EFFICIENT SIMULATION FOR TAIL PROBABILITIES OF GAUSSIAN RANDOM FIELDS

Robert J. Adler

Faculty of Industrial Engineering and Management
and Faculty of Electrical Engineering
Technion-Israel Institute of Technology
Haifa, 32000, ISRAEL

Jose Blanchet

Department of Industrial Engineering
and Operations Research
Columbia University
New York, NY 10027, U.S.A.

Jingchen Liu

Department of Statistics
Columbia University
New York, NY 10027, U.S.A.

ABSTRACT

We are interested in computing tail probabilities for the maxima of Gaussian random fields. In this paper, we discuss two special cases: random fields defined over a finite number of distinct point and fields with finite Karhunen-Loève expansions. For the first case we propose an importance sampling estimator which yields asymptotically zero relative error. Moreover, it yields a procedure for sampling the field conditional on it having an excursion above a high level with a complexity that is uniformly bounded as the level increases. In the second case we propose an estimator which is asymptotically optimal. These results serve as a first step analysis of rare-event simulation for Gaussian random fields.

1 INTRODUCTION

Consider a Gaussian random field, $f(t)$, living in a compact subset $T \subset R^d$. Without loss of generality, we assume that $Ef(t) = 0$, and denote the covariance function by

$$C(s,t) = E(f(s)f(t)),$$

$s,t \in T$. We assume that $f(t)$ is almost surely continuously twice differentiable, sufficient conditions for which can be found, for example, in (Adler and Taylor 2007, Adler, Taylor, and Worsley 2009). Consequently, diffusions and other stochastic processes with rough paths are not included in the current discussion. Despite this, not all the results in the current paper require differentiability. Nevertheless, we assume it in preparation for later work where it will be

needed. We are interested in computing

$$P\left(\sup_{t\in T} f(t) > b\right), \qquad (1)$$

as $b \to \infty$. The strategy we adopt is based on importance sampling. Besides computing the probability in (1), the sampler provides weighted samples with which one can efficiently compute conditional expectations of the form

$$E\left(\Gamma(f(\cdot)) | \sup_{t\in T} f(t) > b\right)$$

where $\Gamma(\cdot)$ is some bounded functional mapping from $C^2(T)$ to R. This is one of the advantages of importance sampling over the classic asymptotic approximations of (1).

Tail probabilities of Gaussian and related random fields are of importance both in the theory of probability and statistics and in a wide range of applications. Among the areas of application are the following:

- Physical oceanography and hydrology, where, as in (Adler, Muller, and Rozovskii 1996, Rubin 2002), the spatial parameter set T might be the two dimensional water surface or a three dimensional body of water, and the variable being measured might be water temperature and/or pressure.
- Atmospheric studies, where the random field might provide a model for wind speads or airborne contaminants, with a theory as described in (Daley 1991).
- Geostatistics and other earth sciences, where, as for oceanography, the spatial aspects of parame-

ter spaces may be two or three dimensional. cf. (Christakos 1991, Christakos 2000, Olea 1999).

- Astrophysics, where random field techniques have been heavily used in analyzing the COBE (Cosmic Background Explorer) data, which measures the 'signature radiation' from the universe of 15 billion years ago (Smoot and Davidson 1993, Torres 1994, Vogeley, Park, Geller, Huchin, and Gott 1994). This is directional data, and so is realised as a random field on the two-dimensional sphere. Three dimensional astrophysical data has been generated by the Sloan digital sky survey (Gott, Hambrick, Vogeley, Kim, Park, Choi, Cen, Ostriker, and Nagamine 2008).

- Analysis of functional magnetic resonance imaging (fMRI) data as in (Evans, Marret, Neelin, Collins, Worsley, Dai, Milot, Meyer, , and Bub 1992, Friston, Holmes, Worsley, Polin, Frith, and Frackowiak 1995, Shafie, Sigal, Siegmund, and Worsley 2003, Taylor and Worsley 2007). Here random field techniques have been heavily used in the thresholding problems, which, because of their importance and wide usage in a number of areas we take a moment to describe.

Signal detection and thresholding Consider the classical signal+noise paradigm, which considers the model

$$f(t) = s(t) + \eta(t). \tag{2}$$

The signal s is deterministic, and, purely for reasons of exposition, is assumed to take primarily positive values if, in fact, it is present. The noise η is a mean zero random field of background noise, the statistical properties of which are assumed to be known. The first problem here is determining whether or not the signal is at all present, and, in view of the assumed positivity of s, a useful statistic for testing this is the threshold statistic

$$\sup_{t \in T} f(t).$$

(In fact, the supremum is the maximum likelihood statistics for this test, if η is Gaussian white noise smoothed with a filter with shape matching that of the signal. This is the 'matched filter theorem' of signal processing.)

In order to use this statistic, or, indeed, to perform any statistical thresholding technique, one needs to know how to compute, at least for large values of u, the *excursion probabilities* (1) when no signal is present. Hence our interest in them. (By 'large' we mean large enough for this probability to be less that about 0.10, the level at which one usually begins to talk about statistical significance.)

Computing Gaussian excursion probabilities is a hard problem, and except for a handful of special cases there are no known precise formulae for them. While there do exist a number of good approximations, some of which we will meet below, these are not always easy to apply, and so in this paper we approach the problem from the viewpoint of simulation. A basic problem is that random fields are, in general, represented by an infinite number of random variables, and so one of the challenges is how to represent them in a computer, which can only represent finite dimensional object. We will present an approach in this paper that overcomes this problem, and will then discuss two finite dimensional special cases: random fields defined on finite space T, and random fields with finite Karhunen-Loève expansions. We will provide one estimator for each of the two cases which can then be rigorously proven to be efficient (precise definitions are given in Section 2). These two cases will be further generalized and extended in the journal version of this paper.

Within this framework, the contributions of this paper are as follows. Firstly, for the finite dimension field problem, we develop an importance sampling algorithm that can be shown to have asymptotically zero coefficient of variation as the probability that the maximum of the coordinates is large tends to zero. This estimator also provides an algorithm to sample from the conditional distribution of a multivariate Gaussian distribution conditional on the maximum of the coordinates being greater than some level via von Neumann's acceptance-rejection sampling, and the expected acceptance probability can be shown to converge to one as the level, b, tends to infinity. Secondly, for Gaussian random fields with finite Karhunen-Loève decomposition, we present an estimator which is asymptotically optimal (a property that is also refered to as weak efficiency).

The remainder of this paper is organized as follows. In Section 2 we discuss the efficiency and optimality of rare event simulation. In Section 3 we present several important large deviations results for Gaussian random fields. In Section 4 we present the importance sampling algorithm for the finite dimension field problem along with an analysis of its efficiency. In Section 5 we discuss random fields with finite Karhunen-Loève decomposition. Section 6 presents two examples for which we implement our algorithms.

2   IMPORTANCE SAMPLING AND EFFICIENT RARE-EVENT SIMULATION

Suppose we are interested in estimating $P(Z > b)$, where Z is a random variable taking values in R. Let $F(\cdot)$ be the distribution function of Z. Let $G(\cdot)$ be another distribution function and assume that the likelihood ratio $L(z) = I_{[b, +\infty)}(z) \frac{dF}{dG}(z)$ is well defined. Then,

$$E^G L(Z) = P(Z > b),$$

where $E^G(\cdot)$ is the expectation under $G(\cdot)$. If one chooses

$$G(dz) = P(Z \in dz | Z > b) = \frac{P(Z \in dz, I(Z > b))}{P(Z > b)},$$

then $L(z) = P(Z > b)$. This suggests that if $G(\cdot)$ is chosen to be the conditional distribution of Z given $Z > b$, the estimator will have zero variance. Such a choice is, of course, not implementable, since it requires knowledge of $P(Z > b)$, which is precisely the probability that we do not know how to compute. Nevertheless, the argument suggests that an efficient importance sampling algorithm should have a sampling distribution very close to the conditional distribution of the original law given that the excursion event occurs. Efficient importance sampling algorithms for rare-event simulation which use this idea is well known, see for instance, (Asmussen and Glynn 2007), (Bucklew 2004), (Giordano, Gubinelli, and Pagano 2005), and (Juneja and Shahabuddin 2006). For studies particularly related to multivariate Gaussian distributions, see (Dieker and Mandjes 2006) and (Glasserman 2004).

There are many criteria to evaluate the efficiency of an estimator. A well accepted one in the rare-event simulation literature is to assess efficiency via second moments. Following this line, we define two types of efficiency criteria for estimators of tail probabilities.

Definition 1   *An estimator, L_b, is said to be strongly efficient (for the tail of the random variable Z) if $EL_b = P(Z > b)$, and*

$$\sup_b \frac{VarL_b}{P^2(Z > b)} < \infty.$$

Definition 2   *An estimator, L_b, is said to be asymptotically optimal or weakly efficient (for the tail of the random variable Z) if $EL_b = P(Z > b)$, and*

$$\frac{\log EL_b^2}{2\log P(Z > b)} \to 1,$$

as $b \to \infty$.

It is easy to see that strong efficiency implies weak efficiency.

Suppose we want to simulate m i.i.d. random variables $L_b^1, ..., L_b^m$ (equally distributed with L_b) and use their average

$$\hat{P}_b^m = \frac{1}{m} \sum_{i=1}^m L_b^i$$

as an estimator for $P(Z > b)$. We want to estimate $P(Z > b)$ with relative accuracy $\varepsilon - \delta$, by which we mean

$$P\left(\left| \frac{\hat{P}_b^m - P(Z > b)}{P(Z > b)} \right| > \varepsilon \right) < \delta. \qquad (3)$$

If L_b is strongly efficient, then Chebyshev's inequality shows that, in order to achieve (3), the required number of replications, m, is bounded as $b \to \infty$. If L_b is weakly efficient, one needs $m = \exp(-o(\log P(Z > b)))$. Notice that to achieve the same relative accuracy, crude Monte Carlo needs $m = O\left(P(Z > b)^{-1}\right)$. Therefore, both strongly efficient and asymptotically optimal estimators are noticeably more efficient than crude Monte Carlo. Obviously, however, these efficiency concepts are not enough to evaluate the overall complexity of a computational procedure because one must also take into account the cost required to generate a single replication of the estimator L_b. Nevertheless, relative to single crude Monte Carlo replications, the computational overhead associated to generating samples under the importance sampling distributions that we consider is not significant and therefore we do not explicitly address it in our discussion here.

3   EXCURSION PROBABILITIES OF GAUSSIAN RANDOM FIELDS

As mentioned earlier, excursion probabilities of Gaussian random field have been studied in depth. In this section we shall recall a few relevant results from the area, starting with a version of a basic result due, independently, to (Borell 1975) and (Tsirelson, Ibragimov, and Sudakov 1976).

Theorem 3 (Borell-TIS)   *Let $f(\cdot)$ be a real-valued, separable, continuous Gaussian process. Suppose that*

$$\sigma^2(T) = \sup_T Var(f(t)) < \infty.$$

Let $m = \sup_T Ef(t)$ and choose any a for which

$$P\left(\sup_T (f(t) - m) \geq a \right) \leq \frac{1}{2}.$$

Then, for all b,

$$P\left(\sup_T f(t) > b \right) \leq 2\left(1 - \Phi\left(\frac{b - m - a}{\sigma(T)} \right) \right), \qquad (4)$$

where $\Phi(\cdot)$ is the c.d.f. of the standard Gaussian distribution.

This theorem provides an upper bound to the tail of the maximum of a random field and covers a large class of Gaussian random fields including nondifferentiable ones. However, as one would expect of a universal bound, it is generally too crude to produce useful numerical bounds.

Note that a trivial lower bound is given by

$$P\left(\sup_T f(t) > b \right) \geq 1 - \Phi(b / \sigma(T)). \qquad (5)$$

Together, the two bounds imply that the decay rate of (1) is roughly $\exp\left(-\frac{b^2}{2\sigma^2(T)}\right)$. However, there is quite a gap between (4) and (5), and generally neither of them gives the right exact asymptotics for smooth fields. To obtain sharper approximations, we need to make further assumption on the field.

For example, (Adler 1981), (Taylor and Adler 2003) and (Adler and Taylor 2007) have studied the Euler characteristic, which we denote by χ, of excursion sets, $f^{-1}([b,+\infty))$, where $f^{-1}(A) = \{t \in T : f(t) \in A\}$, for twice differentiable fields with constant variance σ and satisfying other technical conditions. Putting together a precise analytic formula for the expected value of the the Euler characteristic, and the fact that

$$E\left(\chi\left(f^{-1}([b,+\infty))\right)\right) = (1+o(1))P\left(\sup_T f(t) > b\right),$$

as $b \to \infty$, one can obtain the sharp asymptotic approximation

$$P\left(\sup_T f(t) > b\right) = (1+o(1))C(T)b^{d-1}\exp\left(-\frac{b^2}{2\sigma^2}\right),$$

(6)

where $C(T)$ is a constant only depending on the geometry of T and d is its dimension. This approximation is roughly $b^d b^{-1}\exp\left(-b^2/2\sigma^2\right)$, where $b^{-1}\exp\left(-b^2/2\sigma^2\right)$ is the decay rate of the Gaussian tail probability.

Interestingly, (6) implies that the maximum of a smooth homogeneous Gaussian random field is somehow related to b^d Gaussian random variables which are no more than weakly correlated. It further implies that if we want to discretize the space T to approximate $P(\sup f > b)$, then we might do well by partitioning T into b^d cubes of volume $O\left(b^{-d}\right)$, and side length $O\left(b^{-1}\right)$. The number of cubes grows polynomially with b, depending on the dimension of T. We shall use these facts below, partitioning T into small regions of appropriate size and then using the exclusion and inclusion formula to obtain approximations for exceedence probabilities. This is not a new technique, introduced by (Pickands 1969) and used heavily by (Piterbarg 1995).) under the name "double sum method" by (Piterbarg 1995), in wide generality. In the following Section 4 we shall apply it to finite dimensional multivariate Gaussian distributions as b tends to infinity.

4 THE MAXIMUM OF GAUSSIAN VECTORS

In this section, we consider the problem of computing the tail probability of the maximum of a d dimensional Gaussian vector. More precisely, let $X = (X_1,...,X_d) \sim N(0,\Sigma)$, where Σ is a $d \times d$ positive definite matrix. Equivalently, we can consider X as the realization of a random field $f(t)$, where t ranges over the finite set $T = \{t_1,...,t_d\}$. We want to
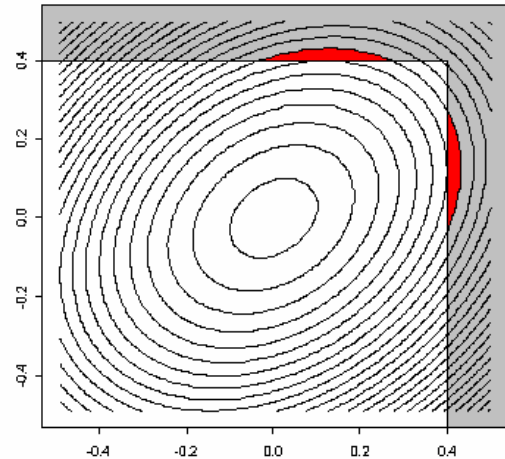

Figure 1: Two Dimensional Example

compute

$$P\left(\max_j X_j > b\right)$$

(7)

as $b \to \infty$.

As noted above, the motivation behind this example lies in approximating a field defined over a general, continuous, parameter space by its values on a finite subset and then using (7), with $X_j = f(t_j)$, to approximate (1). We will not address the important issue of the best choice of the t_j here.

We start with some notation, for which we will use upper case letters for random variables and lower case for non-random quantities. Let $f_j(x_j)$ be the marginal distribution of X_j evaluated at $x_j \in R$. Let x_{-j} be the vector $(x_1,...,x_{j-1},x_{j+1},...,x_d)$. Analogous notation is used for the corresponding random variables. Let $f(x_{-j}|x_j)$ be the conditional density of X_{-j} evaluated at x_{-j} given $X_j = x_j$. The density of X evaluated at x is denoted by $f(x)$. Note that $f(x) = f_j(x_j)f(x_{-j}|x_j)$. Consider now the density function,

$$g(x) = \sum_{j=1}^d p_j(b)\frac{f_j(x_j)f(x_{-j}|x_j)}{P(X_j > b)}I(x_j > b),$$

(8)

where $p_j(b) = P(X_j > b)/\sum_{i=1}^d P(X_i > b)$.

One can think of this density as a mixture. The idea is that samples drawn under g will have a maximum above b, and the event $\{\max_j X_j > b\}$ is most likely to occur by only one of the X_j being large. The mixture probabilities $p_j(b)$, represent the event that is X_j which exceeds the threshold b. A graphical illustration of $d = 2$ is presented in Figure 1,

where the red (dark) regions contain most of the probability of the (grey) region where $\max\{X_1, X_2\} > b$.

The choice of g as an importance sampling density allows us to obtain an estimator with asymptotically zero coefficient of variation as $b \to \infty$. To see this, we compute the likelihood ratio given a realization $X = x$ sampled from $g(\cdot)$ (which includes the event $\max_j X_j > b$ with probability 1) as

$$L = \frac{f(x)}{g(x)} = \frac{\sum_{i=1}^d P(X_i > b)}{\sum_{j=1}^d I(x_j > b)} \le \sum_{i=1}^d P(X_i > b). \quad (9)$$

Therefore,

$$E^g L^2 \le \left(\sum_{i=1}^d P(X_i > b) \right)^2.$$

where $E^g(\cdot)$ is expectation under $g(\cdot)$. Since we can claim an asymptotically zero coefficient of variation if

$$\frac{Var^g(L)}{P(\max_j X_j > b)^2} = \frac{E^g(L^2) - P(\max_j X_j > b)^2}{P(\max_j X_j > b)^2} \to 0,$$

it suffices to show that

$$\frac{\sum_{i=1}^d P(X_i > b)}{P(\max_j X_j > b)} \to 1, \quad (10)$$

as $b \to \infty$. Without loss of generality, assume $X_1, ..., X_{\tilde{d}}$ have the largest variance, where $\tilde{d} \le d$. This implies that $X_1, ..., X_{\tilde{d}}$ are identically distributed and

$$\frac{\sum_{i=1}^{\tilde{d}} P(X_i > b)}{\sum_{i=1}^d P(X_i > b)} \to 1,$$

as $b \to \infty$. By the inclusion-exclusion formula,

$$\sum_{i=1}^{\tilde{d}} P(X_i > b) - \sum_{i \ne j, i, j \le \tilde{d}} P(X_i > b, X_j > b) \quad (11)$$

$$\le P\left(\max_{j \le \tilde{d}} X_j > b \right) \le \sum_{i=1}^d P(X_i > b).$$

Note that for $i \ne j$ and $i, j \le \tilde{d}$,

$$P(X_i > b, X_j > b) = P(X_i > b) P(X_j > b | X_i > b).$$

Now note that as long as the correlation between X_i and X_j is less than 1,

$$P(X_j > b | X_i > b) \to 0,$$

as $b \to \infty$. Therefore,

$$\sum_{i \ne j, i, j \le \tilde{d}} P(X_i > b, X_j > b) = o\left(\sum_{i=1}^{\tilde{d}} P(X_i > b) \right), \quad (12)$$

and

$$\frac{\sum_{i=1}^{\tilde{d}} P(X_i > b)}{P\left(\max_{j \le \tilde{d}} X_j > b \right)} \to 1.$$

Furthermore,

$$1 \ge \frac{P(\max_j X_j > b)}{\sum_{i=1}^d P(X_i > b)} \ge$$

$$\frac{P\left(\max_{j \le \tilde{d}} X_j > b \right)}{\sum_{i=1}^{\tilde{d}} P(X_i > b)} \frac{\sum_{i=1}^{\tilde{d}} P(X_i > b)}{\sum_{i=1}^d P(X_i > b)} \to 1$$

and (10) holds. We summarize these facts as follows.

Theorem 4 *Let X be a d dimensional random vector with the multivariate Gaussian distribution $N(0, \Sigma)$, where Σ is a positive definite matrix. The sampling distribution in (8) provides asymptotically zero relative error for the exceedence probability (7). That is, the corresponding estimator L in (9) satisfies*

$$\frac{Var(L)}{P(\max_j X_j > b)^2} \to 0,$$

as $b \to \infty$. Hence, L is strongly efficient.

Another associated result that comes from the preceeding argument is that one can efficiently simulate the conditional distribution of $(X_1, ..., X_d)$ given $\max_j X_j > b$. It follows from (9) that $g(x) \sum_{i=1}^d P(X_i > b)$ is an upper bound of $f(x)$. By employing von Neumann's acceptance-rejection sampling by simulating X according to the density $g(x)$, then accept X with probability

$$p(x) = \frac{f(x)}{g(x) \sum_{i=1}^d P(X_i > b)}.$$

The expected acceptance probability

$$E^g p(X) = \frac{P(\max_j X_j > b)}{\sum_{i=1}^d P(X_i > b)} \to 1.$$

By this procedure, we can sample from the conditional distribution of X given that $\max_j X_j > b$, asymptotically using basically the same number of computer operations as generating X unconditionally.

Before leaving this example we note that Theorem 4 has limitations when it comes to applying it to random fields. For a start, it assumes that the dimension of X is fixed,

and gives an asymptotic result as $b \to \infty$. For homogeneous random fields, in order to have $\varepsilon - \delta$ accuracy in (3), or ε-accuracy,

$$\left| \frac{P\left(\max_{1 \leq i \leq d} f\left(t_i\right) > b\right)}{P\left(\sup_{t \in T} f\left(t\right) > b\right)} - 1 \right| \leq \varepsilon, \qquad (13)$$

one needs to let the dimension d also tend to infinity, perhaps a rate $d = b^{\alpha}$, for some α depending on the smoothness of the random field or, equivalently, the behavior of its correlation function at the origin. See (Piterbarg 1995). This implies the correlation between some of the X_i and X_j will tend to 1 and (12) may not be true. Nevertheless, the current analysis serves as the first step on the rare-event simulation for Gaussian random fields. Generalizing the results in this section to random fields is one of the directions of our future studies.

5 GAUSSIAN PROCESSES WITH FINITE KARHUNEN-LOÈVE EXPANSIONS

In this section, we consider a Gaussian random field having the form

$$f_Z\left(t\right) = \sum_{i=1}^{n} Z_i \sigma_i\left(t\right) \qquad (14)$$

where the Z_i are i.i.d. $N(0,1)$, and the $\sigma_i\left(\cdot\right)$ are known, deterministic functions. For technical convenience, we further assume $\sigma_i\left(\cdot\right) \in C^2\left(T\right)$, which will also assure $f_Z\left(\cdot\right) \in C^2\left(T\right)$.

We can think of (14) as a truncated version of an infinite Karhunen-Loève expansion, although we do not need to require that the $\sigma_i\left(\cdot\right)$ to be orthogonal in any sense. In that case, to achieve ε-accuracy of the exceedence probability, the choice of n will depend on the decay rates of the $\sigma_i\left(t\right)$. For example, if $\sigma_i\left(t\right)$ decays exponentially as $i \to \infty$, we need to choose $n = O\left(\log b\right)$.

With the representation in (14), to simulate $f_Z\left(t\right)$, we only need to generate i.i.d. Gaussian r.v.'s. In this section, we provide an algorithm that is very easy to implement and asymptotically optimal. Based on the fact that

$$\left(\sum_{i=1}^{n} Z_i \sigma_i\left(t\right)\right)^2 \leq \sum_{i=1}^{n} Z_i^2 \sum_{i=1}^{n} \sigma_i^2\left(t\right),$$

$\sup_t f_Z(t) > b$ implies $\sum_{i=1}^{n} Z_i^2 \geq b^2 / \sigma^2\left(T\right)$. Therefore, the rare event $\sup f > b$ is mainly caused by $\sum_{i=1}^{n} Z_i^2$ being large. Consider the a sampling distribution for $Z_1,...,Z_n$ with the density

$$q_{\theta}\left(z_1,...,z_n\right) = \left(\frac{1}{\sqrt{2\pi}\theta}\right)^n \exp\left(-\frac{1}{2\theta^2}\sum z_i^2\right), \qquad (15)$$

which forms a natural exponential family. In comparison with the underlying standard Gaussian distribution, this sampling distribution inflates all the Z_i by a common factor θ. Now write the $(Z_1,...,Z_n)$ in polar coordinates, (ρ, r), where $\rho = \sqrt{\sum_{i=1}^{n} Z_i^2}$ and $r = \frac{1}{\rho}\left(Z_1,...,Z_n\right)$. If Z_i's are i.i.d standard Gaussian, then it is easy to see that ρ and r are independent. Thus one can think of (15) as corresponding to exponential tilting on ρ^2. The corresponding likelihood is

$$L_b = I\left(\|f_z\| > b\right)\theta^d \exp\left(-\frac{1}{2}\left(1 - \frac{1}{\theta^2}\right)\sum z_i^2\right), \qquad (16)$$

from which it follows that

$$E^{\theta}\left(L_b^2\right) =$$
$$\int_{\{z:\|f_z\|>b\}} \left(\frac{\theta}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}\left(2 - \frac{1}{\theta^2}\right)\sum z_i^2\right) dz_1 \ldots dz_n,$$

where E^{θ} is expectation under q_{θ}. Since $\{z : \sup f_z > b\} \subset \{z : \sigma^2\left(T\right)\sum_{i=1}^{n} z_i^2 \geq b^2\}$, it follows that

$$
\begin{aligned}
& E^{\theta}\left(L_b^2\right) \\
\leq \quad & \left(\frac{\theta}{\sqrt{2\pi}}\right)^n \\
& \int_{\left\{z:\sum_{i=1}^{n} z_i^2 \geq \frac{b^2}{\sigma^2(T)}\right\}} \exp\left(-\left(1 - \frac{1}{2\theta^2}\right)\sum z_i^2\right) dz_1 ... dz_n \\
= \quad & \left(\frac{\theta}{\sqrt{2 - \frac{1}{\theta^2}}}\right)^n P\left(\chi_n^2 \geq \frac{b^2}{\sigma^2\left(T\right)}\left(2 - \frac{1}{\theta^2}\right)\right).
\end{aligned}
$$

Applying a (sharp) asymptotic approximation of the tail probability of a χ_n^2 variable, we have

$$
\begin{aligned}
E^{\theta}\left(L_b^2\right) \quad \leq \quad & (1+o(1))\left(\frac{\theta}{\sqrt{2 - \frac{1}{\theta^2}}}\right)^n \frac{2^{-n/2-1}}{\Gamma\left(n/2\right)} \\
& \left[\frac{b^2}{\sigma^2\left(T\right)}\left(2 - \frac{1}{\theta^2}\right)\right]^{n/2-1} \\
& \exp\left(-\frac{b^2}{\sigma^2\left(T\right)}\left(1 - \frac{1}{2\theta^2}\right)\right) \\
= \quad & (1+o(1))\frac{2^{-n/2-1}}{\Gamma\left(n/2\right)}\frac{\theta^n}{2 - \frac{1}{\theta^2}} \\
& \frac{b^{n-2}}{\sigma^{n-2}\left(T\right)}\exp\left(-\frac{b^2}{\sigma^2\left(T\right)}\left(1 - \frac{1}{2\theta^2}\right)\right).
\end{aligned}
$$

If we choose $\theta = b$, we find

$$
\begin{aligned}
E^{\theta}\left(L_b^2\right) &\leq (1+o(1))C(n)\frac{b^{2n-2}}{2-\frac{1}{b^2}}\exp\left(-\frac{b^2}{\sigma^2(T)}\right) \\
&\triangleq h(b),
\end{aligned}
\tag{17}
$$

where $C(n) = \frac{2^{-n/2-1}\sqrt{e}}{\Gamma(n/2)\sigma^{n-2}(T)}$. This leads to the following theorem.

Theorem 5 *Take f as defined by (14), and choose the sampling distribution q_θ as defined by (15) with $\theta = b$, giving the likelihood ratio L_b of (16). Then the corresponding importance sampling estimate for the exceedence probability (1) is asymptotically efficient.*

Proof. A lower bound of the excursion probability can be easily obtained, that is,

$$
P\left(\sup_T f(t) > b\right) \geq 1 - \Phi\left(\frac{b}{\sigma(T)}\right) \sim \frac{1}{b}\exp\left(-\frac{b^2}{\sigma^2(T)}\right),
\tag{18}
$$

Note that $E^{\theta}L_b^2 \geq \left(E^{\theta}L_b\right)^2$. To show that

$$
\frac{\log E^{\theta}L_b^2}{2\log E^{\theta}(L_b)} \to 1,
$$

as $b \to \infty$, it is sufficient to show that

$$
\begin{aligned}
\frac{\log E^{\theta}L_b^2}{2\log E^{\theta}(L_b)} &\geq \frac{\log E^{\theta}L_b^2}{2\log\left[1-\Phi\left(\frac{b}{\sigma(T)}\right)\right]} \\
&\geq \frac{\log h(b)}{2\log\left[1-\Phi\left(\frac{b}{\sigma(T)}\right)\right]} \to 1.
\end{aligned}
$$

The above inequalities and limit are immediate from (17) and (18). ∎

We close this section by noting that there are many ways to improve the efficiency of (15). For example, instead of sampling from a Gaussian distribution, it is more efficient to simulate from a truncated Gaussian with restriction that $\sum_{i=1}^{n}Z_i^2 \geq b^2/\sigma^2(T)$. This will reduce the variance of L by a constant factor (asymptotically free of b). Another more important improvement lies in twisting the distribution of r. Instead of taking r uniformly distributed on the unit sphere, it may be more efficient to let r be concentrated on smaller regions. To apply this idea effectively requires additional knowledge of the behavior of the σ_i. For example, consider the two cases in which $\sigma^2(t) \triangleq \sum_{i=1}^{n}\sigma_i^2(t)$ is either constant or has finite many local modes in T. If σ^2 is constant, then taking r to be uniformly distributed on the unit sphere is efficient. However, if σ^2 has finitely many local modes, let r be more concentrated in a few regions, the position of which is determined by these modes and the extent of which is determined by the behavior of σ^2 in their immediate

Table 1: Simulation results for finite fields. All results are based on 1000 independent simulations.

b	Est.	Std. Er.
3	6.01E-3	4.70E-5
4	1.53E-4	6.55E-7
5	1.43e-06	2.26E-9

neighborhoods. This improvement will be included in the journal version of this paper by employing state-dependent importance sampling and Lyapunov function technique.

6 SIMULATION EXERCISE

In this section, we provide one example for each of the two cases discussed in the preceding two sections.

Example 6 *We compute $P(\max_i X_i > b)$, where $X = (X_1,...,X_5)$ is multivariate Gaussian with mean zero, marginal variance one. They share a common correlation 0.5. Table 1 shows the simulation results.*

As an important aside, we note that during the implementation, it was clear that the estimator of $E^g L^2$ is not as efficient. As we noted earlier,

$$
L = \frac{\sum_{i=1}^{d}P(X_i > b)}{\sum_{i=1}^{d}I(X_i > b)}.
$$

Thus, in order to observe variation in L, we need to observe at least two of the X_i being greater than b. This event is of very small probability even after the change of measure.

Example 7 *Let $f(t) = X\cos t + Y\sin t$ and $T = [0, 3/4]$ where X and Y are i.i.d. standard Gaussian r.v.. We compute $P(\|f\| > b)$ with the algorithm in Section 5. This probability is known in closed form (cf. (Adler and Taylor 2007)) and is given by*

$$
P(\|f\| > b) = 1 - \Phi(b) + \frac{3}{8\pi}e^{-b^2/2}.
\tag{19}
$$

Table 2 shows the (remarkably accurate) simulation results.

REFERENCES

Adler, R. 1981. *The geometry of random fields*. Chichester, U.K.; New York, U.S.A.: Wiley.

Adler, R., and J. Taylor. 2007. *Random fields and geometry*. Springer.

Adler, R., J. Taylor, and K. Worsley. 2009. *Applications of random fields and geometry, foundations and case studies*. In preparation. Relevant chapters available at http://iew3.technion.ac.il/~radler/publications.html.

Table 2: Simulation results for cosine process. All results are based on 10^5 independent simulations. The "True Value" is computed using (19)

b	True Value	Est.	Std. Er.
1.5	0.1055	0.1057	7.96E-04
2.0	0.0389	0.0392	4.02E-04
2.5	0.0115	0.0118	1.57E-04
3	2.68E-03	2.60E-03	4.31E-05
5	7.31E-07	7.56E-07	2.33E-08
10	3.06E-23	3.15E-23	2.17E-24

Adler, R. J., P. Muller, and B. Rozovskii. (Eds.) 1996. *Stochastic modelling in physical oceanography*. Boston: Birkhaüser.

Asmussen, S., and P. Glynn. 2007. *Stochastic simulation: Algorithms and analysis*. New York, NY, USA: Springer.

Borell, C. 1975. The Brunn-Minkowski inequality in Gauss space. *Inventiones Mathematicae*.

Bucklew, J. 2004. *Introduction to rare event simulation*. New York, NY, USA: Springer.

Christakos, G. 1991. *Random field models in earth sciences*. Academic.

Christakos, G. 2000. *Modern spatiotemporal geostatistics*. Oxford: Oxford Univ. Press.

Daley, R. 1991. *Atmospheric data analysis*. Cambridge: Cambridge Univ. Press.

Dieker, T., and M. Mandjes. 2006. Fast simulation of overflow probabilities in a queue with Gaussian input. *ACM Trans. Modeling Comp. Simulation* 16:119–151.

Evans, A. C., S. Marret, P. Neelin, L. Collins, K. J. Worsley, W. Dai, S. Milot, E. Meyer, , and D. Bub. 1992. Anatomical mapping of functional activation in stereotactic coordinate space. *NeuroImage* 1:43–53.

Friston, K. j., A. P. Holmes, J. K. Worsley, J.-P. Polin, D. Frith, and R. S. J. Frackowiak. 1995. Statistical parametric maps in functional imaging: A general linear approach human brain mapping. *Human Brain Mapping* 2:189–210.

Giordano, S., M. Gubinelli, and M. Pagano. 2005. Bridge Monte-Carlo: a novel approach to rare events of Gaussian processes. In *Proc. 5th Workshop on Simulation*, 281–286.

Glasserman, P. 2004. *Monte Carlo methods in financial engineering*. New York: Springer-Verlag.

Gott, J. R., C. D. Hambrick, M. S. Vogeley, J. Kim, C. Park, Y.-Y. Choi, R. Cen, J. P. Ostriker, and K. Nagamine. 2008. Genus topology of structure in the sloan digital sky survey: Model testing. *The Astrophysical Journal* 675:16–28.

Juneja, S., and P. Shahabuddin. 2006. Rare event simulation techniques: An introduction and recent advances. *Handbook on Simulation*:291–350.

Olea, R. A. 1999. *Geostatistics for engineers and earth scientists*. Boston, MA: Kluwer Acad. Publ.

Pickands, J. 1969. Upcrossing probabilities for stationary gaussian processes. *Transactions of the American Mathematical Society* 145:51–73.

Piterbarg, V. I. 1995. *Asymptotic methods in the theory of gaussian processes and fields*. Providence, R.I.: American Mathematical Society.

Rubin 2002. *Applied stochastic hydrogeology*. New York: Oxford Univ. Press.

Shafie, K., B. Sigal, D. Siegmund, and K. J. Worsley. 2003. Rotation space random fields with an application to fMRI data. *Ann. Statist.* 31:1732–1771.

Smoot, G., and K. Davidson. 1993. *Wrinkles in Time*. New York: William Morrow.

Taylor, J., and R. Adler. 2003. Euler characteristics for gaussian fields on manifolds. *The Annals of Probability*.

Taylor, J. E., and K. J. Worsley. 2007. Detecting sparse signals in random fields, with an application to brain mapping. *J. AmerStatist. Assoc.* 479:913–928.

Torres, S. 1994. Topological analysis of COBE-DMR cosmic microwave background maps. *Astrophysical Journal* 423 (1).

Tsirelson, B., I. Ibragimov, and V. Sudakov. 1976. Norms of Gaussian sample functions. *Proceedings of the Third Japan-USSR Symposium on Probability Theory (Tashkent, 1975)* 550:20–41.

Vogeley, M. S., C. Park, M. J. Geller, J. P. Huchin, and J. R. Gott. 1994. Topological analysis of the CfA redshift survey. *Astrophysical Journal* 420:525–544.

AUTHOR BIOGRAPHIES

ROBERT J. ADLER holds the Louis and Samuel Seiden Academic Chair at the Technion-Israel Institute of Technology in Haifa and has a joint appointment between the Faculty of Industrial Engineering and Management and the Faculty of Electrical Engineering. He holds degrees form Sydney University (BSc), the Australian National Univeristy (MSc) and the University of New South Wales (PhD). He has published extensively in a wide variety of topics in pure and applied probability, including three books on random fields and Gaussian processes, with another one underway with Jonathan Taylor and Keith Worsley. He was the editor of *Stochastic Processes and Their Applications* from 1993-1995 and the *Annals of Applied Probability* from 2003-2005. His email address is `<robert@ieadler.technion.ac.il>`.

JOSE BLANCHET is an Assistant Professor in the Department of Industrial Engineering and Operations

Research at Columbia University. Jose holds a M.Sc. in Engineering- Economic Systems and Operations Research and a Ph.D. in Management Science and Engineering, both from Stanford University. He also holds two B.Sc. degrees: one in Actuarial Science and another one in Applied Mathematics from ITAM (Mexico). Jose worked for two years as an analyst in Protego Financial Advisors, a leading investment bank in Mexico. He has research interests in applied probability, computational finance, performance engineering, queueing theory, risk management, rare-event analysis, statistical inference, stochastic modelling, and simulation. His email address is `<jose.blanchet@columbia.edu>`.

JINGCHEN LIU is Assistant Professor in the Department of Statistics at Columbia University. Jingchen received his Ph.D. in statistics from Harvard University in June, 2008. His email address is `<jcliu@stat.columbia.edu>`.