



HAL
open science

Estimating orthant probabilities of high dimensional Gaussian vectors with an application to set estimation

Dario Azzimonti, David Ginsbourger

► **To cite this version:**

Dario Azzimonti, David Ginsbourger. Estimating orthant probabilities of high dimensional Gaussian vectors with an application to set estimation. 2016. hal-01289126v1

HAL Id: hal-01289126

<https://hal.science/hal-01289126v1>

Preprint submitted on 16 Mar 2016 (v1), last revised 30 Nov 2018 (v3)

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.

Estimating orthant probabilities of high dimensional Gaussian vectors with an application to set estimation

Dario Azzimonti ^{*} David Ginsbourger ^{†*}

Abstract

The computation of Gaussian orthant probabilities has been extensively studied for low dimensional vectors. Here we focus on the high dimensional case and we present a two step procedure relying on both deterministic and stochastic techniques. The proposed estimator relies indeed on splitting the probability into a low dimensional term and a remainder. While the low dimensional probability can be estimated by fast and accurate quadrature, the remainder requires Monte Carlo sampling. We show that an estimator obtained with this technique has higher efficiency than standard Monte Carlo methods. We further refine the estimation by using a novel asymmetric nested Monte Carlo algorithm for the remainder and we highlight cases where this approximation brings substantial efficiency gains. Finally this method is applied to derive conservative estimates of excursion sets of expensive to evaluate deterministic functions under a Gaussian random field prior without requiring a Markov assumption.

Keywords: Gaussian probabilities; Monte Carlo; Gaussian random fields; Conservative set estimation.

1 Introduction

Assume that $X = (X_1, \dots, X_d)$ is a random vector with Gaussian distribution $N_d(\mu, \Sigma)$. We are interested in estimating, for any fixed $t \in \mathbb{R}$, the following probability

$$\pi(t) = P(X \leq (t, \dots, t)). \quad (1)$$

The general problem of evaluating $\pi(t)$, which, for a full rank matrix Σ , is the integral of the multivariate normal density $\phi(\cdot; \mu, \Sigma)$ over the one-sided d -dimensional rectangle $(-\infty, t]^d$, has been extensively studied in moderate dimensions with many different methods. In low dimensions tables are available (see, e.g., Owen (1956) for $d = 2$). Furthermore, when the dimension is smaller than 20, there exist methods (see, e.g., Abrahamson (1964), Moran (1984) and Miwa et al. (2003)) exploiting the specific orthant structure of the probability in (1). Currently, however, most of the literature uses numerical integration techniques to approximate the quantity. In moderate dimensions fast reliable methods are established to approximate $\pi(t)$ (see, e.g. Cox and Wermuth (1991)) and more recently the methods introduced in Schervish (1984);

^{*}IMSV, Department of Mathematics and Statistics, University of Bern, Alpeneggstrasse 22, 3012 Bern, Switzerland

[†]Idiap Research Institute, Centre du Parc, Rue Marconi 19, PO Box 592, 1920 Martigny, Switzerland.

Genz (1992) and Hajivassiliou et al. (1996) (see also Genz and Bretz (2002), Ridgway (2014) and the book Genz and Bretz (2009) for a broader overview) provide state-of-the-art algorithms when $d < 100$. Those techniques rely on fast quasi Monte Carlo (qMC) methods and are very accurate for moderate dimensions. However, when d is larger than 1000 they are not computationally efficient or become intractable. Commonly used alternative methods are standard Monte Carlo (MC) techniques (see Tong (2012), Chapter 8 for an extensive review), for which getting accurate estimates can be computationally prohibitive.

We propose here a two step method that exploits the power of qMC quadratures and the flexibility of stochastic simulation. We rely on the following equivalent formulation.

$$\pi(t) = 1 - P(\max X > t),$$

where $\max X$ denotes $\max_{i=1,\dots,d} X_i$. In the following we fix t and denote $p = P(\max X > t)$.

The central idea here is using a moderate dimensional subvector of X to approximate p and then correcting bias by MC. Let us fix $q \ll d$ and define the active dimensions as $E_q = \{i_1, \dots, i_q\} \subset \{1, \dots, d\}$. Let us further denote with X^q the q dimensional vector $X^q = (X_{i_1}, \dots, X_{i_q})$ and with X^{-q} the $(d - q)$ dimensional vector $X^{-q} = (X_j)_{j \in E \setminus E_q}$. Then,

$$\begin{aligned} p &= P(\max X > t) = p_q + (1 - p_q)R_q, \\ p_q &= P(\max X^q > t), \\ R_q &= P(\max X^{-q} > t \mid \max X^q \leq t). \end{aligned} \tag{2}$$

The quantity p_q is always smaller or equal to p as $E_q \subset \{1, \dots, d\}$. Selecting a non-degenerate vector X^q , we propose to estimate p_q with the QRSVN algorithm (Genz et al., 2012) which is efficient as we choose a number of active dimensions q much smaller than d .

In Chevalier (2013), Chapter 6, the similar problem of approximating the non-exceedance probability of the maximum of a Gaussian random field Z based on a few well-selected points is presented. In that setting each component of X stands for the value of Z at one point of a discretization of the index set. Active dimensions (i.e. the well-selected points) were chosen by numerically maximizing p_q , and the remainder was not accounted for. Here a full optimization of the active dimensions is not needed as we, instead, exploit the decomposition in (2) to correct the error introduced by p_q . For this task, the remainder R_q is estimated with a standard MC technique and a novel asymmetric nested Monte Carlo (anMC) algorithm. The anMC technique draws samples by taking into account the computational cost, resulting in a more efficient estimator.

In the remainder of the paper, we propose an unbiased estimator for p and we compute its variance in Section 2. In Section 3 we introduce the anMC algorithm in the more general setting of estimating expectations depending on two vectors with different simulation costs. It is then explicitly applied to efficiently estimate R_q . Finally, in Section 4, we show an implementation of this method to compute conservative estimates of excursion sets for expensive to evaluate functions under non-necessarily Markovian Gaussian random field priors. In Appendix B, we present two heuristic methods to select active dimensions. All proofs are in Appendix A.

2 The estimator properties

2.1 An unbiased estimator for p

Equation (2) gives us a decomposition that can be exploited to obtain an unbiased estimator for p . In the following proposition we define the estimator and we compute its variance.

Proposition 1. *Consider \widehat{p}_q and \widehat{R}_q , independent unbiased estimators of p_q and R_q respectively, then $\widehat{p} = \widehat{p}_q + (1 - \widehat{p}_q)\widehat{R}_q$ is an unbiased estimator for p . Moreover its variance is*

$$\text{var}(\widehat{p}) = (1 - R_q)^2 \text{var}(\widehat{p}_q) + (1 - p_q)^2 \text{var}(\widehat{R}_q) + \text{var}(\widehat{p}_q) \text{var}(\widehat{R}_q). \quad (3)$$

In what follows we present options for \widehat{p}_q and \widehat{R}_q that form an efficient computational strategy.

2.2 Quasi Monte Carlo estimator for p_q

The quantity p_q can also be computed as

$$p_q = 1 - P(X^q \leq t_q),$$

where t_q denotes the q dimensional vector (t, \dots, t) . The approximation of p_q thus requires only an evaluation of the cumulative distribution function of X^q , selected as non-degenerate. Since we assume that $q \ll d$, then the dimension is low and we propose to estimate p_q with the estimator \widehat{p}_q that uses the method QRSVN introduced in Genz (1992), Hajivassiliou et al. (1996).

This method computes a randomized quasi Monte Carlo integration of the normal density. The estimate's error is approximated with the variance of the randomized integration. The quantity \widehat{p}_q^G obtained with this procedure is an unbiased estimator of p_q , see Genz and Bretz (2009).

Figure 1 shows the boxplots of 30 replications of an experiment where \widehat{p}_q^G is used to approximate p . The dimension of the vector X is $d = 1000$, the threshold is fixed at $t = 11$. In particular the vector X comes from the discretization of a six dimensional Gaussian random field on the first 1000 points of the Sobol' sequence (Bratley and Fox, 1988). The Gaussian random field was chosen with tensor product Matérn ($\nu = 5/2$) covariance kernel and a non constant mean function m . The hyperparameters of the covariance kernel were fixed as $\theta = [0.5, 0.5, 1, 1, 0.5, 0.5]^T$ and $\sigma^2 = 8$, see Rasmussen and Williams (2006), Chapter 4, for details on the parametrization. The active points were chosen with Method 1 reviewed in Appendix B. As the number of active dimensions increases both the variance and the error of \widehat{p}_q^G with respect to p decrease. \widehat{p}_q^G gives an inexpensive estimation of most of the probability mass with as few as 40 active dimensions, however it is intrinsically biased as an estimator of p . Estimating R_q enables to correct the bias of this first step.

2.3 Monte Carlo estimator for R_q

Debiasing \widehat{p}_q^G as an estimator of p can be done at the price of estimating

$$R_q = P(\max X^{-q} > t \mid \max X^q \leq t).$$

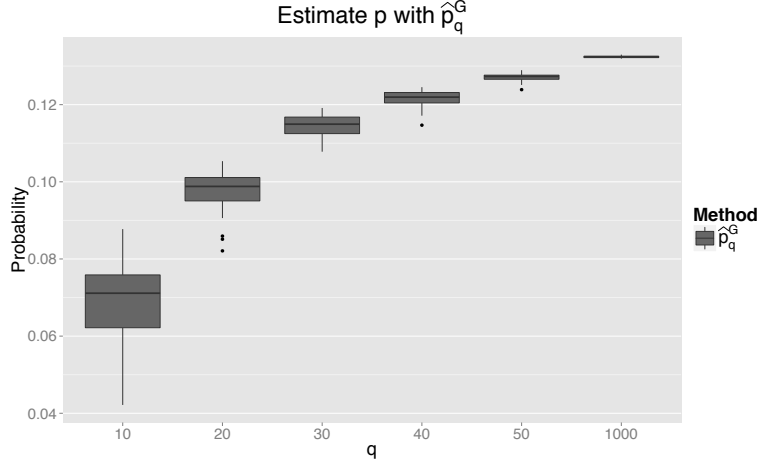


Figure 1: Estimate of $p = P(\max X \geq t)$ with \hat{p}_q^G for different values of q , the number of active dimensions.

There is no close formula for R_q , so it is approximated here via MC. Since X is Gaussian then so are X^q , X^{-q} and $X^{-q} \mid X^q = x^q$, for any deterministic vector $x^q \in \mathbb{R}^q$.

In order to estimate $R_q = P(\max X^{-q} > t \mid X_{i_1} \leq t, \dots, X_{i_q} \leq t)$, we first generate n realizations x_1^q, \dots, x_n^q of X^q such that $X^q \leq t_q$. Second, we compute the mean and covariance matrix of X^{-q} conditional on each realization x_l^q , $l = 1, \dots, n$ with the following formulas

$$\mu^{-q \mid x_l^q} = \mu^{-q} + \Sigma^{-q, q} (\Sigma^q)^{-1} (x_l^q - \mu^q), \quad \Sigma^{-q \mid x_l^q} = \Sigma^{-q} - \Sigma^{-q, q} (\Sigma^q)^{-1} \Sigma^{q, -q}, \quad (4)$$

where μ^q, Σ^q and μ^{-q}, Σ^{-q} are the mean vector and covariance matrix of X^q and X^{-q} respectively, $\Sigma^{-q, q}$ is the cross-covariance between the dimensions $E \setminus E_q$ and E_q , $\Sigma^{q, -q}$ is the transpose of $\Sigma^{-q, q}$. Given the mean and covariance matrix conditional on each sample x_l^q , we can easily draw a realization $y_l^{-q \mid q}$ from $X^{-q} \mid X^q = x_l^q$. Once n couples $(x_l^q, y_l^{-q \mid q}), l = 1, \dots, n$ are drawn from the respective distributions, an estimator for R_q is finally obtained as follows

$$\widehat{R}_q^{\text{MC}} = \frac{1}{n} \sum_{l=1}^n \mathbf{1}_{\max y_l^{-q \mid q} > t}.$$

The realizations of X^q are obtained with a rejection sampling algorithm (Robert, 1995; Horrace, 2005). This step is computationally expensive as many draws of the vector X^q might be rejected depending on the size of the d -dimensional rectangle $[-\infty, t]^d$. Drawing samples from the distribution of $X^{-q} \mid X^q = x_l^q$ is instead less expensive. The computation of the mean vector and covariance matrix requires only linear algebra operations as described in (4) and realizations of $X^{-q} \mid X^q = x_l^q$ can be generated by sampling from a multivariate normal distribution.

The difference in computational cost between the first step and the second step of the MC procedure can be exploited to reduce the variance. In Section 3 we present a new MC procedure that at a fixed computational cost reduces the variance of the estimate.

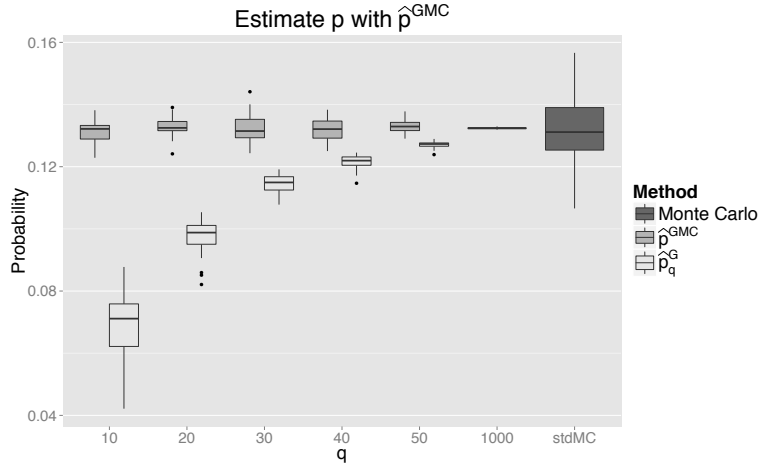


Figure 2: Estimate of p with \hat{p}^{GMC} for different values of q . A full MC estimation of the same quantity is shown for comparison

We denote with \hat{p}^{GMC} the unbiased estimator of p defined as

$$\hat{p}^{\text{GMC}} = \hat{p}_q^{\text{G}} + (1 - \hat{p}_q^{\text{G}}) \widehat{R}_q^{\text{MC}}.$$

Figure 2 shows the box plots of 30 replications of an experiment where p is approximated with \hat{p}^{GMC} . The set-up is the same as in Fig. 1. The core of the probability is approximated with \hat{p}_q^{G} and the active dimensions are chosen with Method 1. The residual R_q is estimated with $\widehat{R}_q^{\text{MC}}$. The remainder allows to correct the bias of \hat{p}_q^{G} even with a small number of active dimensions. As comparison the results of the same experiment with a full MC estimator for p are also shown. For all experiments and for each method the number of samples was chosen in order to have approximately the same computational cost. The estimator \hat{p}^{GMC} exploits an almost exact method to estimate the largest part of the probability p , therefore the MC estimator $\widehat{R}_q^{\text{MC}}$ has less variance than a full MC procedure for a fixed computational cost.

3 Estimation of the residual with asymmetric nested Monte Carlo

In section 2, R_q was estimated by $\widehat{R}_q^{\text{MC}}$. There exists many methods to reduce the variance of such estimators, including antithetic variables (Hammersley and Morton, 1956), importance sampling (Kahn, 1950; Kahn and Marshall, 1953) or conditional Monte Carlo (Hammersley, 1956) among many others, see, Lemieux (2009), Chapter 4, and Robert and Casella (2013), Chapter 4, for a broader overview. Here we propose a so-called asymmetric nested Monte Carlo (anMC) estimator for R_q that reduces the variance by a parsimonious multiple use of conditioning data.

The idea is to use an asymmetric sampling scheme that assigns the available computational resources by taking into account also the actual cost of simulating each component. This type of asymmetric sampling scheme was already introduced in the particular case of comparing the performance of stopping times for a real-valued stochastic

process in discrete times in Dickmann and Schweizer (2014). Here we introduce this procedure in a general fashion and then we detail how to use it as variance reduction for $\widehat{R}_q^{\text{MC}}$. Consider two random elements $W \in \mathcal{W}$ and $Z \in \mathcal{Z}$, defined on the same probability space and not independent. We are interested in estimating

$$G = \mathbb{E}[g(W, Z)], \quad (5)$$

where $g : \mathcal{W} \times \mathcal{Z} \rightarrow \mathbb{R}$ is a measurable function, assumed integrable with respect to (W, Z) 's probability measure. Let us also assume that it is possible to draw realizations from the marginal distribution of W , Z and from the conditional distribution of $Z | W = w_i$, for each w_i sample of W . We can then obtain realizations (w_i, z_i) , $i = 1, \dots, n$ of (W, Z) by simulating w_i from the distribution of W and then z_i from the conditional distribution $Z | W = w_i$, leading to:

$$\widehat{G} = \frac{1}{n} \sum_{i=1}^n g(w_i, z_i). \quad (6)$$

This MC estimator can actually be seen as the result of a two step nested MC procedure where, for each realization w_i , one inner sample z_i is drawn from $Z | W = w_i$. Note that the estimator $\widehat{R}_q^{\text{MC}}$ used in Section 2 is a particular case of Equation (6) with $W = X^q | X^q \leq t_q$, $Z = X^{-q}$ and $g(x, y) = \mathbf{1}_{\max y > t}$. As noted in Section 2, drawing realizations of $X^q | X^q \leq t_q$ has a higher computational cost than simulating X^{-q} because rejection sampling is required. More generally, if we denote with $C_W(n)$ the cost of n realizations of W and with $C_{Z|W}(m; w_i)$ the cost of drawing m conditional simulations from $Z | W = w_i$, then sampling several conditional realizations for a given w_i might bring savings if $C_W(1)$ is much higher than $C_{Z|W}(1; w_i)$.

In the proposed asymmetric sampling scheme for each realization w_i we sample m realizations $z_{i,1}, \dots, z_{i,m}$ from $Z | W = w_i$. Assume that we sample with this scheme the couples $(w_i, z_{i,j})$, $i = 1, \dots, n$, $j = 1, \dots, m$, then we can write the following estimator for G

$$\widetilde{G} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m g(w_i, z_{i,j}). \quad (7)$$

For a fixed number of samples, the estimator \widetilde{G} may have a higher variance than \widehat{G} due to the dependency between pairs sharing the same replicate of W . However, in many cases, it may be more relevant to focus on obtaining good estimates within a fixed time. If we set the computational budget instead of the number of samples and if $C_{Z|W}$ is smaller than C_W , then anMC may lead to an overall variance reduction thanks to an increased number of simulated pairs. We show in the remainder of this section that, in the case of an affine cost function $C_{Z|W}$, there exists an optimal number of inner simulations m diminishing the variance of \widetilde{G} below that of \widehat{G} . Assume

$$C_W(n) = cn \text{ and, for each sample } w_i \\ C_{Z|W}(m; w_i) = C_{Z|W}(m) = \alpha + \beta m,$$

with $c, \alpha, \beta \in \mathbb{R}_+$ dependent on the simulators chosen for W and $Z | W$. The second equation entails that the cost of conditional simulations does not depend on the conditioning value.

If $W = X^q | X^q \leq t_q$, $Z = X^{-q}$ as in Section 2, then $Z | W$ is Gaussian with mean and covariance matrix described in (4). In this case the cost for sampling $Z | W$

is affine, with α describing pre-calculation times and β random number generation and algebraic operations.

Denote with W_1, \dots, W_n replications of W . For each W_i we consider the conditional distribution $Z \mid W_i$ and $Z_{1,i}, \dots, Z_{m,i}$ replications from it. We study here the properties of \tilde{G} when the total simulation budget, denoted $C_{\text{tot}}(n, m)$ is fixed to $C_{\text{fix}} \in \mathbb{R}_+$. First observe that

$$C_{\text{tot}}(n, m) = n(c + \alpha + \beta m).$$

Then we can derive the number of replications of W as a function of m :

$$N_{C_{\text{fix}}}(m) = \frac{C_{\text{fix}}}{c + \alpha + \beta m}.$$

The following proposition shows a decomposition of $\text{var}(\tilde{G})$ that is useful to find the optimal number of simulations m^* under a fixed simulation budget $C_{\text{tot}}(n, m) = C_{\text{fix}}$.

Proposition 2. *Consider n independent copies W_1, \dots, W_n of W and, for each W_i , m copies $Z_{i,j} = Z_j \mid W_i$ $j = 1, \dots, m$, independent conditionally on W_i . Then,*

$$\text{var}(\tilde{G}) = \frac{1}{n} \text{var}(g(W_1, Z_{1,1})) - \frac{m-1}{nm} \mathbb{E}[\text{var}(g(W_1, Z_{1,1}) \mid W_1)]. \quad (8)$$

Corollary 1. *Under the same assumptions, \tilde{G} has minimal variance when*

$$m = \tilde{m} = \sqrt{\frac{(\alpha + c)B}{\beta(A - B)}},$$

where $A = \text{var}(g(W_1, Z_{1,1}))$ and $B = \mathbb{E}[\text{var}(g(W_1, Z_{1,1}) \mid W_1)]$. Moreover denote with $\varepsilon = \tilde{m} - \lfloor \tilde{m} \rfloor$, then the optimal integer is $m^* = \lfloor \tilde{m} \rfloor$ if

$$\varepsilon < \frac{(2\tilde{m} + 1) - \sqrt{4(\tilde{m})^2 + 1}}{2} \quad (9)$$

or $m^* = \lceil \tilde{m} \rceil$ otherwise.

Proposition 3. *Under the same assumptions, if $m^* > \frac{2(\alpha+c)B}{(c+\alpha)B+\beta(A-B)}$ then $\text{var}(\tilde{G}) = \text{var}(\hat{G}) [1 - \eta]$, where $\eta \in (0, 1)$.*

3.1 Algorithmic considerations

In order to compute m^* , we need to know $A = \text{var}(g(W_1, Z_{1,1}))$ and $B = \mathbb{E}[\text{var}(g(W_1, Z_{1,1}) \mid W_1)]$ and the constants c , α and β . A and B depend on the specific problem at hand and are usually not known in advance. Part of the total computational budget is then needed to estimate A and B . This preliminary phase is also used to estimate the system dependent constants c and β . Algorithm 1 reports the pseudo-code for anMC.

Input : $\mu_W, \mu_Z, \Sigma_W, \Sigma_Z, \Sigma_{WZ}, g, C_{\text{tot}}$
Output: \tilde{G}

Part 0: estimate c, β, α ;
initialize compute the conditional covariance $\Sigma_{Z|W}$ and initialize n_0, m_0 ;
Part 1: **for** $i \leftarrow 1$ **to** n_0 **do**
estimate A, B | simulate w_i from the distribution of W ;
| compute the conditional mean $\mu_{Z|W=w_i}$;
| draw m_0 simulations $z_{i,1}, \dots, z_{i,m_0}$ from the conditional distribution $Z | W = w_i$;
| estimate $\mathbb{E}[g(W, Z) | W = w_i]$ with $\tilde{E}_i = \frac{1}{m_0} \sum_{j=1}^{m_0} g(w_i, z_{i,j})$;
| estimate $\text{var}(g(W, Z) | W = w_i)$ with $\tilde{V}_i = \frac{1}{m_0-1} \sum_{j=1}^{m_0} (g(w_i, z_{i,j}) - \tilde{E}_i)^2$;
| compute $\tilde{m} = \sqrt{\frac{(\alpha+c) \frac{1}{n_0} \sum_{i=1}^{n_0} \tilde{V}_i}{\beta \frac{1}{n_0-1} \sum_{i=1}^{n_0} (\tilde{E}_i - \frac{1}{n_0} \sum_{i=1}^{n_0} \tilde{E}_i)^2}}$, m^* as in Corollary 1 and $n^* = N_{C_{\text{fix}}}(m^*)$;
Part 2: **for** $i \leftarrow 1$ **to** n^* **do**
compute \tilde{G} | **if** $i \leq n_0$ **then**
| | **for** $j \leftarrow 1$ **to** m^* **do**
| | | **if** $j \leq m_0$ **then**
| | | | use previously calculated \tilde{E}_i and \tilde{V}_i ;
| | | **else**
| | | | simulate $z_{i,j}$ from the distribution $Z | W = w_i$;
| | | | compute $\tilde{E}_i = \frac{1}{m^*} \sum_{j=1}^{m^*} g(w_i, z_{i,j})$;
| | **else**
| | | simulate w_i from the distribution of W ;
| | | compute the conditional mean $\mu_{Z|W=w_i}$;
| | | **for** $j \leftarrow 1$ **to** m^* **do**
| | | | simulate $z_{i,j}$ from the conditional distribution $Z | W = w_i$;
| | | | compute $\tilde{E}_i = \frac{1}{m^*} \sum_{j=1}^{m^*} g(w_i, z_{i,j})$;
| estimate $\mathbb{E}[g(W, Z)]$ with $\tilde{G} = \frac{1}{n^*} \sum_{i=1}^{n^*} \tilde{E}_i$;

Algorithm 1: Asymmetric nested Monte Carlo.

3.2 Estimate p with \hat{p}^{GanMC}

The anMC algorithm can be used to reduce the variance compared to R_q 's MC estimate proposed in Section 2.3. In fact, let us consider $W = X^q | X^q \leq t_q$ and $Z = X^{-q}$. We have that W is expensive to simulate as it requires rejection sampling while, for a given sample w_i , $Z | W = w_i$ is Gaussian with mean and covariance matrix described in Equation (4). It is much cheaper to obtain samples from $Z | W = w_i$ than from W . Moreover, as noted earlier, R_q can be written in the form of Equation (5) with $g(x, y) = \mathbf{1}_{\max y > t}$. We can then use Algorithm 1 to calculate m^* , sample n^* realizations w_1, \dots, w_{n^*} of W and for each realization w_i obtain m^* samples $z_{i,1}, \dots, z_{i,m^*}$ of $Z | W = w_i$. Then we can estimate R_q via

$$\widehat{R}_q^{\text{anMC}} = \frac{1}{n^* m^*} \sum_{i=1}^{n^*} \sum_{j=1}^{m^*} \mathbf{1}_{\max z_{i,j} > t}.$$

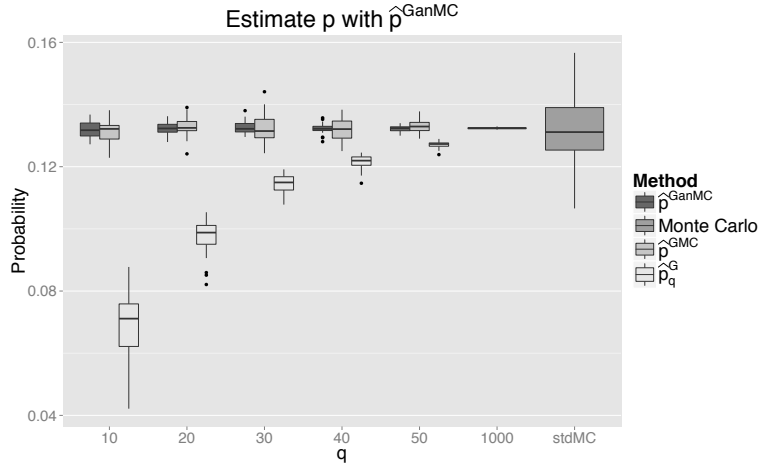


Figure 3: Comparison of results with \hat{p}_q^G , \hat{p}^{GMC} and \hat{p}^{GanMC} on the example introduced in Fig. 1.

Finally plugging in \hat{R}_q^{anMC} and \hat{p}_q^G in Equation (2), we obtain

$$\hat{p}^{\text{GanMC}} = \hat{p}_q^G + (1 - \hat{p}_q^G) \hat{R}_q^{\text{anMC}}.$$

Figure 3 shows a comparison of results using 30 replications of the experiment presented in Section 2.3. Results obtained with a MC estimator are shown for comparison.

While the simulations of all experiment were obtained under the constraint of a fixed computational cost, the actual time to obtain the simulations was not exactly the same. In order to be able compare the methods in more general settings we further rely on the notion of efficiency. For an estimator \hat{p} , we define the efficiency (Lemieux (2009), Section 4.2) as

$$\text{Eff}[\hat{p}] = \frac{1}{\text{var}(\hat{p}) \text{time}[\hat{p}]},$$

where $\text{time}[\hat{p}]$ denotes the computational time of the estimator \hat{p} .

Figure 4 shows a comparison of the efficiency of \hat{p}^{GMC} and \hat{p}^{GanMC} with a full Monte Carlo estimator. With as few as $q = 50$ active dimensions we obtain an increase in efficiency of around 10 times on average over the 30 replications of the experiment with the estimator \hat{p}^{GMC} . The estimator \hat{p}^{GanMC} shows a higher median efficiency than the others for all $q \geq 20$.

4 Application: efficient computation of conservative estimates

We show here that anMC is key in conservative excursion set estimation relying on Gaussian field models. We consider an expensive to evaluate system described by a continuous function $f : D \subset \mathbb{R}^\ell \rightarrow \mathbb{R}$, $\ell \geq 1$, where D is a compact domain, and we focus on estimating, for some fixed threshold $t \in \mathbb{R}$, the set

$$\Gamma^* = \{x \in D : f(x) \leq t\}.$$

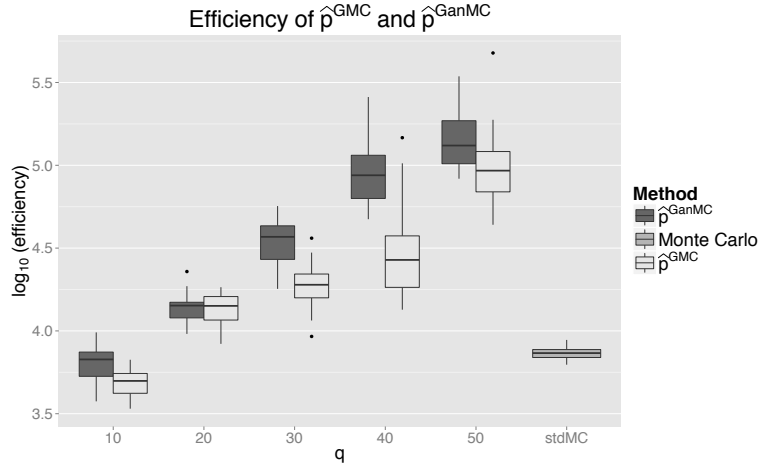


Figure 4: Efficiency of the estimators \hat{p}^{GMC} and \hat{p}^{GanMC} compared with the efficiency of a standard MC estimator on 30 replications of the experiment from Fig. 3. Values in logarithmic scale.

Such problems arise in many applications such as reliability engineering (see, e.g., Picheny et al. (2013), Chevalier et al. (2014a)) climatological studies (Bolin and Lindgren, 2015; French and Sain, 2013) or in natural sciences (Bayarri et al., 2009). Often f can only be evaluated with computer simulations and is seen as expensive to evaluate black-box (Sacks et al., 1989). In practice we assume here that f was only evaluated at points $\chi_k = \{x_1, \dots, x_k\} \subset D$ and the associated responses are denoted with $f_k = (f(x_1), \dots, f(x_k)) \in \mathbb{R}^k$ and we are interested in giving an estimate of Γ^* starting from these k evaluations.

In a Bayesian framework we consider f as one realization of a Gaussian random field $(Z_x)_{x \in D}$ with prior mean function m and covariance kernel \mathfrak{K} . A prior distribution of the excursion set is hence obtained by thresholding Z , thus obtaining the following random closed set

$$\Gamma = \{x \in D : Z_x \leq t\}.$$

Denoting with Z_{χ_k} the random vector $(Z_{x_1}, \dots, Z_{x_k})$, we can then condition Z on the observations f_k and obtain a posterior distribution for the field $Z_x \mid Z_{\chi_k} = f_k$. This gives rise to a posterior distribution for Γ . Different definitions of random closed set expectation (Molchanov (2005), Chapter 2) can be used to summarize the posterior distribution of Γ and to provide estimates for Γ^* . In Chevalier et al. (2013), for example, the Vorob'ev expectation was introduced in this setting. Let us recall here the basic tools needed to compute this estimate. We denote with $p_{\Gamma,k} : D \rightarrow [0, 1]$ the coverage function of the posterior set $\Gamma \mid Z_{\chi_k} = f_k$, defined as

$$p_{\Gamma,k}(x) = P_k(x \in \Gamma), \quad x \in D,$$

where $P_k(\cdot) = P(\cdot \mid Z_{\chi_k} = f_k)$. This function associates to each point in D its probability of being inside the posterior excursion set. The function $p_{\Gamma,k}$ gives rise to a family of excursion set estimates; in fact, for each $\rho \in [0, 1]$ we can define the posterior ρ -level Vorob'ev quantile of Γ

$$Q_\rho = \{x \in D : p_{\Gamma,k}(x) \geq \rho\}.$$

The Vorob'ev expectation of Γ (Molchanov, 2005) is the quantile Q_{ρ_V} that satisfies $|Q_\rho| \leq \mathbb{E}_k[|\Gamma|] \leq |Q_{\rho_V}|$ for all $\rho \geq \rho_V$, where $|A|$ denotes the volume of a set $A \subset \mathbb{R}^l$. This set expectation consists of the points that have high enough marginal probability of being inside the excursion set. The lower limit for the marginal probability is chosen in order to approximate well the volume of the set. In some applications, however, it is important to provide confidence statements on the whole set estimate. Conservative estimates introduced in Bolin and Lindgren (2015) for Gaussian Markov random fields address this issue. A conservative estimate of Γ^* is

$$C_{\Gamma,k} = \arg \max_{C \subset D} \{ |C| : P_k(C \subset \{Z_x \leq t\}) \geq \alpha \}, \quad (10)$$

where $|C|$ denotes the volume of C . This definition however leads to major computational issues.

First of all we need to select a family of sets to use for the optimization procedure in Equation (10). Here we follow Bolin and Lindgren (2015) and select the Vorob'ev quantiles as family of sets. This family has the advantage that it is parametrized by one real number ρ and thus it renders the optimization straightforward. We use here the dichotomy algorithm detailed in Algorithm 2.

Second, for each candidate Q we need to evaluate $P_{\text{next}} = P_k(Q \subset \{Z_x \leq t\})$, the probability that Q is inside the excursion. In fact, this quantity is a high dimensional orthant probability. For the Vorob'ev quantile $Q_{\rho'}$, consider the discretization over the points c_1, \dots, c_r , then

$$P_k(Q_{\rho'} \subset \{Z_x \leq t\}) = P_k(Z_{c_1} \leq t, \dots, Z_{c_r} \leq t) = 1 - P_k(\max_{i=1, \dots, r} Z_{c_i} > t).$$

Thus we use the estimator \hat{p}^{GanMC} to approximate $1 - P_k(Q_{\rho'} \subset \{Z_x \leq t\})$. The use of anMC allows a discretization of the Vorob'ev quantiles at resolutions that seem out of reach otherwise.

Input :

- m_k, \mathfrak{K}_k , conditional mean and covariance of $Z \mid Z_{\chi_k} = f_k$;
- fine discretization design G ;

Output: Conservative estimate for Γ^* at level α .

Part 0: sort the points in G in decreasing order of $p_{\Gamma,k}$, with indices $G_s = \{i_1, \dots, i_m\}$;

compute i_B, i_T find the highest index i_T such that $\prod_{j=1}^T p_{\Gamma,k}(G_s)[i_j] \geq \alpha$;
find the highest index i_B such that $p_{\Gamma,k}(G_s)[i_B] \geq \alpha$;
evaluate mean and covariance matrix $m_k(i_B)$ and Σ_{i_B, i_B} ;

Part 1: initialize $i_L = i_T, i_R = i_B$;

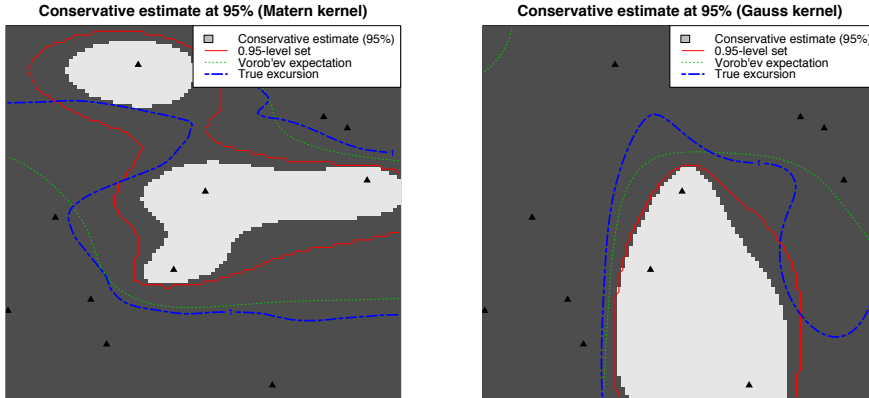
Initialize dichotomy compute $P_L = P_k(Q_{\rho_{i_L}} \subset \{Z_x \leq t\}), P_R = P_k(Q_{\rho_{i_R}} \subset \{Z_x \leq t\})$;

Part 2: **while** $P_R < \alpha$ and $(i_R - i_L) \geq 2$ **do**

optimization next evaluation $i_{\text{next}} = \frac{i_L + i_R}{2}$;
 compute $P_{\text{next}} = P_k(Q_{\rho_{i_{\text{next}}}} \subset \{Z_x \leq t\})$;
 if $P_{\text{next}} \geq \alpha$ **then**
 | $i_L = i_{\text{next}}, i_R = i_R$;
 else
 | $i_L = i_L, i_R = i_{\text{next}}$;

Algorithm 2: Conservative estimates algorithm.

We apply Algorithm 2 to a two dimensional artificial test case. We consider as function f a realization of a Gaussian field $(Z_x)_{x \in D}$, where $D \subset \mathbb{R}^2$ is the unit square.



(a) Realization obtained with a Matérn kernel. (b) Realization obtained with Gaussian kernel.

Figure 5: Conservative estimates at 95% (white region) for the excursion below $t = 1$. Both models are based on 15 evaluations of the function (black triangles). The true excursion level is plotted in blue, the Vorob'ev expectation in green and the 0.95-level set in red.

We consider two parametrizations for the prior covariance kernel: a tensor product Matérn covariance kernel with $\nu = 5/2$, variance $\sigma^2 = 0.5$ and range parameters $\theta = [0.4, 0.2]$ and a Gaussian covariance kernel with variance $\sigma^2 = 0.5$ and range parameters $\theta = [0.2, 0.4]$. In both cases we assume a prior constant mean function. We are interested in the set Γ^* with $t = 1$. For both cases we consider $k = 15$ evaluations of f at the same points chosen by Latin hypercube sampling. Figures 5a and 5b show the conservative estimate at level 95% compared with the true excursion, the Vorob'ev expectation and the 0.95-quantile for the Matérn and the Gaussian kernel. In both cases we see that the 0.95-quantile does not guarantee that the estimate is included in the true excursion with probability 0.95. The conservative estimates instead are guaranteed to be inside the true excursion with probability $\alpha = 0.95$. They correspond to Vorob'ev quantiles at levels 0.998 and 0.993 for Matérn and Gaussian respectively. The conservative estimates were obtained with a 100×100 discretization of the unit square. Such high resolution grids lead to very high dimensional probability calculations. In fact, the dichotomy algorithm required 11 computations of the probability $1 - P_k(Q_{\rho'} \subset \{Z_x \leq t\})$ for each case. The discretization's size for Q_{ρ} varied between 1213 and 3201 points in the Matérn kernel case and between 1692 and 2462 points in the Gaussian case. Such high dimensional probabilities cannot be computed with the current implementation of the algorithm by Genz, however they could also be computed with a standard Monte Carlo at very high computational costs. Instead, with the proposed method, the total computational time on a laptop with Intel Core i7 1.7GHz CPU and 8GB of RAM was equal to 365 and 390 seconds respectively for Matérn and Gaussian kernel.

5 Discussion

In this paper we introduced a new method to approximate high dimensional orthant Gaussian probabilities. The procedure resulted in estimators with greater efficiency than standard Monte Carlo, scalable on dimensions larger than 1000. The remainder R_q in the decomposition of Equation (2) was estimated with standard Monte Carlo and with a novel asymmetric nested Monte Carlo (anMC) procedure. The anMC estimator showed an increase in efficiency in the simulation studies presented. The efficiency of the overall method depends on the general structure of the Gaussian vector and there are situations where it brings only moderate improvements over a standard Monte Carlo approach. A study of the relationships between the covariance structure and the efficiency of the method might be beneficial for understanding this behaviour. The issue of choosing q , the number of active dimensions, was not addressed here however further studies in this direction might lead to a more automated estimation procedure. In the application section we showed that the estimator \hat{p}^{GanMC} made possible the computation of conservative estimates of excursion sets with general Gaussian random field priors. All code was developed in R.

Acknowledgement

The first author acknowledges the support of the Swiss National Science Foundation, grant number 146354.

A Proofs

Proof of Proposition 1

Proof. We have that $\mathbb{E}[\hat{p}_q] = p_q$ and $\mathbb{E}[\widehat{R}_q] = R_q$. Then we have

$$\text{var}(\hat{p}) = \text{var}(\hat{p}_q) + \underbrace{\text{var}((1 - \hat{p}_q)\widehat{R}_q)}_{=\blacksquare} + 2 \underbrace{\text{cov}(\hat{p}_q, (1 - \hat{p}_q)\widehat{R}_q)}_{=\blacktriangle}. \quad (11)$$

We can write the variance \blacksquare and the covariance \blacktriangle as

$$\begin{aligned} \blacksquare &= \text{var}((1 - \hat{p}_q)\widehat{R}_q) = (1 - p_q)^2 \text{var}(\widehat{R}_q) + R_q^2 \text{var}(\hat{p}_q) + \text{var}(\hat{p}_q) \text{var}(\widehat{R}_q), \\ \blacktriangle &= \text{cov}[\hat{p}_q, (1 - \hat{p}_q)\widehat{R}_q] = -\text{var}(\hat{p}_q)R_q, \end{aligned}$$

respectively, by exploiting the independence of \hat{p}_q and \widehat{R}_q . By plugging in those expressions in Equation (11) we obtain the result in Equation (3). \square

Proof of Proposition 2

Proof.

$$\begin{aligned}
\text{var}(\tilde{G}) &= \frac{1}{n^2 m^2} \text{var} \left(\sum_{i=1}^n \sum_{j=1}^m g(W_i, Z_{i,j}) \right) = \frac{1}{nm^2} \text{var} \left(\sum_{j=1}^m g(W_1, Z_{1,j}) \right) \\
&= \frac{1}{nm^2} \sum_{j=1}^m \sum_{j'=1}^m \text{cov} (g(W_1, Z_{1,j}), g(W_1, Z_{1,j'})) \\
&= \frac{1}{nm^2} \left[m \text{var}(g(W_1, Z_{1,1})) + m(m-1) \text{cov}(g(W_1, Z_{1,1}), g(W_1, Z_{1,2})) \right] \\
&= \frac{1}{nm^2} [m \text{var}(g(W_1, Z_{1,1})) + m(m-1)\diamond]. \tag{12}
\end{aligned}$$

where the first equality is a consequence of the independence of W_1, \dots, W_n and the third equality is a consequence of the independence of $Z_{i,j}$ and $Z_{i,j'}$ conditionally on W_i . Moreover the covariance denoted by \diamond in (12) can be written as follows.

$$\begin{aligned}
\diamond &= \underbrace{\mathbb{E}[\text{cov}(g(W_1, Z_{1,1}), g(W_1, Z_{1,2}) \mid W_1)]}_{=0 \text{ } Z_{1,1}, Z_{1,2} \text{ independent conditionally on } W_1} + \underbrace{\text{cov}(\mathbb{E}[g(W_1, Z_{1,1}) \mid W_1], \mathbb{E}[g(W_1, Z_{1,2}) \mid W_1])}_{=\text{var}(\mathbb{E}[g(W_1, Z_{1,1}) \mid W_1])} \\
&= \text{var}(\mathbb{E}[g(W_1, Z_{1,1}) \mid W_1]) = \text{var}(g(W_1, Z_{1,1})) - \mathbb{E} \left[\text{var}(g(W_1, Z_{1,1}) \mid W_1) \right]. \tag{13}
\end{aligned}$$

Equations (12) and (13) give the result (8). \square

Proof of Corollary 1

Proof. Denote with $e = \beta(A-B)$, $f = (\alpha+c)(A-B) + \beta B$, $g = (c+\alpha)B$, $h = C_{\text{tot}}$, then

$$\text{var}(\tilde{G})(m) = \frac{em^2 + fm + g}{hm}. \tag{14}$$

Observe that the first and second derivatives of $\text{var}(\tilde{G})$ with respect to m are respectively

$$\frac{\partial \text{var}(\tilde{G})}{\partial m} = \frac{1}{h} \left[e - \frac{g}{m^2} \right], \quad \frac{\partial^2 \text{var}(\tilde{G})}{\partial m^2} = \frac{2g}{hm^3}.$$

The second derivative is positive for all $m > 0$ then $\text{var}(\tilde{G})$ is a convex function for $m > 0$ and the point of minimum is equal to the zero of $\partial \text{var}(\tilde{G})/\partial m$, which is $m = \sqrt{g/e} = \tilde{m}$.

Since $\text{var}(\tilde{G})$ is convex in m , the integer that realizes the minimal variance is either $\lfloor \tilde{m} \rfloor$ or $\lceil \tilde{m} \rceil$. By plugging in $m = \tilde{m} - \varepsilon = \sqrt{g/e} - \varepsilon$ and $m = \tilde{m} - \varepsilon + 1 = \sqrt{g/e} - \varepsilon + 1$ in Equation (14), we obtain the condition in (9). \square

Proof of Proposition 3

Proof. First of all notice that the total cost of sampling \hat{G} is $C_{\text{tot}} = n(c + C_{Z|W}) = n(c + \alpha + \beta)$. By isolating n in the previous equation we obtain $n = \frac{C_{\text{tot}}}{c + \alpha + \beta}$ and, by computations similar to those in Proposition 2 we obtain

$$\text{var}(\hat{G}) = \frac{c + \alpha + \beta}{C_{\text{tot}}} \text{var}(g(W_1, Z_{1,1})) = \frac{c + \alpha + \beta}{C_{\text{tot}}} A,$$

where $A = \text{var}(g(W_1, Z_{1,1}))$. In the following we will also denote $B = \mathbb{E}[\text{var}(g(W_1, Z_{1,1}) | W_1)]$ as in Corollary 1. Let us now substitute $N_{C_{\text{fix}}}(m^*)$ in equation (8), thus obtaining

$$\begin{aligned} \text{var}(\tilde{G}) &= \frac{(c + \alpha + \beta m^*)Am^* - (m^* - 1)(c + \alpha + \beta m^*)B}{C_{\text{tot}}m^*} \\ &= \text{var}(\hat{G}) \frac{(m^*)^2\beta(A - B) + m^*[(c + \alpha)(A - B) + \beta B] + (c + \alpha)B}{A(c + \alpha + \beta)m^*} \\ &= \text{var}(\hat{G}) \frac{2(\alpha + c)B + m^*[(c + \alpha)(A - B) + \beta B]}{A(c + \alpha + \beta)m^*}, \end{aligned} \quad (15)$$

where in (15) we substituted $(m^*)^2$ from Corollary 1. By further rearranging the terms, we obtain

$$\text{var}(\tilde{G}) = \text{var}(\hat{G}) \left[1 - \frac{(m^* - 2)(c + \alpha)B + m^*\beta(B - A)}{A(c + \alpha + \beta)m^*} \right] = \text{var}(\hat{G}) [1 - \eta].$$

Since $A - B, B, c, \beta, \alpha$ are always positive, then $\eta < 1$ for all $m^* > 0$. Moreover $\eta > 0$ if

$$m^* > \frac{2(\alpha + c)B}{(\alpha + c)B + \beta(A - B)}.$$

□

B Choice of active dimensions

B.1 Basics

The choice of active dimensions E_q plays a crucial role in the approximation of p because it determines the error $\hat{p}_q - p$. Since this error is always negative, we implement procedures to select E_q that exploit this property.

Selecting E_q such that $P(\max X^q > t)$ is numerically maximized, as in Chevalier (2013), optimally reduces the bias of \hat{p}_q as an estimator for p . Here we are not interested in such optimal bias reduction, as we completely remove this error with a second step. However we aim at fast heuristics methods to select E_q in such a way that the error $\hat{p}_q - p$ is lowered.

The basic tool used here to select active dimensions is the excursion probability function:

$$p_t(i) = P(X_i > t) = \Phi \left(\frac{\mu_i - t}{\sqrt{\Sigma_{i,i}}} \right).$$

This function is widely used in spatial statistics (Bolin and Lindgren, 2015) and Bayesian optimization (Kushner, 1964; Bect et al., 2012). In our setting it can be used to identify the dimensions where we have a high probability of exceeding the threshold. The indices that realize a high value for p_t enable identifying dimensions that actively contribute to the maximum. We propose the following methods.

Method 1: sample q indices with probability given by p_t .

Method 2: sample q indices with probability given by $p_t(1 - p_t)$.

These methods require only μ and Σ , and are thus very fast to compute. Both methods were already introduced for sequential evaluations of expensive to evaluate functions, see, e.g., Chevalier et al. (2014b).

Figure 6 shows a comparison of the estimates p_q obtained with different methods to select E_q . The two methods clearly outperform a random choice of active dimensions.

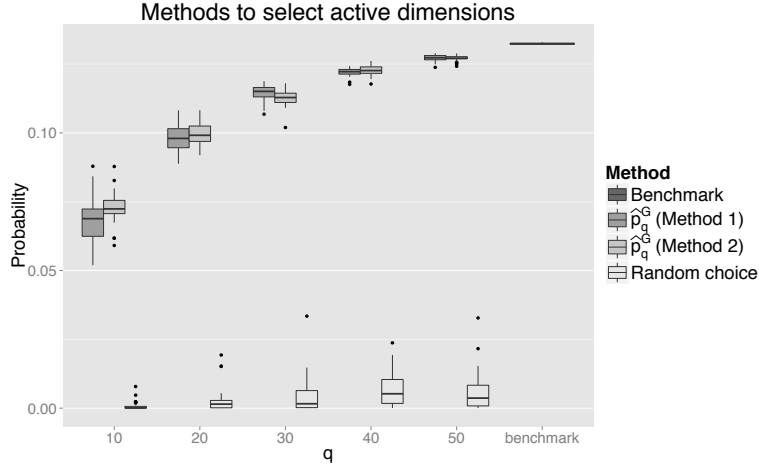


Figure 6: Distribution of \hat{p}_q^G estimates obtained with different choices of active dimensions.

B.2 Choice of active dimensions: spatial methods

In many situations the random vector X comes from a discretization of a Gaussian random field over a set of points $E \subset \mathbb{R}^l$. Let us denote with e_1, \dots, e_d the points in the discretization of E . In this case, we can exploit the spatial information to select the active dimensions with the following two methods.

Method 3: select the first dimension by sampling the index with probability given by p_t . The j th dimension is sampled from the set of indices with probability given by $p_t \delta_j$, where

$$\delta_j = \prod_{k=2}^j \text{dist}(e_k, E \setminus \{e_k\}) \quad (j = 2, \dots, d)$$

with $\text{dist}(e_k, E \setminus \{e_k\})$ denoting the vector of Euclidean distances between e_k and each point in $E \setminus \{e_k\}$.

Method 4: select the first dimension by sampling the index with probability given by $p_t(1 - p_t)$. The j th dimension is sampled from the set of indices with probability given by $p_t(1 - p_t)\delta_j$.

Methods 3 and 4 give better results than Method 1 and 2 when the probability p is not too small because in such situation the function p_t has several local maxima and the spatial correction helps selecting dimensions in each of the modes of p_t .

References

- Abrahamson, I. (1964). Orthant probabilities for the quadrivariate normal distribution. *The Annals of Mathematical Statistics*, 35(4):1685–1703.
- Bayarri, M. J., Berger, J. O., Calder, E. S., Dalbey, K., Lunagomez, S., Patra, A. K., Pitman, E. B., Spiller, E. T., and Wolpert, R. L. (2009). Using statistical and computer models to quantify volcanic hazards. *Technometrics*, 51(4):402–413.

- Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E. (2012). Sequential design of computer experiments for the estimation of a probability of failure. *Stat. Comput.*, 22 (3):773–793.
- Bolin, D. and Lindgren, F. (2015). Excursion and contour uncertainty regions for latent Gaussian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):85–106.
- Bratley, P. and Fox, B. L. (1988). Algorithm 659: Implementing Sobol’s quasirandom sequence generator. *ACM Trans. Math. Software*, 14(1):88–100.
- Chevalier, C. (2013). *Fast uncertainty reduction strategies relying on Gaussian process models*. PhD thesis, University of Bern.
- Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., and Richet, Y. (2014a). Fast kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics*, 56(4):455–465.
- Chevalier, C., Ginsbourger, D., Bect, J., and Molchanov, I. (2013). Estimating and quantifying uncertainties on level sets using the Vorob’ev expectation and deviation with Gaussian process models. In Uciński, D., Atkinson, A., and Patan, C., editors, *mODa 10 Advances in Model-Oriented Design and Analysis*. Physica-Verlag HD.
- Chevalier, C., Picheny, V., and Ginsbourger, D. (2014b). The KrigInv package: An efficient and user-friendly R implementation of kriging-based inversion algorithms. *Comput. Statist. Data Anal.*, 71:1021–1034.
- Cox, D. R. and Wermuth, N. (1991). A simple approximation for bivariate and trivariate normal integrals. *International Statistical Review*, 59(2):263–269.
- Dickmann, F. and Schweizer, N. (2014). Faster comparison of stopping times by nested conditional Monte Carlo. *arXiv preprint arXiv:1402.0243*.
- French, J. P. and Sain, S. R. (2013). Spatio-temporal exceedance locations and confidence regions. *Ann. Appl. Stat.*, 7(3):1421–1449.
- Genz, A. (1992). Numerical computation of multivariate normal probabilities. *J. Comput. Graph. Statist.*, 1(2):141–149.
- Genz, A. and Bretz, F. (2002). Comparison of methods for the computation of multivariate t probabilities. *Journal of Computational and Graphical Statistics*, 11(4):950–971.
- Genz, A. and Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities*. Lecture Notes in Statistics 195. Springer-Verlag.
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., Bornkamp, B., and Hothorn, T. (2012). *mvtnorm: Multivariate Normal and t Distributions*. R package version 0.9-9992.
- Hajivassiliou, V., McFadden, D., and Ruud, P. (1996). Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results. *Journal of econometrics*, 72(1):85–134.

- Hammersley, J. (1956). Conditional Monte Carlo. *Journal of the ACM (JACM)*, 3(2):73–76.
- Hammersley, J. and Morton, K. (1956). A new monte carlo technique: antithetic variates. In *Mathematical proceedings of the Cambridge philosophical society*, volume 52, pages 449–475. Cambridge Univ Press.
- Horrace, W. C. (2005). Some results on the multivariate truncated normal distribution. *Journal of Multivariate Analysis*, 94(1):209–221.
- Kahn, H. (1950). Random sampling (Monte Carlo) techniques in neutron attenuation problems–I. *Nucleonics*, 6(5):27–33.
- Kahn, H. and Marshall, A. W. (1953). Methods of reducing sample size in monte carlo computations. *Journal of the Operations Research Society of America*, 1(5):263–278.
- Kushner, H. J. (1964). A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *Journal of Basic Engineering*, 86(1):97–106.
- Lemieux, C. (2009). *Monte Carlo and quasi-Monte Carlo sampling*. Springer.
- Miwa, T., Hayter, A., and Kuriki, S. (2003). The evaluation of general non-centred orthant probabilities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1):223–234.
- Molchanov, I. (2005). *Theory of Random Sets*. Springer, London.
- Moran, P. (1984). The monte carlo evaluation of orthant probabilities for multivariate normal distributions. *Australian Journal of Statistics*, 26(1):39–44.
- Owen, D. B. (1956). Tables for computing bivariate normal probabilities. *The Annals of Mathematical Statistics*, 27(4):1075–1090.
- Picheny, V., Ginsbourger, D., Richet, Y., and Caplin, G. (2013). Quantile-based optimization of noisy computer experiments with tunable precision. *Technometrics*, 55(1):2–13.
- Rasmussen, C. E. and Williams, C. K. (2006). *Gaussian processes for machine learning*. MIT Press.
- Ridgway, J. (2014). Computation of gaussian orthant probabilities in high dimension. *arXiv preprint arXiv:1411.1314*.
- Robert, C. and Casella, G. (2013). *Monte Carlo statistical methods*. Springer.
- Robert, C. P. (1995). Simulation of truncated normal variables. *Statistics and computing*, 5(2):121–125.
- Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statist. Sci.*, 4(4):409–435.

Schervish, M. J. (1984). Algorithm AS 195: Multivariate normal probabilities with error bound. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 33(1):81–94.

Tong, Y. L. (2012). *The multivariate normal distribution*. Springer.