \beta\,,$$

where $\mathbf{c} = \mathbb{E}\left(\Phi\left(\frac{\rho - m_{D_n}(\mathbf{X})}{\sqrt{K_{D_n}(\mathbf{X},\mathbf{X})}}\right)(1 - \mathbb{I}_{\hat{R}_\rho}(\mathbf{X}))\right)$.

Proof

The mean of $\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{X}))(1 - \mathbb{I}_{\hat{R}_\rho}(\mathbf{X}))|F^{D_n})$ can be computed in the same fashion than the mean of $\Pi_\rho^{D_n}$ in Proposition 2.1. It gives

$$\mathbb{E}\left(\mathbb{E}(\mathbb{I}_{]-\infty,\rho[}(F^{D_n}(\mathbf{X}))(1 - \mathbb{I}_{\hat{R}_\rho}(\mathbf{X}))|F^{D_n})\right) = \mathbb{E}\left(\Phi\left(\frac{\rho - m_{D_n}(\mathbf{X})}{\sqrt{K_{D_n}(\mathbf{X},\mathbf{X})}}\right)(1 - \mathbb{I}_{\hat{R}_\rho}(\mathbf{X}))\right)\,.$$

Then, Markov inequality is applied which completes the proof. □

Finally, by gathering the results of Proposition 3.1 and Proposition 3.2, a stochastic upper bound is found on $\Pi_\rho^{D_n}$.

Proposition 3.3. *For $\alpha, \beta \in]0,1[$ such that $\alpha + \beta < 1$, it holds*

$$\mathbb{P}\left(\Pi_\rho^{D_n} \leq \mathbf{b}\mathbb{P}_{\mathbf{X}}(\hat{R}_\rho) + \frac{\mathbf{c}}{\beta}\right) \geq 1 - (\alpha + \beta)\,, \tag{3.6}$$

where $\mathbf{b}$ *and* $\mathbf{c}$ *have been defined above.*

The proof is obvious.
If $\hat{R}_\rho$ is chosen as proposed in (3.1), the bound $\mathbf{c}$ is:

$$\mathbf{c} = c(\kappa) = \mathbb{E}\left(\Phi\left(\frac{\rho - m_{D_n}(\mathbf{X})}{\sqrt{K_{D_n}(\mathbf{X},\mathbf{X})}}\right)\mathbb{I}_{]-\infty,-\kappa[}\left(\frac{\rho - m_{D_n}(\mathbf{X})}{\sqrt{K_{D_n}(\mathbf{X},\mathbf{X})}}\right)\right)\,.$$

4 Numerical experiments

4.1 A toy example

The function $f : [-10, 10]^2 \to \mathbb{R}_+$ is assumed to describe a physical model:

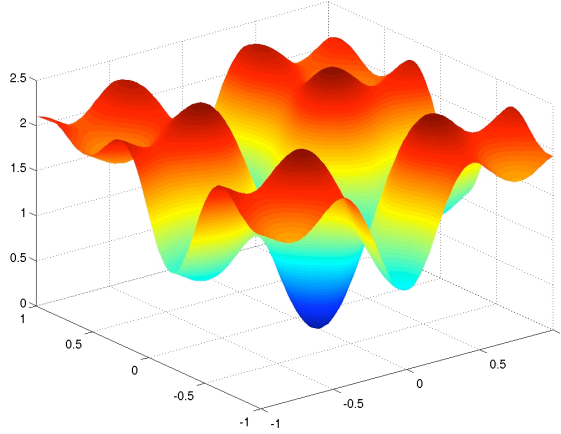$$f(x_1, x_2) = -\frac{\sin(x_1)}{x_1} - \frac{\sin(x_2 + 2)}{x_2 + 2} + 2 \,.$$

Figure 1: The function f

The input vector $\mathbf{X}$ is supposed to have a uniform distribution on $[-10, 10]^2$. The threshold is set to $\rho = 0.01$ which corresponds to the probability
$\mathbb{P}_{\mathbf{X}} \left(f(\mathbf{X}) < \rho \right) = 4.72 \cdot 10^{-4}$. This probability was computed thanks to a massive Monte Carlo integration. In the case where only $N = 100$ calls to f are available, the two strategies are tested. A maximin design with 100 points for the Bayesian strategy and one with 50 points for importance sampling strategy are computed thanks to a simulated annealing algorithm. Kriging metamodels are built with an intercept as the regression function and a Gaussian correlation function is chosen as the correlation function of the Gaussian process ζ i.e. $\forall \mathbf{x} \in E$, $h(\mathbf{x}) = 1$ and $\forall \mathbf{x}, \mathbf{x}' \in E$, $K(\mathbf{x}, \mathbf{x}) = \exp\left(-\theta \|\mathbf{x} - \mathbf{x}'\|^2\right)$ are set for the model given by equation (2.1). The Bayesian estimate of π_ρ is $4.63 \cdot 10^{-4}$. It was computed by a Monte Carlo integration on a 10^7-sample using the result of Proposition 2.1. Yet, we were not able to determine the posterior variance in a reasonable time. The importance sampling estimate of π_ρ, constructed on a 50-sample, is $6.13 \cdot 10^{-4}$. The probability $\mathbb{P}_{\mathbf{X}}(\hat{R}_{\rho,\kappa})$ (and also the bound on the bias, given in Proposition 3.2) was also computed by a Monte Carlo integration on a 10^7-sample and $\kappa = 3$ has been set.

Then, the stochastic bounds on $\Pi_\rho^{D_n}$ are focused on. A thousand iterations of Algorithm 2.1 where the points $\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_{\tilde{n}}$ have been chosen to be a grid of one hundred points in $[-1, 1]^2$

and the numerical integration at step 3 is performed with a 10^5-sample, are done. In order to prevent the covariance matrix of the posterior process to be ill-conditioned an eye matrix multiplied by a small coefficient (here 10^{-5}) is added. It is a regularization of the solution known as a nugget effect in the Kriging literature. The estimates of the posterior quantiles are $1.2 \cdot 10^{-3}$ at level 90% and $2.1 \cdot 10^{-3}$ at level 98%. The bounds found with importance sampling are $1.5 \cdot 10^{-3}$ at level 90% ($\alpha = \beta = 5\%$) and $2.1 \cdot 10^{-3}$ at level 98% ($\alpha = \beta = 1\%$). If a crude Monte Carlo scheme is used here with only $N = 100$ calls, the estimator is equal to 0 with probability greater than 0.95 and in this case, the upper confidence bounds are 0.023 and 0.038 respectively at levels 90% and 98%.

There are sources of variability on the estimators and the bounds due to the choice in the designs. Indeed, the designs are computed to be maximin by using a finite number of iterations of a simulated annealing algorithm. Moreover, there exist symmetries within the class of maximin designs. Concerning the importance sampling strategy, the sampling which gives $\mathbf{Z}_{1:m}$ induces variability.

In order to test the sensitivity of the estimators and the bounds to these sources of variability, each of the two strategies as described just above, are repeated one hundred times. Figure 2 displays a boxplot of one hundred estimates obtained with the Bayesian method on the left-hand side and a boxplot of one hundred estimates obtained with the importance sampling method on the right-hand side.
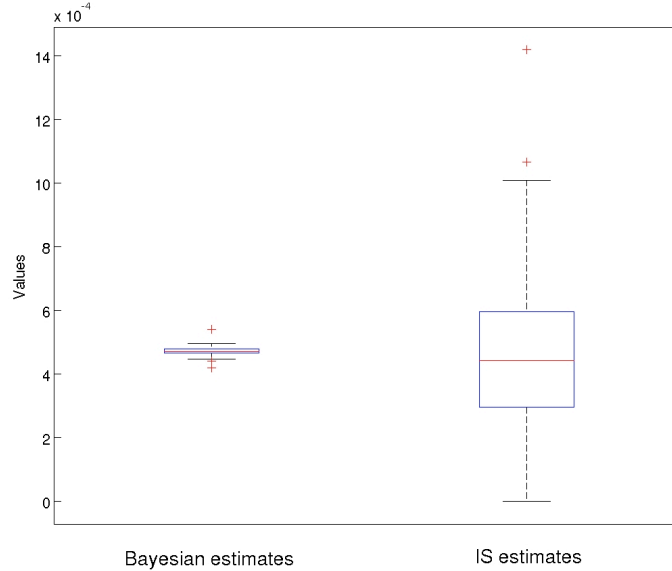

Figure 2: Estimates of π_ρ

Figure 3 displays the boxplots concerning bounds at level 90% and at level 98% given by the Bayesian method (left-hand side) and the importance sampling method (right-hand side). Table 1 summarizes the estimates and Table 2 summarizes the bounds.
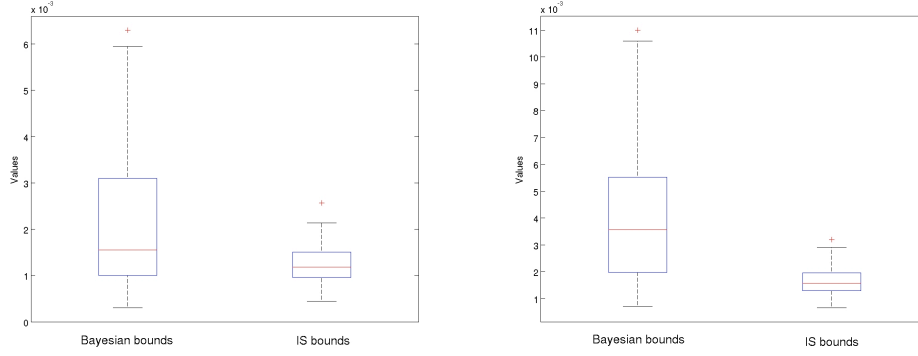
Figure 3: Bounds on π_ρ at level 90% (left) and at level 98% (right)

These results show that the Bayesian method is very reliable for estimating π_ρ while the importance sampling method provides the sharpest upper bounds. The Bayesian method suffers from the fact that the posterior quantiles are estimated thanks to an algorithm which relies on a discretization of the space and is burdensome which implies a limited number of possible iterations. The importance sampling methods which splits into two terms the probability to bound is much more efficient. As these methods depend on the Kriging model hypothesis (2.1), a leave-one-out cross validation as proposed by Jones et al. (1998) can be performed to check if this hypothesis is sensible. It consists of building n metamodels with posterior mean and variance denoted respectively by $m_{D_n^{-i}}$ and $\sigma^2_{D_n^{-i}}$, from designs

$$D_n^{-i} = \left\{ \mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n \right\},$$

where $i = 1, \ldots, n$.
Then, the values

$$\frac{|f(\mathbf{x}_i) - m_{D_n^{-i}}(\mathbf{x}_i)|}{\sigma^2_{D_n^{-i}}(\mathbf{x}_i)}, \tag{4.1}$$

are computed. If something like 99.7% of them lies in the interval $[-3, 3]$, the Kriging hypothesis is not rejected. In our toy example, all of the tests which were made give that all these values are in $[-2, 2]$.

	Bayesian estimates	IS estimates
Minimum	4.19	0
Maximum	5.40	14
Mean	4.72	4.72

Table 1: Estimates of π_ρ multiplied by 10^{-4}

	Bayesian bounds		IS bounds	
	90%	98%	90%	98%
Minimum	3	7	4.5	6.5
Maximum	63	110	26	32
Mean	20	39	12	16

Table 2: Bounds on π_ρ multiplied by 10^{-4}

4.2   A real case study: release enveloppe clearance

4.2.1   Context

When releasing an airborne load, a critical issue is the risk that its trajectory could collide the aircraft. The behaviour of such a load after release depends on many variables. Some are under control of the crew: mach, altitude, load factor etc. We call them *controlled variables* and note C their variation domain. The others are *uncontrolled variables*: let E be the set of their possible values. The *release enveloppe clearance* problem consists of exploring the set C to find a subset where the release is safe, whatever the uncontrolled variables are. To investigate this problem, we can use a simulator which computes the trajectory of the carriage when the values of all the variables are given. Moreover, for $\mathbf{x}_C \in C$ and $\mathbf{x} \in E$, besides the trajectory $\tau(\mathbf{x}_C, \mathbf{x})$, the program delivers a *dangerousness score* $f(\mathbf{x}_C, \mathbf{x})$ to be interpreted as an "algebraic distance": a negative value characterizes a collision trajectory. To assess the safety of release at a given point of C, we suppose that the values of the uncontrolled variables are realizations of a random variable $\mathbf{X} \in E$ that can be simulated. Therefore, for a given value $\mathbf{x}_C \in C$, and $\rho \geq 0$ the ρ-collision risk is the probability

$$\pi_\rho(\mathbf{x}_C) = \mathbb{P}(f(\mathbf{x}_C, \mathbf{X}) < \rho) \,.$$

We do not aim at estimating accurately this risk.
We would rather classify the points into three categories: according to the position of 0-risk $\pi_0(\mathbf{x}_C)$ with respect to the two markers 10^{-5} and 10^{-2}, $\mathbf{x}_C$ is said to be

1. **totally safe** if $\pi_0(\mathbf{x}_C) \leq 10^{-5}$,

2. **relatively safe** if $10^{-5} < \pi_0(\mathbf{x}_C) < 10^{-2}$,

3. **unsafe** if $\pi_0(\mathbf{x}_C) \geq 10^{-2}$.

In this example, there are 5 controlled and 26 uncontrolled variables, so that $C \subset \mathbb{R}^5, E \subset \mathbb{R}^{26}$. From budget point of view, experts consider that a set of about 400 representative points of C are enough to cover consistently the domain C. On the other hand, the computation of 800000 trajectories takes about 4 days which is considered reasonable. On the basis of these indications, the maximum amount of available calls to the simulator is $N = 2000$ per point.

4.2.2 Estimation strategy

Our estimation strategy which applies iteratively to each point of the set of representative points, has two steps each of which uses half of the calls budget: $m = n = \frac{N}{2} = 1000$. Let $\mathbf{x}_C \in C$ be the current point of interest that we suppose fixed. For any $\mathbf{x} \in E$, $f(\mathbf{x}) = f(\mathbf{x}_C, \mathbf{x})$ is set, recovering the notation introduced in the first part of the paper.

1. At the first stage, a Gaussian process is built as explained in (2), on the basis of evaluations $f(\mathbf{x}_1), \cdots, f(\mathbf{x}_n) \in \mathbb{R}^n$ of f on $D_n = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$. From Proposition 2.1, we know that π_ρ is a realization of the random variable $\Pi_\rho^{D_n}$ whose mean

$$\mathbb{E}(\Pi_\rho^{D_n}) = \mathbb{E}\left(\Phi\left(\frac{\rho - m_{D_n}(\mathbf{X})}{\sqrt{K_{D_n}(\mathbf{X}, \mathbf{X})}}\right)\right),$$

   can be computed accurately.
   As stated by (2.7), applying Markov inequality gives, for any $\alpha \in]0; 1[$,

$$\mathbb{P}\left(\Pi_\rho^{D_n} \leq \frac{\mathbb{E}(\Pi_\rho^{D_n})}{\alpha}\right) \geq 1 - \alpha.$$

   According to the value of $\mathbb{E}(\Pi_\rho^{D_n})$, we, then, take the following decisions:

   - if $\mathbb{E}(\Pi_\rho^{D_n}) \leq \frac{1}{2}10^{-10}$ which leads by (2.7) to $\mathbb{P}\left(\Pi_\rho^{D_n} \leq \frac{10^{-5}}{2}\right) \geq 1 - \frac{10^{-5}}{2}$, we qualify the current point $\mathbf{x}_C \in C$ as totally safe,
   - if $\mathbb{E}(\Pi_\rho^{D_n}) \geq 10^{-2}$, we conservatively classify $\mathbf{x}_C$ as unsafe,
   - if $\frac{1}{2}10^{-10} < \mathbb{E}(\Pi_\rho^{D_n}) < 10^{-2}$ we use a second stage procedure to refine the risk assessment.

2. A million-sample $\mathbf{x}_1, \cdots, \mathbf{x}_M$ of $\mathbf{X}$ is drawn from which we tune κ in such a way that $m = 1000$ of these million elements of E are in $\hat{R}_{\rho,\kappa}$. The resulting points $\mathbf{z}_1, \cdots, \mathbf{z}_m$ are a m-sample $\mathbf{z}_{1:m}$ of realizations of the random variable $\mathbf{Z}$ which follows the importance distribution,

$$\mathbb{P}_{\mathbf{Z}} : A \mapsto \mathbb{P}_{\mathbf{X}}(A | \hat{R}_{\rho,\kappa}).$$

   By using m calls to the simulator, $\Gamma(f, \mathbf{z}_{1:m}, \rho)$ is computed. Drawn from Proposition 3.3 with setting $\alpha = \beta$, we obtain the bound

$$b(\Gamma(f, \mathbf{z}_{1:m}, \rho), m, \alpha)\mathbb{P}_{\mathbf{X}}(\hat{R}_{\rho,\kappa}) + \frac{c(\kappa)}{\alpha},$$

   which is a decreasing function of α.
   Let define $\alpha_0 = \min\{\alpha : b(\Gamma(f, \mathbf{z}_{1:m}, \rho), m, \alpha)\mathbb{P}_{\mathbf{X}}(\hat{R}_{\rho,\kappa}) + \frac{c(\kappa)}{\alpha} \leq 2\alpha\}$. For such an α_0, Proposition 3.3 states:

$$\mathbb{P}\left(\Pi_\rho^{D_n} \leq b(\Gamma(F^{D_n}, \mathbf{Z}_{1:m}, \rho), m, \alpha_0)\mathbb{P}_{\mathbf{X}}(\hat{R}_\rho) + \frac{c(\kappa)}{\alpha_0}\right) \geq 1 - 2\alpha_0,$$

   which provides $2\alpha_0$ as a $1 - 2\alpha_0$ confidence upper bound on π_ρ.

4.2.3 Experiments

Three points of C have been experienced. Of these cases the first one is known to be a null 0-risk point, while the third one is very unsafe and the second one is in-between.

For benchmarking purpose, besides the simulator calls budget required for the estimation process described in 4.2.2, a 10000-samples of $f(\mathbf{x}_E, \mathbf{X})$ has been computed for each of the three examples. For each case, we began by estimating a Gaussian process on the basis of f-values computed on the points of a 1000 points maximin latin hypercube design $D_n = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$. Figures 4, 5 and 6 show the predictive performance of the processes when applied to the benchmark points. These points, which appear in red, are sorted according to their process mean values while the blue curves mark the predicted 3 standard deviation positions around the means. As it appears rather clearly, the dispersion of ther eal values is underestimated by the model: they overflow the blue zone with a frequency ($\sim 5\%$) higher than expected (0.27%). The worse case is the first one, for which large deviations appear for benchmark points with low values of f. In order to obtain bounds from (2.7), we then
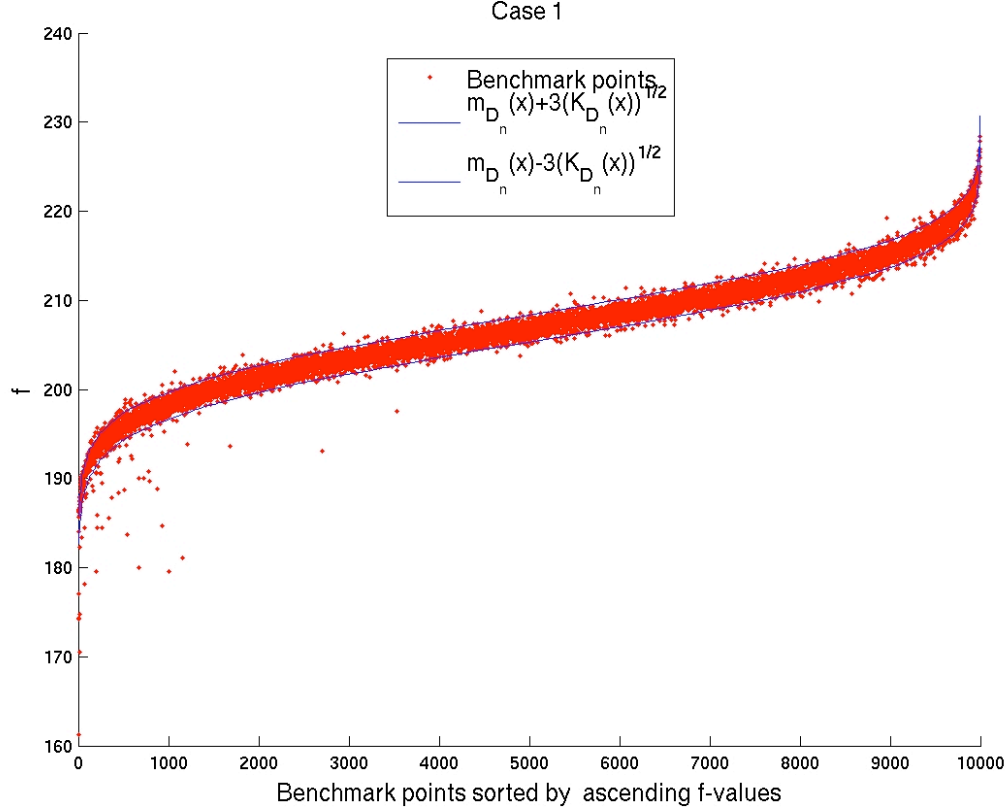

Figure 4: Prediction performance case 1
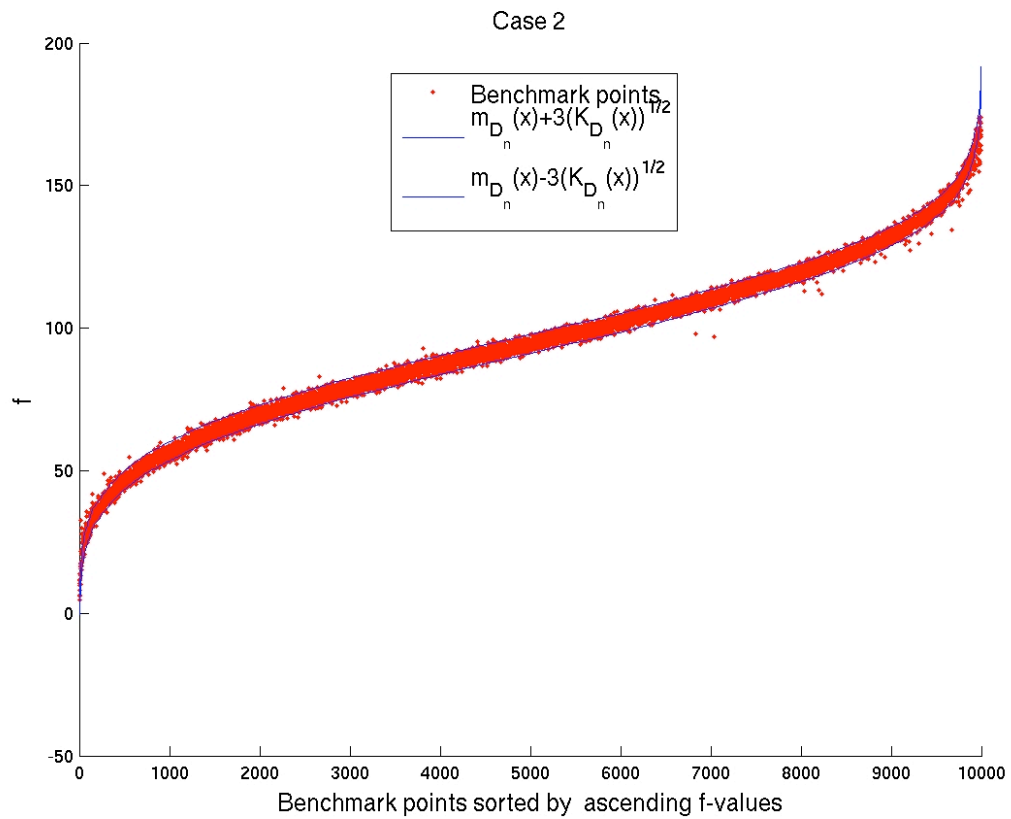
computed $\mathbb{E}(\Pi_0^{D_n})$ using (2.5):

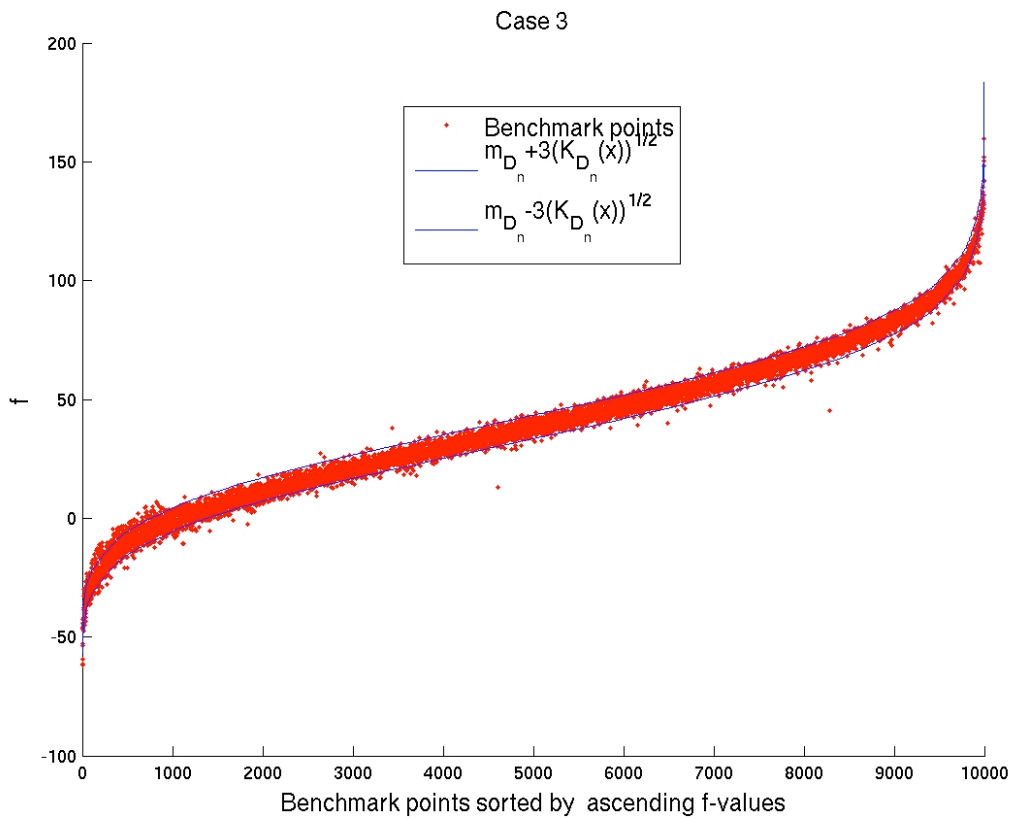Figure 5: Prediction performance case 2

Figure 6: Prediction performance case 3

- In the first case, the massive Monte Carlo procedure leads to a numerically null evaluation of $\mathbb{E}(\Pi_0^{D_n})$ and, as a consequence, to the classification of the related C point as totally safe.

- In the second example, $\mathbb{E}(\Pi_0^{D_n})$ beeing evaluated at $1.68 \ 10^{-4}$, we need to proceed the second step.

- $\mathbb{E}(\Pi_0^{D_n}) = 0.103$ in case 3 which is consistent with the 90% confidence interval $[0.0999; 0.1101]$, obtained on benchmark data.

We now applied the procedure second stage to refine collision probability estimation: the obtained confidence upper bound is $1.2 \ 10^{-5}$ at confidence level $1 - 1.2 \ 10^{-5}$. The benchmark data do not show collision case: a 90% confidence upper bound is $2.3 \ 10^{-4}$.

5 Dicussion

In this paper, two methods were proposed to estimate and to bound the probability of a rare event which depends on an expensive black-box function. They are both based on a Kriging hypothesis which induces an random interpretation of the probability to estimate. That is why the Bayesian context is natural in this problem and leads to a very accurate estimator. As it is hard to reach the posterior quantiles, it does not achieve as tight upper bounds as the importance sampling method does. The importance sampling method relies on a split in the possible calls to f. We have proposed to use half of the calls to compute a metamodel and half of the calls to draw a sample according to the importance distribution; yet, other way of splitting can be investigated.

As it was noticed on the toy example, there is a variability due to the choice in the design. To reduce it, some points can be added where uncertainties on the prediction of the metamodel are high ($K_{D_n}(\mathbf{x}, \mathbf{x})$ is large) and the probability that f is smaller than ρ is high. It can consist of adding sequentially points of $\hat{R}_{\rho,\kappa}$ where the variance of prediction is the largest, as in Vazquez and Bect (2009); Bect et al. (2011) and in Ranjan et al. (2008); Picheny et al. (2010) for contour estimation.

We have dealt with a cross validation method to assess the Kriging hypothesis. However, in the case where the cross validation leads to reconsider this hypothesis, a solution is to extend the confidence interval on the prediction by tuning at hand the parameter σ^2 in equation (2.1). In Bayesian words, it can be called using a less informative prior distribution on f.

We have not manage to compute the posterior variance (given by Proposition 2.1) by using a massive Monte Carlo integration in our examples since it is very small. However, other rare events methods can be investigated since the variance does not depend anymore on f.

References

Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E. (2011). Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*. to appear.

Bingham, D., Hengartner, N., Higdon, D., and Kenny, Q. Y. (2006). Variable Selection for Gaussian Process Models in Computer Experiments. *Technometrics*, 48(4):478–490.

Cannamela, C., Garnier, J., and Iooss, B. (2008). Controlled stratification for quantile estimation. *The Annals of Applied Statistics*, 2(4):1554–1580.

Cérou, F. and Guyader, A. (2007a). Adaptive multilevel splitting for rare event analysis. *Stochastic Analysis and Applications*, 25(2):417–443.

Cérou, F. and Guyader, A. (2007b). Adaptive particle techniques and rare event estimation. In *Conference Oxford sur les méthodes de Monte Carlo séquentielles*, volume 19 of *ESAIM Proc.*, pages 65–72. EDP Sci., Les Ulis.

Del Moral, P. and Garnier, J. (2005). Genealogical particle analysis of rare events. *The Annals of Applied Probability*, 15(4):2496–2534.

Dette, H. and Pepelyshev, A. (2010). Generalized Latin Hypercube Design for Computer Experiments. *Technometrics*, 52(4):421–429.

Fang, K.-T., Li, R., and Sudjianto, A. (2006). *Design and Modeling for Computer Experiments*. Computer Science and Data Analysis. Chapman & Hall/CRC.

Heidelberg, P. (1995). Fast simulation of rare events in queuing and reliability models. *ACM Transactions on Modeling and Computer Simulation*, 5:43–85.

Jones, D. R., Schonlau, M., and Welch, W. J. (1998). Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13(4):455–492.

Joseph, V. R. (2006). Limit Kriging. *Technometrics*, 48(4):458–466.

Koehler, J. and Owen, A. (1996). Computer experiments. In *Design and analysis of experiments*, volume 13 of *Handbook of Statistics*, pages 261–308. North Holland, Amsterdam.

L'Ecuyer, P., Demers, V., and Tuffin, B. (2007). Rare events, splitting, and quasi-monte carlo. *ACM Trans. Model. Comput. Simul.*, 17(2).

Li, R. and Sudjianto, A. (2005). Analysis of Computer Experiments using Penalized Likelihood in Gaussian Kriging Models. *Technometrics*, 47(2):111–120.

Mease, D. and Bingham, D. (2006). Latin Hyperrectangle Sampling for Computer Experiments. *Technometrics*, 48(4):467–477.

Morris, M. and Mitchell, T. (1995). Exploratory designs for computer experiments. *Journal of Statistical Planning and Inference*, 43:381–402.

Munoz Zuniga, M., Garnier, J., Remy, E., and de Rocquigny, E. (2010). Adaptative directional stratification for controlled estimation of the probability of a rare event. Technical report.

Oakley, J. (2004). Estimating percentiles of uncertain computer codes outputs. *Applied Statistics*, 53:83–93.

Picheny, V., Ginsbourger, D., Roustant, O., and Haftka, R. (2010). Adaptive designs of experiments for accurate approximation of a target region. *Journal of Mechanical Design*, 132(7).

Ranjan, P., Bingham, D., and Michailidis, G. (2008). Sequential Experiment Design for Contour Estimation From Complex Computer Codes. *Technometrics*, 50(4):527–541.

Santner, T., Williams, B., and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. Springer-Verlag.

Shahabuddin, P. (1995). Rare event simulation in stochastic models. In *WSC'95: Proceedings of the 27th conference on Winter simulation*.

Vazquez, E. and Bect, J. (2009). A sequential Bayesian algorithm to estimate a probability of failure. In *15th IFAC SYmposium on System IDentification (SYSID 2009)*.

Welch, W. J., Buck, R. J., Sack, J., Wynn, H. P., Mitchell, T. J., and Morris, M. D. (1992). Screening, Predicting, and Computer Experiments. *Technometrics*, 34(1):15–25.