

A sequential Bayesian algorithm to estimate a probability of failure

Emmanuel Vazquez, Julien Bect

► **To cite this version:**

Emmanuel Vazquez, Julien Bect. A sequential Bayesian algorithm to estimate a probability of failure. 15th IFAC Symposium on System Identification, SYSID 2009, Jul 2009, Saint-Malo, France. 5 p. hal-00368158v2

HAL Id: hal-00368158

<https://hal-supelec.archives-ouvertes.fr/hal-00368158v2>

Submitted on 26 Mar 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A sequential Bayesian algorithm to estimate a probability of failure

Emmanuel Vazquez and Julien Bect

SUPELEC, Gif-sur-Yvette, France
(e-mail: firstname.lastname@supelec.fr)

Abstract: This paper deals with the problem of estimating the probability of failure of a system, in the challenging case where only an expensive-to-simulate model is available. In this context, the budget for simulations is usually severely limited and therefore classical Monte Carlo methods ought to be avoided. We present a new strategy to address this problem, in the framework of sequential Bayesian planning. The method uses kriging to compute an approximation of the probability of failure, and selects the next simulation to be conducted so as to reduce the mean square error of estimation. By way of illustration, we estimate the probability of failure of a control strategy in the presence of uncertainty about the parameters of the plant.

Keywords: Sequential Bayesian planning; Design of experiments; Kriging; Probability of failure

1. INTRODUCTION

Uncertainty on the parameters of a system have to be taken into account during its design to insure a desired level of reliability. In particular, it is important to estimate the probability of the system to work under abnormal or dangerous operating conditions due to unknown random dispersions of its parameters. The *probability of failure* of a system is usually expressed as the probability of the excursion set of a function of performance above a fixed threshold. More precisely, let f be a measurable real function defined over a probability space $(\mathbb{X}, \mathcal{X}, \mathbb{P}_{\mathbb{X}})$ and let $u \in \mathbb{R}$ be a threshold. The problem to be considered in this paper is the estimation of the volume under $\mathbb{P}_{\mathbb{X}}$ of the excursion set

$$A(f) := \{x \in \mathbb{X} : f(x) \geq u\} \quad (1)$$

of the function f above the level u . In the context of robust design, the volume $\alpha(f) := \mathbb{P}_{\mathbb{X}}(A(f))$ can be viewed as the probability of failure of a system. Then, \mathbb{X} plays the role of an input or factor space, the probability $\mathbb{P}_{\mathbb{X}}$ models the uncertainty on the inputs or the factors, and f is some function of performance of the outputs of the system, whose evaluation may involve complex and time-consuming simulations or fabrications of prototypes. When f is expensive to evaluate, the estimation of $\alpha(f)$ must be carried out with a *restricted number of evaluations* of f , generally excluding the estimation of the probability of excursion by a Monte Carlo approach. Indeed, consider the empirical estimator

$$\alpha_m(f) := \frac{1}{m} \sum_{i=1}^m \mathbf{1}_{\{f(X_i) \geq u\}}, \quad (2)$$

where the X_i s are independent random variables with distribution $\mathbb{P}_{\mathbb{X}}$. According to the strong law of large

numbers, the estimator $\alpha_m(f)$ converges to $\alpha(f)$ almost surely when m increases. Moreover, it is an unbiased estimator of $\alpha(f)$, i.e. $\mathbb{E}(\alpha_m(f)) = \alpha(f)$. Its mean square error is

$$\mathbb{E}((\alpha_m(f) - \alpha(f))^2) = \frac{1}{m} \alpha(f)(1 - \alpha(f)).$$

If the probability of failure $\alpha(f)$ is small, then the standard deviation of $\alpha_m(f)$ is approximately $(\alpha(f)/m)^{1/2}$. To achieve a given standard deviation $\delta\alpha(f)$ thus requires approximately $1/(\delta^2\alpha(f))$ evaluations, which can be prohibitively high if δ is small. Of course, many importance sampling methods have been proposed to improve over the basic Monte Carlo convergence rate — for instance, methods based on cross-entropy (Rubinstein, 1999), or subset sampling (Au and Beck, 2001). They will not be considered here for the sake of brevity and because the required number of function evaluations is still very high.

Methods to estimate a probability of failure that do not require a large number of evaluations of f are usually based on an approximation of the frontier $\partial A(f)$ of $A(f)$ by a simple geometrical shape. Compared to Monte Carlo methods, this class of methods does achieve a large reduction in the required number of system evaluations, but the accuracy of the estimator depends on the actual shape of $\partial A(f)$ and its resemblance to the approximating shape. The popular first-order reliability method (FORM) uses a hyperplane as an approximation of $\partial A(f)$ and the second-order reliability method (SORM) uses a paraboloid (see, e.g., Bjerager, 1990). Nevertheless, $\partial A(f)$ may depart significantly from such elementary shapes. In fact, this class of methods does not provide *statistically consistent* estimators of the probability of failure.

This paper presents a new approach. The method proposed, which is outlined in Section 2, is in essence a Bayesian sequential search algorithm similar in spirit to Bayesian algorithms for global optimization (Zilinskas, 1992; Mockus, 1994) such as the Expected Improvement algorithm (Schonlau and Welch, 1996) and the IAGO algorithm Villemonteix et al. (to appear.). Here, we transpose the main ideas and propose an algorithm to choose *sequentially* the evaluations of f to produce a consistent and rapidly converging estimation of $\alpha(f)$. This algorithm is detailed in Section 3. Section 4 will illustrate the potential evaluation savings of the methodology on a simple but archetypal example, in which the probability of failure of a control strategy of an uncertain plant is estimated.

2. SEQUENTIAL BAYESIAN SEARCH ALGORITHM

2.1 Sequential Bayesian optimal design

Let (G_n) be any given sequence of functions, $G_n : (\mathbb{X} \times \mathbb{R})^n \rightarrow \mathbb{R}_+$, such that $G_n((x_1, f(x_1)), \dots, (x_n, f(x_n)))$ yields an approximation of $\alpha(f)$. Our goal is to choose locations $X_n = X_n(f)$ sequentially in such a way that the approximations

$$\hat{\alpha}_n := G_n((X_1, f(X_1)), \dots, (X_n, f(X_n)))$$

will converge as fast as possible to the true value $\alpha(f)$. By doing so, it is hoped that a reasonably good approximation of $\alpha(f)$ will be obtained using a limited number of evaluations. Obviously, in such a procedure, the n^{th} location $X_n = X_n(f)$ is only allowed to depend on the unknown function f through the past evaluations $f(X_i)$, $1 \leq i < n$. For the sake of clarity, the explicit dependence on f will be dropped in the rest of the paper.

We formulate the problem in the framework of sequential Bayesian optimal design, where the information from the past evaluations can be combined with prior information on the system in order to take the “best” sequence of decisions (in a sense to be defined later). Let $F = \mathbb{R}^{\mathbb{X}}$ be the space of all real functions $\mathbb{X} \rightarrow \mathbb{R}$ and denote by \mathcal{U} the cylinder σ -algebra on F . Prior information on the unknown function f is taken into account in the Bayesian approach under the form of a probability measure \mathbb{P} on (F, \mathcal{U}) . This amounts to considering f as a random function, the distribution of which is given by the *prior probability* \mathbb{P} before evaluations are made, and by the *posterior probability* $\mathbb{P}\{\cdot | \mathcal{F}_n\}$ after n evaluations, where \mathcal{F}_n is the σ -algebra generated by $f(X_1), \dots, f(X_n)$. For each $n \geq 0$, the location X_{n+1} is assumed to be chosen measurably from the past evaluations (in other words, X_{n+1} is an \mathcal{F}_n -measurable random variable). It must be pointed out that despite the semantic shift we are still dealing with deterministic algorithms: randomness only comes from the fact that we are now considering f as a random function.

Given a cost function $C : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}$ and a budget of N evaluations of f , a strategy $X = (X_1, \dots, X_N)$ is said to be optimal if it minimizes the Bayes risk

$\mathbb{E}(C(\alpha(f), \hat{\alpha}_N))$ associated to the prior \mathbb{P} . It is well known (Berger, 1985) that, at least in principle, the optimal strategy can be found using a form of dynamic programming. In practice, finding the optimal strategy is usually not feasible, especially in problems like the one at hand where there is a continuum of possible states and decisions at each stage n . In this paper we handle the difficulty with the following suboptimal greedy strategy (sometimes called a *one-step look-ahead procedure*) :

$$X_n = \underset{x_n \in \mathbb{X}}{\operatorname{argmin}} \mathbb{E}\{C(\alpha(f), \hat{\alpha}_n) | \mathcal{F}_{n-1}\}, \quad (3)$$

for all $n \in \{1, \dots, N\}$, where the predictor $\hat{\alpha}_n$ is computed using the past evaluations $f(X_i)$, $1 \leq i < n$ and the uncertain outcome $f(x_n)$ of an evaluation at x_n .

The reader is referred to Chaloner and Verdinelli (1995) for a survey on both sequential and non-sequential Bayesian experimental design (from a statistical point of view) and to Pronzato (2008) for a survey on the connections between the fields of design of experiments and automatic control.

2.2 Gaussian process priors and kriging

The proposed method will now be specified following the Bayesian approach explained in Section 2.1, using a *Gaussian process prior* as the probability \mathbb{P} . The idea of modelling an unknown deterministic function f by a stochastic process has originally been introduced in the field of geostatistics (see Chilès and Delfiner, 1999, and the references therein), and also plays a central role in the design and analysis of “computer experiments” (see, e.g., Sacks et al., 1989; Currin et al., 1991; Welch et al., 1992).

The distribution \mathbb{P} of a Gaussian process f is uniquely determined by its mean $m(x) := \mathbb{E}(f(x))$ and its covariance function $k(x, y) := \mathbb{E}((f(x) - m(x))(f(y) - m(y)))$. Restricting our framework to Gaussian processes makes it possible to deal with the conditional distributions $\mathbb{P}\{\cdot | \mathcal{F}_n\}$ that appear in equation (3). Another benefit of using such a prior is that the posterior mean

$$f_n(x) := \mathbb{E}(f(x) | \mathcal{F}_n), \quad (4)$$

also called the *kriging predictor* of $f(x)$, can easily be computed using linear algebra (see, e.g., Chilès and Delfiner, 1999) and provides a cheap *surrogate model* for the true function f . Computing the probability of failure $\alpha(f_n)$ of the surrogate model is a natural choice for the approximations G_n of subsection 2.1, which leads us to set $\hat{\alpha}_n = \alpha(f_n)$. Another important benefit of using Gaussian process priors is that the posterior variance $\sigma_n^2(x)$ is also easily computed, which allows to provide prediction intervals.

Practically, the probability $\alpha(f_n)$ can be approximated by a Monte Carlo estimator such as $\alpha_m(f_n)$, which requires m evaluations of f_n . Since the evaluation of the surrogate model is much faster than the evaluation of the true model, m can be chosen large enough to make $\alpha_m(f_n) - \alpha(f_n)$ negligible with respect to $\alpha(f_n) - \alpha(f)$. Thus, we shall only focus on the convergence of $\alpha(f_n)$ to $\alpha(f)$ and assume

that $\alpha(f_n)$ can be computed with the required accuracy in a reasonable time.

Finally, choosing the quadratic cost function $C(\alpha, \alpha') = (\alpha - \alpha')^2$, the proposed strategy is to sequentially minimize the mean square error of approximation:

Program 1

$$X_n = \operatorname{argmin}_{x_n \in \mathbb{X}} \Upsilon_n(x_n) := \mathbb{E}\{(\alpha(f) - \alpha(\tilde{f}_n))^2 \mid \mathcal{F}_{n-1}\}.$$

Here and subsequently, we denote by $\tilde{f}_n(x)$ the kriging predictor of $f(x)$ based on $f(X_1), \dots, f(X_{n-1})$ and $f(x_n)$. This strategy can be seen as an instance of the so-called *stepwise uncertainty reduction* principle (Geman and Jedyak, 1996; Villemonteix et al., to appear.). The criterion $\Upsilon_n(x_n)$ provides a measure of the expected approximation error if a new evaluation is performed at x_n . It depends of course on the past evaluations $f(X_1), \dots, f(X_{n-1})$, since it is defined as a conditional expectation with respect to \mathcal{F}_n , but this dependency is kept implicit for simplicity of notation. It must be noted that the actual result $f(x_n)$ of the additional evaluation at x_n is integrated out by the conditional expectation with respect to \mathcal{F}_{n-1} .

3. TOWARD A NUMERICALLY TRACTABLE SEQUENTIAL ALGORITHM

This section presents a numerically tractable approximation of Program 1, proceeding in two steps. In the first step an upper bound Υ'_n on the criterion Υ_n is derived. This upper bound is simpler than Υ_n but still not in closed form. In the second step we provide a computationally efficient procedure to approximate Υ'_n , the result of which is our final criterion Υ''_n .

3.1 Upper bound on the criterion Υ_n

Using the chain rule for conditional expectations, observe that $\Upsilon_n(x_n)$ can also be written as

$$\Upsilon_n(x_n) = \mathbb{E}\left\{ \Gamma_n(x_n, f(x_n))^2 \mid \mathcal{F}_{n-1} \right\}, \quad (5)$$

where the function Γ_n is defined by

$$\Gamma_n(x_n, z_n) := \tilde{\mathbb{E}}_n \left\{ \left(\alpha(f) - \alpha(\tilde{f}_n) \right)^2 \right\}^{1/2}.$$

For abbreviation, the notation $\tilde{\mathbb{E}}_n$ has been introduced to denote the conditional expectation operator

$$\tilde{\mathbb{E}}_n := \mathbb{E}\{ \cdot \mid \mathcal{F}_{n-1}, f(x_n) = z_n \}.$$

Note that Equation (5) is simply a one-dimensional integral with respect to the value of $f(x_n)$. Unfortunately, $\Gamma_n(x_n, z_n)$ has no analytical expression. A numerical approximation could in principle be obtained using a Monte Carlo approach, but would require a very large number of conditional simulations (see Chilès and Delfiner, 1999, Chap. 7) of the random function f .

Instead, we choose to replace $\Gamma_n(x_n, z_n)$ by an upper-bound which will be easier to compute and minimize. Remark that

$$\alpha(f) - \alpha(\tilde{f}_n) = \int_{\mathbb{X}} \left(\mathbf{1}_{f(x) > u} - \mathbf{1}_{\tilde{f}_n(x) > u} \right) \mathbb{P}_{\mathbb{X}}(dx).$$

An upper bound is then readily provided by the generalized Minkowski inequality (see, e.g., Vestrup, 2003, Section 10.7):

$$\Gamma_n(x_n, z_n) \leq \Gamma'_n(x_n, z_n) := \int_{\mathbb{X}} v_n(x; x_n, z_n)^{\frac{1}{2}} \mathbb{P}_{\mathbb{X}}(dx),$$

with

$$v_n(x; x_n, z_n) := \tilde{\mathbb{E}}_n \left\{ \left(\mathbf{1}_{f(x) > u} - \mathbf{1}_{\tilde{f}_n(x) > u} \right)^2 \right\}.$$

Replacing Γ_n by Γ'_n in Program 1 yields the following modified strategy:

Program 2

$$X_n = \operatorname{argmin}_{x_n \in \mathbb{X}} \Upsilon'_n(x_n) := \mathbb{E}\left\{ \Gamma'_n(x_n, f(x_n))^2 \mid \mathcal{F}_{n-1} \right\}.$$

This strategy is indeed simpler than the previous one, since $v_n(x; x_n, f(x_n))$ is given by

$$v_n(x; x_n, f(x_n)) = \Psi \left(\left| \frac{u - \tilde{f}_n(x)}{\tilde{\sigma}_n(x)} \right| \right),$$

where Ψ is the complementary cumulative distribution function of the normal distribution and $\tilde{\sigma}_n(x)$ the standard deviation of $\tilde{f}_n(x)$ conditionally to \mathcal{F}_{n-1} and $f(x_n)$.

An informal interpretation of Program 2 is that X_n minimizes, conditionally to the past evaluations, the expected posterior error of prediction of $\mathbf{1}_{f(x) > u}$ by $\mathbf{1}_{\tilde{f}_n(x) > u}$, averaged over \mathbb{X} (using the distribution $\mathbb{P}_{\mathbb{X}}$).

3.2 Approximation of the upper bound

In the following paragraphs, we show that an approximate version of Program 2 can be solved with an acceptable computational complexity.

A first approximation is required for the computation of Γ'_n . Assuming that the distribution $\mathbb{P}_{\mathbb{X}}$ is easy to sample from, we propose to use a simple Monte Carlo approximation for the integral with respect to $\mathbb{P}_{\mathbb{X}}$:

$$\Gamma'_n(x_n, z_n) \approx \Gamma''_n(x_n, z_n) := \frac{1}{m} \sum_{i=1}^m v_n(Y_i; x_n, z_n)^{1/2},$$

where $(Y_i)_{i \geq 1}$ is a sequence of independent and identically distributed variables with common distribution $\mathbb{P}_{\mathbb{X}}$.

Another approximation is required for the computation of

$$\Upsilon'_n(x_n) \approx \mathbb{E}\left\{ \Gamma''_n(x_n, f(x_n))^2 \mid \mathcal{F}_{n-1} \right\}. \quad (6)$$

Using a tagged partition $\{(\Delta_n^j, z_n^j), 1 \leq j \leq Q\}$ of \mathbb{R} , we replace the random variable $f(x_n)$ in the right-hand side of (6) by the quantized variable $\xi_n := \sum_j z_n^j \mathbf{1}_{f(x_n) \in \Delta_n^j}$. The proposed algorithm is thus

Program 3

$$X_n = \operatorname{argmin}_{x_n \in \mathbb{X}} \Upsilon''_n(x_n) := \mathbb{E}\left\{ \Gamma''_n(x_n, \xi_n)^2 \mid \mathcal{F}_{n-1} \right\}.$$

Practically, the design criterion is computed as

$$\Upsilon''_n(x_n) = \sum_{j=1}^Q \Gamma''_n(x_n, z_n^j) \mathbb{P}\{f(x_n) \in \Delta_n^j \mid \mathcal{F}_{n-1}\},$$

using the fact that $f(x_n)$ is normally distributed conditionally to \mathcal{F}_{n-1} , with mean $f_{n-1}(x_n)$ and variance $\sigma_{n-1}^2(x_n)$.

4. NUMERICAL EXAMPLES

4.1 Illustration

This section provides a compendious one-dimensional illustration of the algorithm. We wish to estimate $\alpha(f)$, where $f(x)$ is a given function defined over \mathbb{R} , which is endowed with the probability distribution $\mathbb{P}_{\mathbb{X}} = \mathcal{N}(0, \sigma^2)$, $\sigma = 0.4$ (see Figure 1). After a few iterations, the unknown function f has been sampled so that the probability of excursion $\mathbb{P}(\{f : f(x) > u\} | \mathcal{F}_n)$ is either 0 or 1 in the region where the density of $\mathbb{P}_{\mathbb{X}}$ is high (as shown on Figure 1). This example illustrates the effectiveness of the proposed algorithm. Note that in practice, a parametrized covariance has to be chosen for specifying \mathbb{P} and its parameters should be estimated from the data, using, for instance, a maximum likelihood approach (e.g. Stein, 1999).

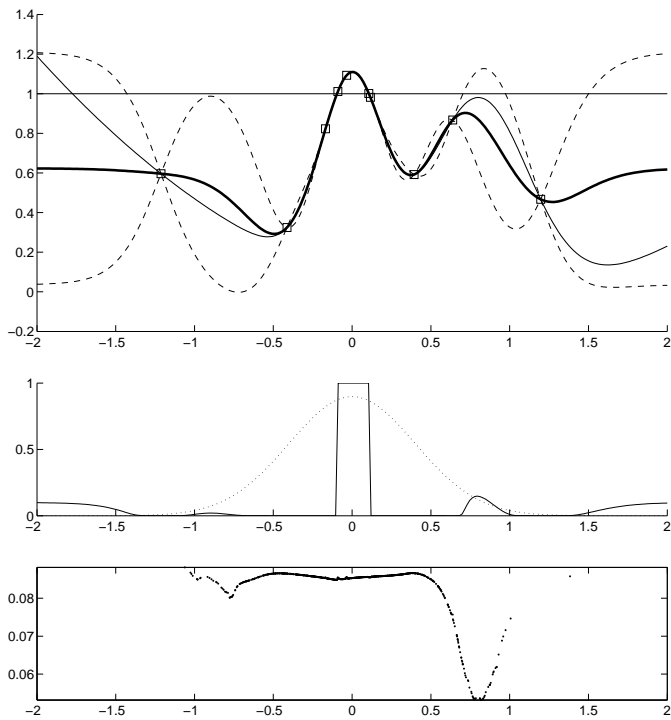


Fig. 1. Top: threshold $u = 1$ (horizontal solid line); function f (thin line); $n=10$ evaluations (squares) as obtained by the proposed algorithm using $m = 800$ for the Monte Carlo integration and $Q = 20$ for the tagged partition; kriging approximation f_n (thick line); 95% confidence intervals computed from the kriging variance (dashed lines). Middle: probability of excursion (solid line); probability density of $\mathbb{P}_{\mathbb{X}}$ (dotted line). Bottom: graph of $\Upsilon''(x_i)$, $i = 1, \dots, m = 800$, the minimum of which indicates where the next evaluation of f should be done (i.e., at approximately 0.75).

4.2 Probability of failure of a regulated system

A typical example for which our method is particularly relevant is the estimation of a probability of failure of

a system described by partial differential equations with complex boundary conditions and a few uncertain parameters.

We choose, however, to consider a much simpler illustrative problem, for two reasons. First, it is possible, within the space allowed, to give enough details to allow the reader to use it and to compare the performance of our approach with other methods not considered here. Second, nothing is lost by considering such an example, as the methodology would be strictly the same for a more expensive-to-simulate model.

We consider a simple continuous-time dynamical system, corresponding to a two-compartment model (Figure 2).

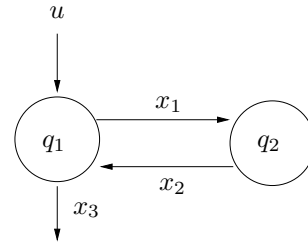


Fig. 2. A two-compartment model.

The state variables q_1 and q_2 represent an amount of material in each compartment, the positive real numbers x_1, x_2, x_3 correspond to exchange rates and u is the control input of the system. The state vector is governed by the ordinary differential equation

$$\begin{pmatrix} \dot{q}_1 \\ \dot{q}_2 \end{pmatrix} = \begin{pmatrix} -x_1 - x_3 & x_2 \\ x_1 & -x_2 \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} + u \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

We assume that x_3 is known and equal to 0.35. However, we suppose that the parameters x_1 and x_2 are uncertain. We model this uncertainty by a probability distribution $\mathbb{P}_{\mathbb{X}}$ centered on $(x_1, x_2) = (0.6, 0.15)$ (see Figure 3).

Our objective is to regulate the output variable q_2 , by means of a proportional-integral controller, i.e. a linear feedback corrector whose transfer function can be written as

$$G(s) = K \frac{a + s}{s}.$$

The value of K and a of the corrector were chosen to insure satisfactory phase and gain margins for $(x_1, x_2) = (0.6, 0.15)$. We obtained $K = 0.93$ and $a = 0.23$. To measure the performance of the system, we consider the settling time $f(x_1, x_2)$ of the step response of system — the response is considered to be settled at time s when $|q_2(t) - \lim_{t \rightarrow \infty} q_2(t)|$ becomes smaller than a fraction equal to 0.02 of its final value, $\forall t > s$. For instance, we have $f(0.6, 0.15) = 10.95$. We would like to estimate the probability $\alpha(f) = \mathbb{P}_{\mathbb{X}}(A(f))$ of the excursion set

$$A(f) = \{(x_1, x_2) \in \mathbb{R}^2 : f(x_1, x_2) \geq u = 12.5\},$$

i.e. the probability that the system has a settling time greater than 12.5, given the probability distribution $\mathbb{P}_{\mathbb{X}}$ on the uncertain parameters. The result of our sequential algorithm is presented on Figure 3. After $n = 40$ iterations,

the unknown settling-time function f has been sampled so that the probability of excursion $P(\{f : f(x_1, x_2) > u\} | \mathcal{F}_n)$ is very close to either 0 or 1 in the region where the density of $P_{\mathbf{x}}$ is high.

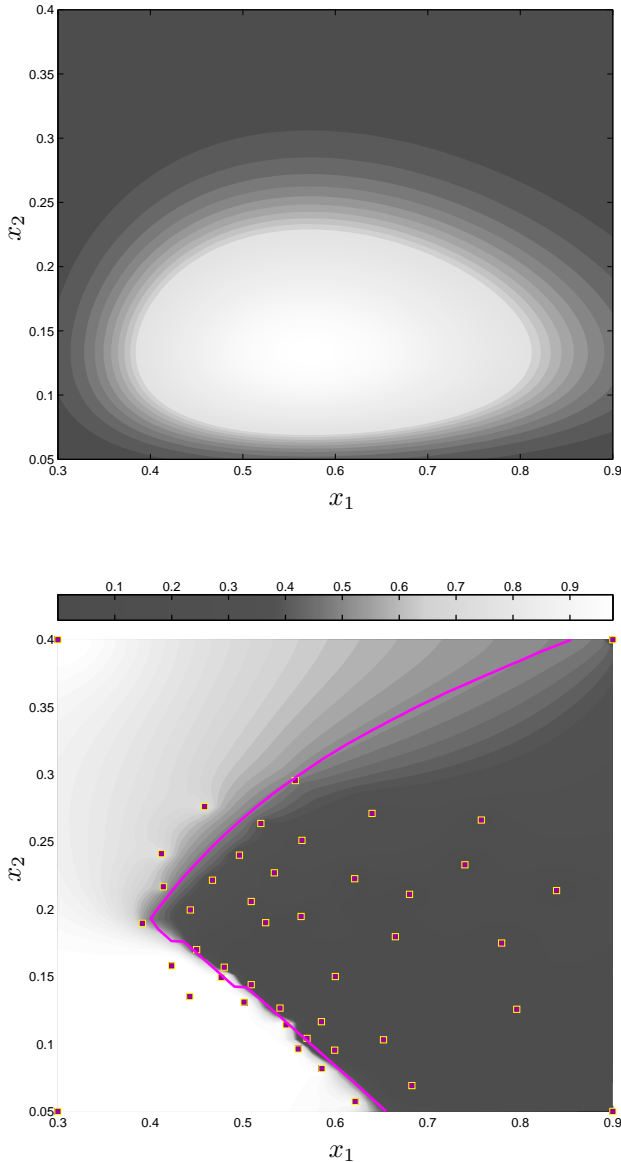


Fig. 3. Top: gray-level plot of the probability density of $P_{\mathbf{x}}$ along the parameters x_1 and x_2 of the two-compartment model. The bright areas correspond to high values of the density. Bottom: gray-level plot of the probability of excursion conditioned on the evaluations of the settling time, whose locations are indicated by square dots. The solid line indicates the frontier ∂A of the failure domain. One can see that the probability of excursion is very close to either 0 or 1 in the region where the density of $P_{\mathbf{x}}$ is high.

ACKNOWLEDGEMENTS

This work was partially supported by the French *Agence Nationale de la Recherche* in the context of the OPUS-COPRIN project (ref. ANR-07-CIS7-010 and ANR-07-TLOG-015).

REFERENCES

- S. K. Au and J. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Probab. Engrg. Mechan.*, 16(4):263–277, 2001.
- J. O. Berger. *Statistical Decision Theory and Bayesian Analysis (second edition)*. Springer, 1985.
- P. Bjerager. On computational methods for structural reliability analysis. *Structural Safety*, 9:76–96, 1990.
- K. Chaloner and I. Verdinelli. Bayesian experimental design: A review. *Statist. Sci.*, 10(3):273–304, 1995.
- J.-P. Chilès and P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York, 1999.
- C. Currin, T. Mitchell, M. Morris, and D. Ylvisaker. Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association*, 86(416):953–963, 1991.
- D. Geman and B. Jedynek. An active testing model for tracking roads in satellite images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(1):1–14, 1996.
- J. Mockus. Application of Bayesian approach to numerical methods of global and stochastic optimization. *J. Global Optim.*, 4:347–365, 1994.
- L. Pronzato. Optimal experimental design and some related control problems. *Automatica*, pages 303–325, 2008.
- R. Y. Rubinstein. The cross-entropy method for combinatorial and continuous optimization. *Methodol. Comput. Appl. Probab.*, 2:127–190, 1999.
- J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409–435, 1989.
- M. Schonlau and W. Welch. Global optimization with non-parametric function fitting. In *Proc. Section on Physical and Engineering Sciences*, pages 183–186. American Statistical Association, 1996.
- M. L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York, 1999.
- E. M. Vestrup. *The Theory of Measures and Integration*. Wiley, 2003.
- J. Villemonteix, E. Vazquez, and E. Walter. An informational approach to the global optimization of expensive-to-evaluate functions. *J. Global Optim.*, to appear.
- W. J. Welch, R. J. Buck, J. Sacks, H. P. Wynn, T. J. Mitchell, and M. D. Morris. Screening, predicting and computer experiments. *Technometrics*, 34:15–25, 1992.
- A. Zilinskas. A review of statistical models for global optimization. *J. Global Optim.*, 2:145–153, 1992.