Applied Probability Trust (31 August 2017)

UNIFORMLY EFFICIENT SIMULATION FOR EXTREMES OF GAUSSIAN RANDOM FIELDS

XIAOOU LI,* *University of Minnesota*

GONGJUN XU,** *University of Michigan*

Abstract

This paper considers the problem of simultaneously estimating rare-event probabilities for a class of Gaussian random fields. A conventional rare-event simulation method is usually tailored to a specific rare event and consequently would lose estimation efficiency for different events of interest, which often results in additional computational cost in such simultaneous estimation problem. To overcome this issue, we propose a uniformly efficient estimator for a general family of Hölder continuous Gaussian random fields. We establish the asymptotic and uniform efficiency of the proposed method and also conduct simulation studies to illustrate its effectiveness.

Keywords: Rare event, Importance Sampling, Gaussian Random Field

2010 Mathematics Subject Classification: Primary 65C05

Secondary 60G15

1. Introduction

Consider a continuous Gaussian random field $\{f(t) : t \in T\}$ with zero mean and unit variance, living on a d-dimensional compact set $T \subset R^d$; that is, for every finite subset of $\{t_1, ..., t_n\} \subset T$, $(f(t_1), ..., f(t_n))$ is a multivariate Gaussian random vector with $Ef(t_i) = 0$ and $Var(f(t_i)) = 1$ for $i = 1, \cdots, n$. We are interested in estimating the tail probability

$$w_{\sigma,\mu}(b) = P\left(\sup_{t\in T}\{\sigma(t)f(t) + \mu(t)\} > b\right), \text{ as } b \to \infty,$$

simultaneously for a class of continuous mean and variance functions $\mu(t)$ and $\sigma^2(t)$, where the functions $\mu(t)$ and $\sigma^2(t)$ may be unspecified and only known to be in certain ranges.

The extremes of Gaussian random fields have wide applications in finance, spatial analysis, physical oceanography, and many other disciplines [4, 5]. Tail probabilities of the extremes have been extensively studied in the literature, with its focus mostly on the development of approximations and bounds for the

* Postal address: School of Statistics, 224 Church ST SE, Minneapolis, MN, USA, 55455. Email: lixx1766@umn.edu

** Postal address: Department of Statistics, 1085 South University, Ann Arbor, MI, USA, 48109. Email: gongjun@umich.edu

suprema [e.g., 3, 7, 9–13, 18, 19, 29–33]. Tail probabilities of other convex functions of Gaussian random fields have also been studied; see [21, 23, 25, 28].

Most of the sharp theoretical approximations developed in the literature require the evaluation of certain constants that are hard to estimate, such as the Lipschitz-Killing curvatures and Pickands' constant. Moreover, although the asymptotic results may provide good approximations for large tail values as $b \to \infty$, evaluation of the approximation results for finite b may be challenging and it is often unclear how large the tail values are required to ensure the approximations within an acceptable range relative to the quantity of interest. Therefore, to evaluate the tail probabilities, rare-event simulation serves as an appealing alternative from a computational point of view. In particular, the design and the analysis do not require very sharp approximations of the tail probabilities. Importance sampling based efficient simulation procedures have been proposed in the literature to estimate the tail probabilities. Numerical methods for rare-event analysis of the suprema were studied in [1, 2]; see also [8, 20, 24, 26–28, 34] for related studies.

To design asymptotically efficient importance sampling estimator, one needs to construct a change of measure that is tailored to a specific event. Such construction usually requires detailed information of the Gaussian random fields, such as $\mu(t)$ and $\sigma(t)$ whose computations themselves are sometimes intensive. In addition, the specific form of the change of measure is sensitive to $\mu(t)$ and $\sigma(t)$ in the sense that the entire simulation needs to be redone even if there is a tiny change of the system. This often leads to additional computational overhead especially at the exploratory stage when one often needs to tune different model parameters. This motivates us to seek for a single Monte Carlo scheme that is efficient for a class of distributions. An advantage of such uniformly efficient methods is that there is no need to regenerate samples if there is a change in the original system and one just needs to recompute the importance weights. This could save substantial computational time. Moreover, this can help researchers efficiently estimate many probabilities for a certain range of mean and variance parameter values, which are often of practical importance. For instance, in finance risk analysis, there is often uncertainty surrounding the true population values for the mean and variance; portfolio credit risk management may require the estimation of the tail probabilities of extremes for a family of Gaussian processes; in physical system reliability analysis, we may need to evaluate the failure probability for a range of system parameters.

To address the above issues, this study focuses on the problem of simultaneous efficient estimation of $w_{\sigma,\mu}(b)$ for all possible $\mu(t) \in [\mu_l, \mu_u]$ and $\sigma^2(t) \in [\sigma_l^2, \sigma_u^2]$, $t \in T$, where $\mu_l \le \mu_u \in R$ and $\sigma_l \le \sigma_u \in (0, \infty)$ are constants that are prespecified. We propose a mixture type change of measure that yields uniformly efficient estimation (criterion defined in Section 2). In particular, the uniform efficiency result holds for general Hölder continuous Gaussian random fields and therefore it is applicable to most of the practical problems.

The remainder of the paper is organized as follows. In Section 2 we introduce some notions of efficiency and computational complexity under the setting of rare-event simulation. Section 3 provides the construction of our importance sampling estimator and shows the main properties of our algorithm. Numerical simulations are conducted in Section 4 and detailed proofs of our main theorems are given in Section 5.

2. Efficiency Criteria

2.1. Efficiency of rare-event simulation and importance sampling

We first introduce some general notions of rare-event simulations. Given that the tail probability $w_{\sigma,\mu}(b)$ converges to zero, it is usually meaningful to consider the relative error of a Monte Carlo estimator $L(b)$ with respect to $w_{\sigma,\mu}(b)$. This is because a trivial estimator $L^*(b) \equiv 0$ has an error $|L^*(b) - w_{\sigma,\mu}(b)| = w_{\sigma,\mu}(b) \to 0$. In the literature of rare-event simulation (e.g., [2, 6, 17]), one usually employs the concept of *polynomially efficiency* as an efficiency criterion.

Definition 1. (*Polynomial efficiency.*) An estimator $L(b)$ is said to be *polynomially efficient with the order* q in estimating $w_{\sigma,\mu}(b)$ if $EL(b) = w_{\sigma,\mu}(b)$ and there exist constants $q \geq 0$ and $b_0 \geq 0$ such that

$$\sup_{b \geq b_0} \frac{Var(L(b))}{|\log w_{\sigma,\mu}(b)|^q w^2_{\sigma,\mu}(b)} < \infty. \tag{1}$$

When $q = 0$, $L(b)$ is also called strongly efficient.

To illustrate this efficiency criterion, we compare a polynomially efficient estimator with a standard Monte Carlo estimator. Suppose that we want to estimate $w_{\sigma,\mu}(b)$ with certain relative accuracy with a high probability. That is, we would like to have an estimator $Z(b)$ such that for some prescribed $\varepsilon, \delta > 0$,

$$P\left(|Z(b)/w_{\sigma,\mu}(b) - 1| > \varepsilon\right) < \delta. \tag{2}$$

If a standard Monte Carlo simulation method is used, then it requires at least $n = O(\varepsilon^{-2}\delta^{-1}w^{-1}_{\sigma,\mu}(b))$ i.i.d. replicates, according to the central limit theorem. By the Borell-TIS lemma (Lemma 3), we know $w_{\sigma,\mu}(b) \leq \exp\{-(1 + o(1))b^2/(2\sup_{t\in T}\sigma^2(t))\}$. Therefore, n has to grow at an exponential rate in b^2. On the contrary, suppose that a polynomially efficient estimator of $w_{\sigma,\mu}(b)$ has been obtained, denoted by $L(b)$. Let $\{L^{(j)}(b) : j = 1, ..., n\}$ be n i.i.d. copies of $L(b)$. Then the averaged estimator $Z(b) = \frac{1}{n}\sum_{j=1}^n L^{(j)}(b)$ has a mean squared error (MSE) $E(Z(b) - w_{\sigma,\mu}(b))^2 = Var(L(b))/n$. A direct application of Chebyshev's inequality yields

$$P(|Z(b)/w_{\sigma,\mu}(b) - 1| \geq \varepsilon) \leq \frac{Var(L(b))}{n\varepsilon^2 w^2_{\sigma,\mu}(b)}. \tag{3}$$

Thus, if $L(b)$ is a polynomially efficient estimator with the order q, it suffices to simulate $n = \varepsilon^{-2}\delta^{-1}|\log w_{\sigma,\mu}(b)|^q = O(\varepsilon^{-2}\delta^{-1}b^{2q})$ i.i.d. replicates of $L(b)$ to achieve the accuracy in (2). Compared with the standard Monte Carlo simulation, polynomially efficient estimators reduce the computational cost substantially for large b.

Remark 1. In the rare event analysis literature, another widely used efficiency criterion is the *weakly efficient* ([6]). An estimator $L(b)$ is said to be weakly efficient in estimating $w_{\sigma,\mu}(b)$, if $EL(b) = w_{\sigma,\mu}(b)$ and for all positive constants $\varepsilon > 0$,

$$\limsup_{b\to\infty} \frac{Var(L(b))}{w^{2-\varepsilon}_{\sigma,\mu}(b)} = 0.$$

It is easy to verify that if $L(b)$ is polynomially efficient, then $L(b)$ is also weakly efficient. That is, polynomial efficiency is a stronger criterion than the weak efficiency.

To construct polynomially efficient estimators, importance sampling is a commonly used method for the variance reduction. In particular, we have

$$w_{\sigma,\mu}(b) = E\Big[I\Big(\sup_{t\in T}\{\sigma(t)f(t)+\mu(t)\} > b\Big)\Big] = E^Q\Big[\frac{dP}{dQ}I\Big(\sup_{t\in T}\{\sigma(t)f(t)+\mu(t)\} > b\Big)\Big],$$

where $I(\cdot)$ denotes the indicator function, Q is a probability measure that is absolutely continuous with respect to P on the set $\{\sup_{t\in T}\{\sigma(t)f(t)+\mu(t)\} > b\}$, and we use E and E^Q to denote the expectations under the measures P and Q, respectively. Then, the random variable defined by

$$L_{\sigma,\mu}(b) = \frac{dP}{dQ}I\Big(\sup_{t\in T}\{\sigma(t)f(t)+\mu(t)\} > b\Big) \tag{4}$$

is an unbiased estimator of $w_{\sigma,\mu}(b)$ under the measure Q. To have an efficient estimator, we want to choose Q such that the variance $Var^Q(L_{\sigma,\mu}(b))$ is small. It is straightforward to show that the optimal change of measure is the conditional probability $Q^*(\cdot) := P(\cdot \mid \sup_{t\in T}\{\sigma(t)f(t)+\mu(t)\} > b) = P(\cdot \cap \{\sup_{t\in T}\{\sigma(t)f(t)+\mu(t)\} > b\})/w_{\sigma,\mu}(b)$, for which the corresponding importance sampling estimator has a zero variance. However, Q^* cannot be implemented in practice because $w_{\sigma,\mu}(b)$, the probability of interest, is unknown beforehand. Therefore, constructing an efficient change of measure usually involves analysis and approximation of the optimal change of measure Q^*.

2.2. Non-uniformly efficient issue and an example

Various importance sampling estimators for rare-event analysis of the suprema of Gaussian random fields have been studied in [1, 2, 8, 20]. As the measure Q^* depends on the mean and variance function $\sigma(\cdot)$ and $\mu(\cdot)$, the designed measures usually depend on the $\mu(\cdot)$ and $\sigma(\cdot)$ as well. As a consequence, a measure Q that gives an efficient estimator $L_{\sigma,\mu}(b) = \frac{dP}{dQ}I(\sup_{t\in T}\{\sigma(t)f(t)+\mu(t)\} > b)$ for $w_{\sigma,\mu}(b)$ may not be efficient any more for estimating $w_{\sigma',\mu'}(b)$, where $\sigma'(t)$ and $\mu'(t)$ are two different variance and mean functions. That is, the corresponding importance sampling estimator based on Q

$$L_{\sigma',\mu'}(b) := \frac{dP}{dQ}I(\sup_{t\in T}\{\sigma'(t)f(t)+\mu'(t)\} > b)$$

may not be an efficient estimator for $w_{\sigma',\mu'}(b)$.

To illustrate the non-uniform efficiency issue, we take the estimator proposed in [2] as an example. For simplicity, we consider the case when T contains finite points and write $T := \{t_1, \cdots, t_M\}$.

For known μ and σ, [2] proposed the following simulation procedure in Algorithm 1. Let $Q^\dagger$ be the corresponding change of measure. We have

$$\frac{dQ^\dagger}{dP} = \frac{\sum_{i=1}^M I(\sigma(t_i)f(t_i)+\mu(t_i) > b)}{\sum_{i=1}^M P(\sigma(t_i)f(t_i)+\mu(t_i) > b)}.$$

[2] showed that $L_{\sigma,\mu}(b) = \frac{dP}{dQ^\dagger}I(\sup_{t\in T}\{\sigma(t)f(t)+\mu(t)\} > b)$ is a polynomially efficient estimator for $w_{\sigma,\mu}(b)$ with the order $q = 0$. We explain intuitively why this estimator is efficient. First, Algorithm 1 samples a

Algorithm 1: Sampling procedure proposed by [2]

Input: $T = \{t_1, \cdots, t_M\}$.

1 Simulate a random variable $\tau \in \{t_1, \cdots, t_M\}$ according to the following probability measure:

$$P(\tau = t_i) = \frac{P(\sigma(t_i)f(t_i) + \mu(t_i) > b)}{\sum_{j=1}^{M} P(\sigma(t_j)f(t_j) + \mu(t_j) > b)}; \tag{5}$$

2 Given the realized τ, simulate $f(\tau)$ conditional on $\sigma(\tau)f(\tau) + \mu(\tau) > b$;

3 Given $(\tau, f(\tau))$, simulate the rest $\{f(t) : t \neq \tau, t \in T\}$ from the original conditional distribution under P.

Output: $f(t)$ for $t \in T$

random index τ whose distribution is approximating that of $t^* := \arg\max_{t_i}(\sigma(t_i)f(t_i) + \mu(t_i))$. Second, it simulates $f(\tau)$ approximately from the conditional distribution $P(f(t^*) \in \cdot | f(t^*) > b)$. Third, Algorithm 1 simulates the $f(t)$ at $t \neq \tau$ according to the original conditional distribution given $(f(\tau), \tau)$. Combining the three steps, the entire sample path $\{f(t) : t \in T\}$ generated from Algorithm 1 approximately follows the conditional distribution $\{f(t) : t \in T | \max_{t_i}(\sigma(t_i)f(t_i) + \mu(t_i)) > b\}$. According to the discussion on page 4, this conditional probability measure is the optimal change of measure. See [2] for rigorous justifications of the above statements.

Let μ' and σ' be a different mean and variance function. We have Proposition 1 for the estimator

$$L_{\sigma',\mu'}(b) := \frac{dP}{dQ^\dagger} I\Big(\sup_{t\in T}\{\sigma'(t)f(t) + \mu'(t)\} > b\Big).$$

Proposition 1. *Let $\mu'(t) = \mu(t) = 0$ for all $t \in T$.*

(i) If $\sigma'(t) \leq \sigma(t)$ for all $t \in T$ and $\max_{t_i\in T} \sigma'(t_i) < \max_{t_i\in T} \sigma(t_i)$, then for some constant $\varepsilon > 0$,

$$\lim_{b\to\infty} \frac{E^{Q^\dagger}\left[\left(\frac{dP}{dQ^\dagger}\right)^2; \max_{t_i\in T}\sigma'(t_i)f(t_i) > b\right]}{w^{2-\epsilon}_{\sigma',\mu}(b)} = \infty.$$

(ii) If $\max_{t_i\in T} \sigma'(t_i) > \max_{t_i\in T} \sigma(t_i)$, then $\frac{dP}{dQ^\dagger}$ is not well defined on the event $\{\max_{t_i\in T}\sigma'(t_i)f(t_i) > b\}$.

According to the definition of weakly efficient estimator in Remark 1, the first part of the above proposition implies that $L_{\sigma,\mu}(b)$ is not weakly efficient for estimating $w_{\sigma',\mu'}(b)$ if $\max_{t_i\in T} \sigma'(t_i) > \max_{t_i\in T} \sigma(t_i)$, and is therefore not polynomially efficient. The second part of the above proposition implies that the estimator $L_{\sigma,\mu}(b)$ is not well defined when $\max_{t_i\in T} \sigma'(t_i) > \max_{t_i\in T} \sigma(t_i)$. Therefore, for each $L_{\sigma,\mu}(b)$ there always exist mean and variance functions $\mu'(\cdot)$, $\sigma'(\cdot)$ such that $\mu'(t) \in [\mu_l, \mu_u], \sigma'(t) \in [\sigma_l, \sigma_u]$ and $L_{\sigma,\mu}(b)$ is *not* (weakly) efficient for estimating $w_{\sigma',\mu'}(b)$. We use a simple numerical study to further illustrate this.

Example 1. Consider i.i.d. standard normal random variables $\{f(t), t = 1, \cdots, 100\}$. For simplicity, we take $\mu(t) = 0$ and $\sigma(t) = \sigma$ for all t. The probability of interest is $P(\sigma \max_t f(t) > b)$ for $\sigma \in [0.3, 1]$ and $b = 3$. This is equivalent to simulating $P(\max_t f(t) > b)$ for all $b \in [3, 10]$. Table 1 displays the simulation results for $\sigma = 0.3, 0.6$ and 1, from Algorithm 1, where the change of measure is constructed based on $\sigma = 1$. The results are based on 10^4 independent simulations. We report the estimated tail probability (est.), the estimated standard deviation (sd.) of $L_{\sigma,\mu}(b)$, and the coefficient of variation (CV), which is the ratio sd./est.. We also give the theoretical values of the tail probabilities, that is, $P(\max_i f(t_i) > b/\sigma) = 1 - \Phi(b/\sigma)^{100}$ where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$ denotes the left tail probability of the standard Gaussian distribution. We can see that the estimator is more efficient when σ value is equal to the designed value 1 and less for other σ values. In particular, when $\sigma = 0.3$, it gives 0 estimated value.

σ	est.	sd.	CV	Theoretical Value
0.3	0	0	NA	7.62e-22
0.6	1.35e-05	1.35e-03	1.00e+02	2.87e-05
1	1.26e-01	2.32e-02	1.84e-01	1.26e-01

Table 1: Estimates based on Algorithm 1

The above non-uniform efficiency result can be extended, with similar techniques, to the importance sampling estimators in [2] when $\{f(t), t \in T\}$ is a continuous Gaussian random field. It can also be extended to the case when other change of measures are used such as [20]. In general, if the construction of a rare-event change of measure relies heavily on the mean and variance functions, then it would not be efficient for another set of functions.

2.3. Uniform Efficiency

In applications, one is often interested in estimating many probabilities for a certain range of mean and variance parameter values, such as evaluating the tail probabilities of a loss distribution for a range of loss thresholds in portfolio credit risk management (e.g., [15, 16]). This motivates us to construct a change of measure such that the corresponding importance sampling estimator $L_{\sigma,\mu}(b)$ is polynomially efficient for a family of functions μ and σ. In particular, this paper considers μ and σ satisfying the following condition:

C1. For all $t \in T$, $\mu(t) \in [\mu_l, \mu_u]$ and $\sigma^2(t) \in [\sigma_l^2, \sigma_u^2]$. Moreover, μ and σ are Hölder continuous in the sense that there exists positive constants κ_H and $\beta > 0$ such that for all $s, t \in T$ $|\sigma(t) - \sigma(s)| + |\mu(t) - \mu(s)| \leq \kappa_H |s - t|^\beta$.

Denote by $\mathcal{C}(\mu_l, \mu_u, \sigma_l, \sigma_u, \beta, \kappa_H)$ the class of functions $\sigma(\cdot)$ and $\mu(\cdot)$ that satisfy Assumption C1. We introduce the following uniform efficiency criterion.

Definition 2. (*Uniform polynomially efficient change of measure.*) We say that a change of measure Q is uniformly polynomially efficient with the order $q \geq 0$ if there exists a constant $b_0 \geq 0$ such that the importance sampling estimator

$$L_{\sigma,\mu}(b) = \frac{dP}{dQ} I\Big(\sup_{t\in T}\{\sigma(t)f(t)+\mu(t)\} > b\Big)$$

satisfies

$$\sup_{b\geq b_0,\mu,\sigma\in\mathcal{C}(\mu_l,\mu_u,\sigma_l,\sigma_u,\beta,\kappa_H)} \frac{Var(L_{\sigma,\mu}(b))}{|\log w_{\sigma,\mu}(b)|^q w^2_{\sigma,\mu}(b)} < \infty. \tag{6}$$

Similar to the previous discussion, we consider the relative accuracy of a class of the importance sampling estimators corresponding to a uniformly polynomially efficient change of measure. Let the Q be uniformly polynomially efficient for $\sigma(\cdot),\mu(\cdot)\in\mathcal{C}(\mu_l,\mu_u,\sigma_l,\sigma_u,\beta,\kappa_H)$. Then, according to (3), there exists some $\kappa_u > 0$, such that the averaged estimator $Z_{\sigma,\mu}(b) = \frac{1}{n}\sum_{i=1}^n L^{(i)}_{\sigma,\mu}(b)$ based on $n = \kappa_u b^{2q}\delta^{-1}\varepsilon^{-2}$ i.i.d. Monte Carlo samples satisfies

$$\sup_{(\sigma,\mu)\in\mathcal{C}(\mu_l,\mu_u,\sigma_l,\sigma_u,\beta,\kappa_H)} P\left(|Z_{\sigma,\mu}(b) - w_{\sigma,\mu}(b)| > \varepsilon w_{\sigma,\mu}(b)\right) < \delta.$$

Remark 2. Although the current paper focuses on rare-event simulation for the extremes of Gaussian random fields, the uniform efficiency criterion as well as the proposed method can be easily extended to other Gaussian-related rare-event problems, such as the exponential integrals of Gaussian random fields [e.g., 27, 28], where the mean and variance functions are unspecified and we are interested in estimating a family of tail probabilities. Moreover, the proposed method can be extended to the estimation of non-Gaussian tail probabilities. For instance, in statistical hypothesis testing with data generated independently from certain distribution with unknown parameters that are of interest, it often needs to evaluate the test power/error probabilities for a range of model parameters as the sample size increase; see [22] for an example.

Remark 3. In the literature, a similar uniform efficiency definition has been proposed in [16] to design an algorithm that is asymptotically efficient uniformly for a family of probability sets when estimating the tail probabilities of sums of light tailed random variables. Differently from this study, the random variable parameters are assumed to be known in their case.

3. Uniformly Efficient Estimation

3.1. Discrete case

We start with the case when T contains finite points and propose a new change of measure which gives a uniformly efficient estimator. We assume $T := \{t_1,\cdots,t_M\}$. We describe the new measure Q in two ways. First, we specify the sampling scheme of f under Q and then provide its Radon-Nikodym derivative with respect to P. Under the measure Q, $f(t)$ is generated according to the following algorithm.

Algorithm 2: Simulating $f(\cdot)$ under Q

Input: $T = \{t_1, \cdots, t_M\}$, $\delta_b = ab^{-1}$ for some constant $a > 0$.

1 Simulate a random variable ς with respect to some positive continuous density function g on $[\sigma_l, \sigma_u + \delta_b^2]$;

2 Simulate a random variable ν with respect to some positive continuous density function h on $[\mu_l, \mu_u + \delta_b]$;

3 Simulate a random variable τ uniformly over $T = \{t_1, \cdots, t_M\}$;

4 Given the realized ς, ν and τ, simulate $f(\tau)$ conditional on $\varsigma f(\tau) + \nu > b$;

5 Given $(\tau, f(\tau))$, simulate the Gaussian process $\{f(t) : t \neq \tau, t \in T\}$ from the original conditional distribution under P.

Output: $f(t)$ for $t \in T$

For the measure Q defined above, it is not hard to verify that P and Q are mutually absolutely continuous with the Radon-Nikodym derivative being

$$\frac{dQ}{dP} = \int_{\mu_l}^{\mu_u+\delta_b} \int_{\sigma_l}^{\sigma_u+\delta_b^2} \frac{\sum_{i=1}^M I(\varsigma f(t_i) + \nu > b)}{MP(\varsigma f(t_1) + \nu > b)} g(\varsigma) h(\nu) d\varsigma d\nu.$$

This gives the importance sampling estimator

$$\begin{aligned} L_{\sigma,\mu}(b) \quad &= \quad \left(\int_{\mu_l}^{\mu_u+\delta_b} \int_{\sigma_l}^{\sigma_u+\delta_b^2} \frac{\sum_{i=1}^M I(\varsigma f(t_i) + \nu > b)}{MP(\varsigma f(t_1) + \nu > b)} g(\varsigma) h(\nu) d\varsigma d\nu \right)^{-1} \\ &\quad\; \times I(\sup_{i:t_i \in T} \sigma(t_i) f(t_i) + \mu(t_i) > b). \end{aligned} \tag{7}$$

Note that under Q, if $\max_{t_i \in T} \sigma(t_i) f(t_i) + \mu(t_i) > b$, then $\varsigma f(t_i) + \nu > b$ holds for all i, $\varsigma > \max_{t_i \in T} \sigma(t_i)$ and $\nu > \max_{t_i \in T} \mu(t_i)$. Therefore, the change of measure is well defined.

We take a closer look at the proposed change of measure Q by comparing it with the measure $Q^\dagger$ discussed in Section 2.2. We can see that steps 1 and 2 of Algorithm 1 requires the knowledge of the mean and variance function μ and σ. When μ and σ are unknown, running Algorithm 1 with a misspecified μ' and σ' may cause inefficiency. The proposed Algorithm 2 avoids this inefficiency by introducing prior probability density functions g and h. Intuitively, the proposed algorithm explores each possible values of mean and variance of the random field at a random index (steps 1-3), and is a hybrid scheme for all $\sigma(\cdot)$ and $\mu(\cdot)$ that take values in the support of g and h. The next proposition states the uniform efficiency of the proposed change of measure.

Proposition 2. *Let $L_{\sigma,\mu}(b)$ be defined in (7), then there exist constants b_0 and κ_p, independent of $\sigma(\cdot), \mu(\cdot)$ and b and for $b \geq b_0$,*

$$\frac{E^Q(L_{\sigma,\mu}^2(b))}{M^2 b^6 w_{\sigma,\mu}^2(b)} \leq \kappa_p$$

for all μ and σ satisfying C1.

Note that $|\log(w_{\sigma,\mu}(b))| = O(b^2)$. Therefore, the above proposition gives the uniformly polynomial efficiency of Q with the order $q = 3$ for the discrete case.

Remark 4. The parameter δ_b in Algorithm 2 is introduced to control the second moment of the importance sampling estimator. Otherwise, consider the case of constant variance $\sigma \in [\sigma_l, \sigma_u]$ and zero mean $\mu = 0$. Then for σ taking the value of σ_u, denote the corresponding estimator by $L_{\sigma_u,N}(b)$ and the second moment of $L_{\sigma_u,N}(b)$ is lower bounded by

$$\begin{aligned}
E^Q[L^2_{\sigma_u,N}(b)] &= E^Q\left[\left(\frac{dP}{dQ}\right)^2;\ \max_i \sigma_u f(t_i) > b\right] = E\left[\frac{dP}{dQ};\ \max_i \sigma_u f(t_i) > b\right] \\
&= E\left[\left(\int_{\mu_l}^{\mu_u}\int_{\sigma_l}^{\sigma_u} \frac{\sum_{i=1}^M I(\varsigma f(t_i) > b)}{MP(\varsigma f(t_1) > b)} g(\varsigma)h(\nu)d\varsigma d\nu\right)^{-1};\ \max_i \sigma_u f(t_i) > b\right] \\
&\geq P(\sigma_l f(0) > b)P(\max_i \sigma_u f(t_i) > b) \\
&\quad \times E\left[\left(\int_{\mu_l}^{\mu_u}\int_{\sigma_l}^{\sigma_u} I(\max_i f(t_i) > \varsigma^{-1}b)g(\varsigma)h(\nu)d\varsigma d\nu\right)^{-1} \Big| \max_i f(t_i) > \sigma_u^{-1}b\right].
\end{aligned}$$

However, the conditional expectation cannot be controlled and we have the estimator $L_{\sigma_u,N}(b)$ is not efficient for $\sigma = \sigma_u$.

Remark 5. To evaluate the Radon-Nikodym derivative in (7), we need to calculate the integral

$$\int_{\mu_l}^{\mu_u+\delta_b}\int_{\sigma_l}^{\sigma_u+\delta_b^2} \frac{\sum_{i=1}^M I(\varsigma f(t_i) + \nu > b)}{MP(\varsigma f(t_1) + \nu > b)} g(\varsigma)h(\nu)d\varsigma d\nu.$$

Define

$$l(z) = \int_{\mu_l}^{\mu_u+\delta_b}\int_{\sigma_l}^{\sigma_u+\delta_b^2} \frac{I(\varsigma z + \nu > b)}{\bar{\Phi}((b-\nu)/\varsigma)} g(\varsigma)h(\nu)d\varsigma d\nu, \tag{8}$$

where $\bar{\Phi}(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ is the right tail probability of a standard Gaussian distribution, then we have

$$\int_{\mu_l}^{\mu_u+\delta_b}\int_{\sigma_l}^{\sigma_u+\delta_b^2} \frac{\sum_{i=1}^M I(\varsigma f(t_i) + \nu > b)}{MP(\varsigma f(t_1) + \nu > b)} g(\varsigma)h(\nu)d\varsigma d\nu = \frac{1}{M}\sum_{i=1}^M l(f(t_i)).$$

Therefore, we only need to evaluate $l(f(t_i))$ for all $f(t_i)$ simulated by Algorithm 2. We use the following simplification for the function $l(z)$. Let $s = \frac{b-\nu}{\varsigma}$, then

$$l(z) = \int\int_{b-\varsigma s \in I_1, \varsigma \in I_2, s<z} \varsigma/\bar{\Phi}(s) g(\varsigma)h(b - s\varsigma)d\varsigma ds = \int_{s<z} 1/\bar{\Phi}(s) \int_{\varsigma \in (\frac{b}{s} - \frac{1}{s}I_1)\cap I_2} \varsigma h(b - s\varsigma)g(\varsigma)d\varsigma ds \tag{9}$$

where $I_1 = [\mu_l, \mu_u + \delta_b]$, and $I_2 = [\sigma_l, \sigma_u + \delta_b^2]$. We can then choose $h(\cdot)$ and $g(\cdot)$ so that the inner integral in (9) has a closed form expression. In particular, in the numerical examples in this paper, we choose $g(\cdot)$ and $h(\cdot)$ to be the density functions of uniform distributions. In this case, let $r(s) = \frac{1}{2}(\sigma_u + \delta_b^2 - \sigma_l)^{-1}(\mu_u + \delta_b - \mu_l)^{-1}\int_{\varsigma \in (\frac{b}{s} - \frac{1}{s}I_1)\cap I_2} d\varsigma^2$, then $l(z)$ can be further simplified as

$$l(z) = \int_{-\infty}^{z} r(s)/\bar{\Phi}(s) ds,$$

which is a one-dimensional integral and can be evaluated numerically.

3.2. Continuous case

Direct simulation of a continuous random field is typically not a feasible task, and the change of measure proposed in the previous subsection is not directly applicable. Thus, we use a discrete object to approximate the continuous fields for the implementation. In particular, we create a regular lattice covering T in the following way. Let $G_{N,d}$ be a countable subset of R^d: $G_{N,d} = \left\{\left(\frac{i_1}{N}, \frac{i_2}{N}, ..., \frac{i_d}{N}\right) : i_1, ..., i_d \in \mathbb{Z}\right\}$. That is, $G_{N,d}$ is a regular lattice on R^d. Furthermore, let

$$T_N = G_{N,d} \cap T, \tag{10}$$

which is the sub-lattice intersecting with T. Since T is compact, T_N is a finite set. We enumerate the elements in $T_N = \{t_1, \cdots, t_M\}$. Because T is compact, we have $M = O(N^d)$. Let

$$w_{\sigma,\mu,N}(b) = P\left(\sup_{t_i \in T_N} \sigma(t_i)f(t_i) + \mu(t_i) > b\right).$$

We use $w_{\sigma,\mu,N}(b)$ as a discrete approximation of $w_{\sigma,\mu}(b)$. We estimate $w_{\sigma,\mu,N}(b)$ by importance sampling, which is based on the change of measure proposed in Section 3.1. In particular we define Q_N and P_N as the discrete versions (on T_N) of Q and P respectively. Then dQ_N/dP_N takes the form:

$$\frac{dQ_N}{dP_N} = \int_{\mu_l}^{\mu_u+\delta_b} \int_{\sigma_l}^{\sigma_u+\delta_b^2} \frac{\sum_{i=1}^{M} I(\varsigma f(t_i) + \nu > b)}{MP(\varsigma f(t_1) + \nu > b)} g(\varsigma)h(\nu)d\varsigma d\nu. \tag{11}$$

Note that here M depends on N and goes to infinity as $N \to \infty$. This gives importance sampling estimator

$$\begin{aligned} L_{\sigma,\mu,N}(b) \quad := \quad & \left(\int_{\mu_l}^{\mu_u+\delta_b} \int_{\sigma_l}^{\sigma_u+\delta_b^2} \frac{\sum_{i=1}^{M} I(\varsigma f(t_i) + \nu > b)}{MP(\varsigma f(t_1) + \nu > b)} g(\varsigma)h(\nu)d\varsigma d\nu\right)^{-1} \\ & \times I(\sup_{i:t_i \in T_N} \sigma(t_i)f(t_i) + \mu(t_i) > b). \end{aligned}$$

The discretization usually introduces bias. The next two theorems control the bias and variance of the estimator $L_{\sigma,\mu,N}(b)$ under the following assumptions.

C2 There exists a positive constant κ_m such that $\sup_{t\in T} \min_{t' \in T_N} |t - t'| \leq \frac{\kappa_m}{N}$ for all N.

C3 The Gaussian random field f is almost surely continuous.

C4 Define the correlation function $r(s,t) = E(f(s)f(t))$. There exists $\beta' > 0$ and $\kappa'_H > 0$ such that

$$|r(t,s) - r(t',s')| \leq \kappa'_H[|t - t'|^{\beta'} + |s - s'|^{\beta'}] \tag{12}$$

for all $s, t, s', t' \in T$.

Theorem 1. *Let $\beta^* = \min(\beta, \beta')$ and $N_0(\varepsilon, b) = b^{2/\beta^*(\frac{3d}{\beta^*}+2-\varepsilon_0)}\varepsilon^{-2/\beta^*+\varepsilon_0}$. Under Assumptions C1-C4, for any $\varepsilon_0 > 0$, there exist constants κ_0 and b_0 such that for any $\varepsilon \in (0,1)$, if $N \geq N_0(\varepsilon, b)$ and $b > b_0$, then*

$$\frac{|w_{\sigma,\mu,N}(b) - w_{\sigma,\mu}(b)|}{w_{\sigma,\mu}(b)} < \varepsilon$$

uniformly for $\mu, \sigma \in \mathcal{C}(\mu_l, \mu_u, \sigma_l, \sigma_u, \beta, \kappa_H)$.

Theorem 2. *Let $N_0(\varepsilon, b)$ be defined in Theorem 1. Under Assumptions C1-C4, if $N \geq N_0(\varepsilon, b)$, then there exist constants $b_0 > 0$ (depending on ε_0) and $\kappa_c > 0$ such that*

$$\sup_{b \geq b_0, \varepsilon \in (0,1)} \frac{E^{Q_N} L^2_{\sigma,\mu,N}(b)}{b^q w^2_{\sigma,\mu}(b) \varepsilon^{-q_1}} < \kappa_c$$

uniformly for $\mu, \sigma \in \mathcal{C}(\mu_l, \mu_u, \sigma_l, \sigma_u, \beta, \kappa_H)$ with $q = 4d/\beta^*(\frac{3d}{\beta^*} + 2 + \varepsilon_0) + 6$ and $q_1 = 4d/\beta^* + 2d\varepsilon_0$.

We consider the relative accuracy of the importance sampling estimator based on Q_N. Let $L^{(i)}_{\sigma,\mu,N}(b)$ be i.i.d. copies of $L_{\sigma,\mu}(b)$ for $i = 1, .., n$. Let

$$Z_{\sigma,\mu,N}(b) = \frac{1}{n} \sum_{i=1}^{n} L^{(i)}_{\sigma,\mu,N}(b). \tag{13}$$

With the aid of Chebyshev's inequality, we have

$$P(|Z_{\sigma,\mu,N}(b) - w_{\sigma,\mu}(b)| > \varepsilon w_{\sigma,\mu}(b)) \leq \frac{E(Z_{\sigma,\mu,N}(b) - w_{\sigma,\mu}(b))^2}{\varepsilon^2 w^2_{\sigma,\mu}(b)}.$$

The mean squared error $E(Z_{\sigma,\mu,N}(b) - w_{\sigma,\mu}(b))^2$ can be written as

$$E(Z_{\sigma,\mu,N}(b) - w_{\sigma,\mu}(b))^2 = [EZ_{\sigma,\mu,N}(b) - w_{\sigma,\mu}(b)]^2 + Var(Z_{\sigma,\mu,N}(b)).$$

The first and second terms on the right-hand side of the above display is the squared bias and the variance of the estimator $Z_{\sigma,\mu,N}(b)$, respectively. If we choose $N = N_0(\varepsilon\delta^{1/2}, b)$ according to Theorem 1 and let $n = 2\kappa_c b^q \varepsilon^{-q_1-2} \delta^{-\frac{q_1}{2}-1}$ where q and q_1 are defined in Theorem 2, then the MSE is well controlled relative to $w_{\sigma,\mu}(b)$ and so is the relative accuracy. We summarize this result in the next corollary.

Corollary 1. *Under the Assumption C1-C4, let $Z_{\sigma,\mu,N}(b)$ be defined in* (13). *If we choose $n = 2\kappa_c b^q \varepsilon^{-q_1-2} \delta^{-\frac{q_1}{2}-1}$ and $N = N_0(\varepsilon\delta^{1/2}, b)$, then*

$$P\left(|Z_{\sigma,\mu,N}(b)/w_{\sigma,\mu}(b) - 1| > \varepsilon\right) < \delta. \tag{14}$$

Remark 6. The computational complexity for generating $Z_{\sigma,\mu,N}(b)$ is n multiplied by the cost for generating one copy of $L_{\sigma,\mu,N}(b)$. The cost for generating $L_{\sigma,\mu,N}(b)$ is of order $O(M^3) = O(N^{3d})$, which is mainly the cost of generating a multivariate Gaussian vector (line 5 of Algorithm 2). The overall computational cost is also a polynomial in ε, δ and b. Algorithm with such a computation cost to achieve (14) is sometimes referred to as a fully polynomial randomized approximation scheme (FPRAS), see [2] for more details.

4. Simulation Studies

In this section, we present numerical examples to show the performance of the proposed algorithm. All the results are based on $n = 10^4$ independent simulations. The discretization size is chosen as $M = 40$ in Example 2-5. In each numerical example, we report the estimated tail probabilities, which will be referred to as "est.", along with the estimated standard deviations, that is $sd^Q\{L_{\sigma,\mu}(b)\} = \sqrt{Var^Q\{L_{\sigma,\mu}(b)\}}$, which will be referred to as "sd.". The standard error of the estimator with 10^4 Monte Carlo samples is sd./100. We also report the coefficient of variation (CV) of the estimators, which is the ratio sd./est. of the estimators.

We start with the discrete setting in Example 1, where $T = \{1, \cdots, 100\}$ and $\{f(t), t = 1, \cdots, 100\}$ are i.i.d. standard normal random variables. We take $\mu(t) = 0$ and $\sigma(t) = \sigma$ with $\sigma \in [0.3, 1]$ for all $t \in T$, and the probability of interest is $P(\sigma \max_t f(t) > b)$ for $b = 3$. Table 2 displays the simulation results for $\sigma = 0.3, 0.6$ and 1 using the proposed method. For different σ values, the estimates are close to the true values. Compared with the result of Algorithm 1 in Table 1, the proposed method gives better overall performance.

σ	est.	sd.	CV	Theoretical Value
0.3	7.55e-22	5.33e-21	7.05	7.62e-22
0.6	2.93e-05	1.33e-04	4.52	2.87e-05
1	1.26e-01	5.92e-01	4.69	1.26e-01

TABLE 2: Estimates of $w_\sigma(b)$, $sd^Q(L_{\sigma,\mu}(b))$, and $sd^Q(L_{\sigma,\mu}(b))/w_\sigma(b)$. All results are based on 10^4 independent simulations and thus the standard errors of the estimates are $sd^Q(L_{\sigma,\mu}(b))/100$.

We proceed to an example of a continuous Gaussian random field, whose tail probability of the supremum is in a closed-form.

Example 2. Consider the Gaussian random field $f(t) = X \cos t + Y \sin t$, , where X and Y are independent standard Gaussian variables and $T = [0, 3/4]$. We let $b = 4$ and consider the class of constant variance and mean functions: $\sigma(t) = \sigma$ and $\mu(t) = \mu$, with $\sigma \in [0.5, 1]$ and $\mu \in [-0.5, 0.5]$.

For constant mean and variance functions considered in this example, the probability $P(\sup_{t\in T}(\sigma f(t) + \mu) > b)$ is known to be in a closed form [3]:

$$P\left(\sup_{0\leq t\leq 3/4}(\sigma f(t) + \mu) > b\right) = \bar{\Phi}((b-\mu)/\sigma) + \frac{3}{8\pi} e^{-(b-\mu)^2/(2\sigma^2)}. \tag{15}$$

The simulation results for Example 2 are summarized in Table 3. Similar to Example 1, we report the estimated probability, the standard deviation of the estimator, and its coefficient of variation. The theoretical value is computed according to (15). We can see that for all combinations of σ and μ in Table 3, the estimated probabilities are close to the theoretical values. We also see that as the probability of interest decrease from 8.18×10^{-6} to 4.01×10^{-12}, the CV of the estimator does not increase substantially (from 2.7 to 6.2). This finding is consistent with our theoretical efficiency analysis of the proposed estimator.

We proceed to examples where the mean and variance functions are not constants. We consider a continuous and centered Gaussian random field $\{f(t) : 0 \leq t \leq 1\}$, whose covariance function is

$$r(s,t) = E(f(s)f(t)) = e^{-|s-t|}. \tag{16}$$

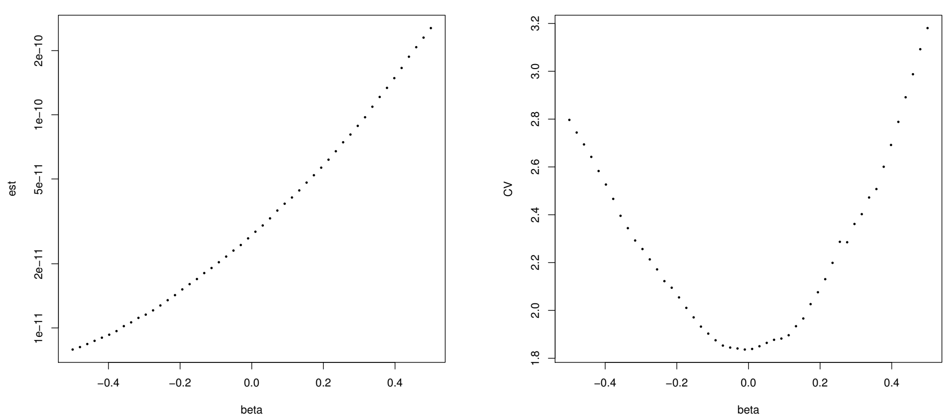
In particular, in Example 3 we consider a Gaussian random field with nonconstant mean and constant variance; in Example 4 we consider a Gaussian field with constant mean and nonconstant variance; and in Example 5, both mean and variance functions are nonconstant.

σ	μ	est.	sd.	CV	theoretical value
0.5	0.5	4.18E-12	2.59E-11	6.2	4.01E-12
0.6	0.3	1.03E-09	4.38E-09	4.2	1.01E-09
0.7	0.1	3.34E-08	1.18E-07	3.5	3.43E-08
0.8	-0.1	3.68E-07	1.19E-06	3.2	3.85E-07
0.9	-0.3	2.10E-06	5.97E-06	2.8	2.20E-06
1	-0.5	8.11E-06	2.20E-05	2.7	8.18E-06

TABLE 3: Simulation result for Example 2 with $b = 4$ and $\delta_b = \frac{1}{b}$. Theoretical values are computed according to (15).

Example 3. Consider the Gaussian random field $f(t)$ defined in (16), and the class of variance and mean functions $\sigma(t) = 1$ and $\mu(t) = \beta_1 t$, for $\beta_1 \in [-0.5, 0.5]$. The probability of interest is $P\left(\sup_{t\in[0,1]} f(t) + \beta_1 t > b\right)$ for $b = 7$.

We summarize the simulation results for Example 3 in Figure 1. Figure 1(a) shows the scatter plot of the estimated probability (y-axis) against β_1 (x-axis). Figure 1(b) shows the scatter plot of the CV (y-axis) against β_1 (x-axis). We see that the probability of interest is an increasing function in β_1. Moreover, when the estimated probability is within the range from 1×10^{-11} to 2×10^{-10}, the CV of the estimator is always controlled within 3.2, showing the good performance of the proposed estimation method.

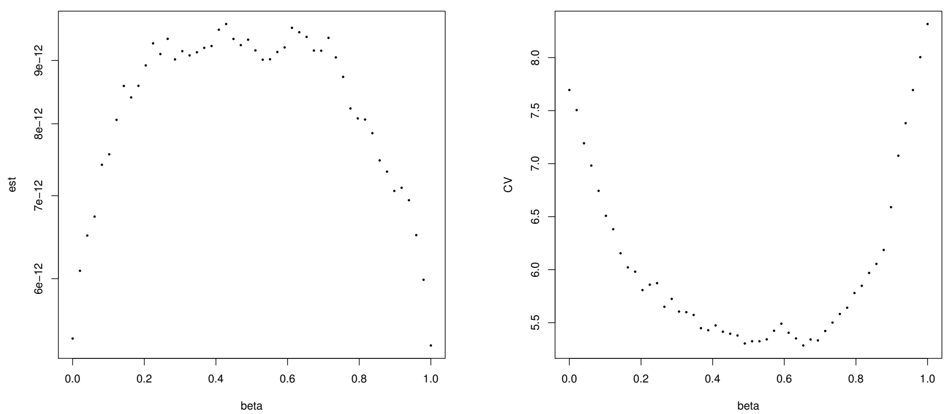


(a) Estimates as a function of β_1 (b) CV as a function of β_1

FIGURE 1: Simulation results for Example 3, where $b = 7$ and $\delta_b = 1/b$.

Example 4. Consider the Gaussian random field $\{f(t), t \in T\}$ defined in (16) and the class of variance and mean functions $\sigma(t) = 1 - 0.5(t - \beta_2)^2$ and $\mu(t) = 0$, where $\beta_2 \in [0, 1]$. The probability of interest is $P\left(\sup_{t\in[0,1]}[1 - 0.5(t - \beta_2)^2]f(t) > b\right)$ for $b = 7$.

For Example 4, the scatter plot of estimated probability and the CV of the estimator are presented in Figure 2. Note that in Example 4, the maximum variance $\max_{t\in T} Var(\sigma(t)f(t)) = Var(\sigma(\beta_2)f(\beta_2)) = 1$. Therefore, for all $\beta_2 \in [0,1]$ the probability of interest has the same exponential decay rate $P(\sup_{t\in[0,1]} \sigma(t)f(t) > b) = e^{-(1+o(1))\frac{b^2}{2\max_{t\in T} Var(\sigma(t)f(t))}} = e^{-(1+o(1))b^2/2}$, as $b \to \infty$. In Figure 2(a), we see that the estimated probability is relatively small when β_2 is close to the boundary values 0 or 1, compared to the case when $\beta_2 \in [0.2, 0.8]$ and is far away from the boundary values. For $\beta_2 \in [0.2, 0.8]$ the estimated probability stays around 9×10^{-12} and does not fluctuate much. For all $\beta_2 \in [0,1]$, the maximum CV of the estimator is controlled within 10. This is again consistent with our theoretical results.



(a) Estimates as a function of β_2 (b) CV as a function of β_2

FIGURE 2: Simulation results for Example 4, where $b = 7$ and $\delta_b = 1/b$.

Example 5. Consider the Gaussian random field $\{f(t), t \in T\}$ defined in (16), and the class of variance and mean functions $\sigma(t) = 1 - 0.5(x - \beta_2)^2$ and $\mu(t) = \beta_1 t$, where $\beta_1 \in [-0.5, 0.5]$ and $\beta_2 \in [0,1]$. The probability of interest is $P(\sup_{t\in[0,1]}\{[1 - 0.5 \times (t-\beta_2)^2]f(t) + \beta_1 t\} > b)$, for $b = 7$.

Table 4 shows the simulated results for different choices of β_1 and β_2. We see that the estimated probabilities range from 4.2×10^{-12} to 1.16×10^{-10}. The maximum CV in Table 4 is 9.9. This means that the standard error of the averaged Monte Carlo estimator with 10^4 samples is controlled within $9.9\% \times E^Q L_{\sigma,\mu}(b)$.

5. Proofs of main results

Throughout the proof, we write $a(b) = O(c(b))$ if there exists a positive constant κ, independent of $b, \sigma(\cdot), \mu(\cdot)$, such that $|a(b)|/|c(b)| \le \kappa$. We also write $a(b) = o(c(b))$ if $|a(b)|/|c(b)| \to 0$ as $b \to \infty$, uniformly in $\sigma(\cdot)$ and $\mu(\cdot)$ satisfying Assumption C1. We will use $\tilde\kappa$ as a generic notation to denote large and not-so-important constants (independent of μ, σ and b) whose value may vary from place to place. Similarly, we use $\tilde\epsilon$ as a generic notation for small positive constants.

Proof of Proposition 1. We start with the proof of Proposition 1 (i). We can see that if $\max_{t_i\in T}\sigma'(t_i)f(t_i) + \mu(t_i) > b$, then $\max_{t_i\in T} \sigma(t_i)f(t_i) + \mu(t_i) > b$ always happens and the change of measure is well defined. We

β_1	β_2	est.	sd.	CV
-0.50	0.00	4.20E-12	4.03E-11	9.6
-0.33	0.17	5.60E-12	3.69E-11	6.6
-0.17	0.33	5.69E-12	3.29E-11	5.8
0.00	0.50	8.78E-12	5.09E-11	5.8
0.17	0.67	2.09E-11	1.27E-10	6.1
0.33	0.83	5.82E-11	4.04E-10	6.9
0.50	1.00	1.16E-10	1.15E-09	9.9

TABLE 4: Simulation results for Example 5, where $b = 7$ and $\delta_b = 2/b$.

have

$$
\begin{aligned}
& E^{Q^\dagger}\left[\left(\frac{dP}{dQ^\dagger}\right)^2; \max\nolimits_{t_i\in T}\sigma'(t_i)f(t_i)+\mu(t_i)>b\right] \\
= \; & E\left[\frac{dQ^\dagger}{dP}\times\left(\frac{dP}{dQ^\dagger}\right)^2; \max\nolimits_{t_i\in T}\sigma'(t_i)f(t_i)+\mu(t_i)>b\right] \\
= \; & E\left[\frac{dP}{dQ^\dagger}; \max\nolimits_{t_i\in T}\sigma'(t_i)f(t_i)+\mu(t_i)>b\right] \\
= \; & E\left[\frac{\sum_{i=1}^M P(\sigma(t_i)f(t_i)+\mu(t_i)>b)}{\sum_{i=1}^M I(\sigma(t_i)f(t_i)+\mu(t_i)>b)}; \max\nolimits_{t_i\in T}\sigma'(t_i)f(t_i)+\mu(t_i)>b\right].
\end{aligned}
$$

Because $\sum_{i=1}^M I(\sigma(t_i)f(t_i)+\mu(t_i)>b)\le M$, the above display is further bounded from below by

$$
\begin{aligned}
= \; & \frac{1}{M}\left(\sum_{i=1}^M P(\sigma(t_i)f(t_i)+\mu(t_i)>b)\right)\times w_{\sigma',\mu}(b) \\
\ge \; & \frac{1}{M}\max_{t_i\in T} P(\sigma(t_i)f(t_i)+\mu(t_i)>b)\times w_{\sigma',\mu}(b) \\
= \; & \exp\left\{-(1+o(1))\frac{b^2}{2\max_{t_i\in T}\sigma(t_i)^2}-(1+o(1))\frac{b^2}{2\max_{t_i\in T}\sigma'(t_i)^2}\right\},
\end{aligned}
$$

where we used the following lemma, whose proof is given in Section 5.1, to obtain that

$$
w_{\sigma',\mu}(b)=\exp\left\{-(1+o(1))\frac{b^2}{2\max_{t_i\in T}\sigma'(t_i)^2}\right\}.
$$

Lemma 1. *Let $\{f(t): t\in T\}$ be a centered, unit variance and continuous Gaussian random field living on a compact set T. Assume that $\sigma(t)>0$ and $\mu(t)$ are continuous functions. Then, there exists positive $\tilde\epsilon$ such that*

$$
P\Big(\sup_{t\in T}\sigma(t)f(t)+\mu(t)>b\Big)=e^{-(1+o(1))\frac{b^2}{2\max_{t\in T}\sigma^2(t)}} \text{ and } P\Big(\sup_{t\in T}\sigma(t)f(t)+\mu(t)>b\Big)\ge\tilde\epsilon b^{-1}\max_{t\in T}e^{-\frac{(b-\mu(t))^2}{2\sigma^2(t)}}.
$$

Under the assumption that $\max_{t_i\in T}\sigma'(t_i)<\max_{t_i\in T}\sigma(t_i)$, we know that for $\varepsilon<\frac{1}{2}(1-\frac{\max\sigma'(t_i)}{\max\sigma(t_i)})$

$$
\frac{E^{Q^\dagger}\left[\left(\frac{dP}{dQ^\dagger}\right)^2; \max_{t_i\in T}\sigma'(t_i)f(t_i)+\mu(t_i)>b\right]}{w_{\sigma',\mu}^{2-\epsilon}(b)}\ge w_{\sigma',\mu}^{-\epsilon}(b),
$$

which tends to infinity as $b \to \infty$.

We proceed to the proof of part (ii). Let $t'_{\max} = \arg\max_{t\in T} \sigma'(t)$. We consider the event $F = \{b/\sigma'(t'_{\max}) < f(t'_{\max}) < \min_{t_i\in T}[b/\sigma(t_i)]\}$. Because $\max_{t_i\in T}\sigma'(t_i) > \max_{t_i\in T}\sigma(t_i)$, F is non-empty and $F \subset \{\max_{t_i\in T}\sigma'(t_i)f(t_i) + \mu'(t_i) > b\}$. Moreover, according to the sampling scheme in Algorithm 1, we have $Q^\dagger(F) > 0$. On the other hand, when the event F happens, $\sum_{i=1}^M I(\sigma(t_i)f(t_i) > b) = 0$, therefore $Q^\dagger(\frac{dP}{dQ^\dagger} = \infty) \geq Q^\dagger(F) > 0$. In other word, $\frac{dP}{dQ^\dagger}$ is not well-defined. $\square$

Proof of Proposition 2. Define the random index $t^* := \arg\max_{t\in T}[\sigma(t)f(t) + \mu(t)]$. We restrict our analysis to the integral over the region $[\mu(t^*), \mu(t^*) + \delta_b] \times [\sigma(t^*), \sigma(t^*) + \delta_b^2]$ and arrive at

$$
\begin{aligned}
&E^Q[L^2_{\sigma,\mu}(b)]\\
=&E^Q\Big[\Big(\int_{\mu_l}^{\mu_u+\delta_b}\int_{\sigma_l}^{\sigma_u+\delta_b^2}\frac{\sum_{i=1}^M I(\varsigma f(t_i)+\nu>b)}{MP(\varsigma f(t_1)+\nu>b)}g(\varsigma)d\varsigma d\nu\Big)^{-2};\max\nolimits_{t_i\in T}\ \sigma(t_i)f(t_i)+\mu(t_i)>b\Big]\\
\leq&E^Q\Big[\Big(\int_{\mu(t^*)}^{\mu(t^*)+\delta_b}\int_{\sigma(t^*)}^{\sigma(t^*)+\delta_b^2}\frac{\sum_{i=1}^M I(\varsigma f(t_i)+\nu>b)}{MP(\varsigma f(t_1)+\nu>b)}g(\varsigma)h(\nu)d\varsigma d\nu\Big)^{-2};\max\nolimits_{t_i\in T}\ \sigma(t_i)f(t_i)+\mu(t_i)>b\Big]\\
=&E^Q\Big[\Big(\int_{\mu(t^*)}^{\mu(t^*)+\delta_b}\int_{\sigma(t^*)}^{\sigma(t^*)+\delta_b^2}\frac{\sum_{i=1}^M I(\varsigma f(t_i)+\nu>b)}{M\bar\Phi(\frac{b-\nu}{\varsigma})}g(\varsigma)h(\nu)d\varsigma d\nu\Big)^{-2};\max\nolimits_{t_i\in T}\ \sigma(t_i)f(t_i)+\mu(t_i)>b\Big]
\end{aligned}
\tag{17}
$$

Note that for all $(\varsigma,\nu) \in [\mu(t^*), \mu(t^*)+\delta_b]\times[\sigma(t^*), \sigma(t^*)+\delta_b^2]$, we have $\varsigma f(t^*)+\nu \geq \max_{t_i\in T}\sigma(t_i)f(t_i)+\mu(t_i)$. Therefore, the event $\max_{t_i\in T}\sigma(t_i)f(t_i)+\mu(t_i) > b$ implies $\varsigma f(t^*)+\nu \geq b$. Consequently, $\sum_{i=1}^M I(\varsigma f(t_i)+\nu > b) \geq 1$ on the event $\max_{t_i\in T}\sigma(t_i)f(t_i) + \mu(t_i) > b$. Therefore, (17) is further bounded from above by

$$
\begin{aligned}
\leq\ & M^2E^Q\Big[\Big(\int_{\mu(t^*)}^{\mu(t^*)+\delta_b}\int_{\sigma(t^*)}^{\sigma(t^*)+\delta_b^2}\frac{g(\varsigma)h(\nu)}{\bar\Phi(\frac{b-\nu}{\varsigma})}d\varsigma d\nu\Big)^{-2};\max\nolimits_{t_i\in T}\sigma(t_i)f(t_i)+\mu(t_i)>b\Big]\\
\leq\ & O(1)M^2E^Q\Big[\Big(\int_{\mu(t^*)}^{\mu(t^*)+\delta_b}\int_{\sigma(t^*)}^{\sigma(t^*)+\delta_b^2}g(\varsigma)h(\nu)be^{\frac{(b-\nu)^2}{2\varsigma^2}}d\varsigma d\nu\Big)^{-2};\max\nolimits_{t_i\in T}\sigma(t_i)f(t_i)+\mu(t_i)>b\Big]
\end{aligned}
\tag{18}
$$

Note that for all $(\varsigma,\nu) \in [\mu(t^*), \mu(t^*)+\delta_b]\times[\sigma(t^*), \sigma(t^*)+\delta_b^2]$, we have $be^{\frac{(b-\nu)^2}{2\varsigma^2}} = O(1)be^{\frac{(b-\mu(t^*))^2}{2\sigma^2(t^*)}}$. Therefore, (18) is bounded from above by

$$
\begin{aligned}
\leq\ & O(1)M^2E^Q\Big[\Big(\int_{\mu(t^*)}^{\mu(t^*)+\delta_b}\int_{\sigma(t^*)}^{\sigma(t^*)+\delta_b^2}g(\varsigma)h(\nu)be^{\frac{(b-\mu(t^*))^2}{2\sigma^2(t^*)}}d\varsigma d\nu\Big)^{-2};\max\nolimits_{t_i\in T}\sigma(t_i)f(t_i)+\mu(t_i)>b\Big]\\
=\ & O(1)M^2\delta_b^{-6}b^{-2}E^Q\Big[e^{-\frac{(b-\mu(t^*))^2}{\sigma^2(t^*)}};\max\nolimits_{t_i\in T}\sigma(t_i)f(t_i)+\mu(t_i)>b\Big]\\
\leq\ & O(1)M^2\delta_b^{-6}b^{-2}\max_{t_i\in T}e^{-\frac{(b-\mu(t_i))^2}{\sigma^2(t_i)}}.
\end{aligned}
\tag{19}
$$

On the other hand, according to Lemma 1 we have

$$
P\Big(\sup_{t_i\in T}\sigma(t_i)f(t_i)+\mu(t_{t_i})>b\Big)\geq\tilde\epsilon b^{-1}\max_{t_i\in T}e^{-\frac{b-\mu(t_i)}{2\sigma^2(t_i)}}.
$$

Combining this and (19), we have that there exists b_0 sufficiently large such that for $b \geq b_0$

$$\frac{E^Q[L^2_{\sigma,\mu}(b); \max_{t_i \in T} \sigma(t_i) f(t_i) + \mu(t_i) > b]}{M^2 b^6 w^2_{\sigma,\mu}(b)} = O(1).$$

This completes our proof. $\square$

Proof of Theorem 1. Note that $\sup_{t\in T} \sigma(t) f(t) + \mu(t) \geq \sup_{t \in T_N} \sigma(t) f(t) + \mu(t)$, we have

$$\begin{aligned}
& \Big| P\Big(\sup_{t\in T} \sigma(t) f(t) + \mu(t) > b\Big) - P\Big(\sup_{t\in T_N} \sigma(t) f(t) + \mu(t) > b\Big)\Big| \\
= \; & P\Big(\sup_{t\in T} \sigma(t) f(t) + \mu(t) > b, \sup_{t\in T_N} \sigma(t) f(t) + \mu(t) \leq b\Big).
\end{aligned}$$

We split the above probability into two parts.

$$\begin{aligned}
& P\Big(\sup_{t\in T} \sigma(t) f(t) + \mu(t) > b, \sup_{t\in T_N} \sigma(t) f(t) + \mu(t) \leq b\Big) \\
= \; & P\Big(b < \sup_{t\in T} \sigma(t) f(t) + \mu(t) \leq b + \frac{\gamma}{b}, \sup_{t\in T_N} \sigma(t) f(t) + \mu(t) \leq b\Big) \\
& + P\Big(\sup_{t\in T} \sigma(t) f(t) + \mu(t) > b + \frac{\gamma}{b}, \sup_{t\in T_N} \sigma(t) f(t) + \mu(t) \leq b\Big),
\end{aligned}$$

which is further bounded from above by

$$P\Big(b < \sup_{t\in T} \sigma(t) f(t) + \mu(t) \leq b + \frac{\gamma}{b}\Big) + P\Big(\sup_{t\in T} \sigma(t) f(t) + \mu(t) > b + \frac{\gamma}{b}, \sup_{t\in T_N} \sigma(t) f(t) + \mu(t) \leq b\Big), \tag{20}$$

where we will choose γ later. We proceed to upper bounds of the above two terms separately. For the first term, we apply the following Lemma.

Lemma 2. (Proposition 6.5 of [2].) *Under Assumptions C1,C3 and C4, for any $v > 0$, let $\beta^* = \min(\beta, \beta')$ and $\rho = \frac{2d}{\beta^*} + dv + 1$, where d is the dimension of T. There exists constants $b_0, \lambda \in (0, \infty)$ so that for all $b \geq b_0 \geq 1$,*

$$P\Big(\max_{t\in T} \sigma(t) f(t) + \mu(t) \leq b + \frac{\gamma}{b} \;\Big|\; \max_{t\in T} \sigma(t) f(t) + \mu(t) > b\Big) \leq \lambda a b^{\rho}. \tag{21}$$

With the aid of the above lemma with $v = \frac{1}{\beta^*}$, we have for $b \geq b_0$

$$\begin{aligned}
& P\Big(b < \sup_{t\in T} \sigma(t) f(t) + \mu(t) \leq b + \frac{\gamma}{b}\Big) \\
= \; & P\Big(\max_{t\in T} \sigma(t) f(t) + \mu(t) > b\Big) P\Big(\max_{t\in T} \sigma(t) f(t) + \mu(t) \leq b + \frac{\gamma}{b} \;\Big|\; \max_{t\in T} \sigma(t) f(t) + \mu(t) > b\Big) \\
\leq \; & \lambda \gamma b^{\rho} P\Big(\max_{t\in T} \sigma(t) f(t) + \mu(t) > b\Big)
\end{aligned}$$

with $\rho = \frac{3d}{\beta^*} + 1$. We choose $\gamma := 2^{-1}\lambda^{-1} b^{-\rho} \varepsilon$, then the above display gives the following upper bound for the first term in (20)

$$P\Big(b < \sup_{t\in T} \sigma(t) f(t) + \mu(t) \leq b + \frac{\gamma}{b}\Big) \leq \frac{\varepsilon}{2} w_{\sigma,\mu}(b).$$

We proceed to the second term in (20). According to Assumption C2, we have

$$\begin{aligned} & P\Big(\sup_{t\in T}\sigma(t)f(t)+\mu(t)>b+\frac{\gamma}{b},\sup_{t\in T_N}\sigma(t)f(t)+\mu(t)\leq b\Big)\\ \leq\ & P\Big(\sup_{t,s\in T,|t-s|\leq\kappa_m/N}|\sigma(t)f(t)+\mu(t)-(\sigma(s)f(s)+\mu(s))|>\frac{\gamma}{b}\Big), \end{aligned}$$

which is further bounded from above by

$$P\Big(\sup_{t,s\in T,|t-s|\leq\kappa_m/N}|\sigma(t)f(t)-\sigma(s)f(s)|+\sup_{t,s\in T,|t-s|\leq\kappa_m/N}|\mu(t)-\mu(s)|>\frac{\gamma}{b}\Big). \tag{22}$$

According to Assumption C1, we have

$$\sup_{t,s\in T,|t-s|\leq\kappa_m/N}|\mu(t)-\mu(s)|=O(\kappa_m^{\beta^*}/N^{\beta^*}).$$

Plugging this into (22), we have

$$\begin{aligned} & P\Big(\sup_{t,s\in T,|t-s|\leq\kappa_m/N}|\sigma(t)f(t)+\mu(t)-\sigma(s)f(s)+\mu(s)|>\frac{\gamma}{b}\Big)\\ \leq & P\Big(\sup_{t,s\in T,|t-s|\leq\kappa_m/N}|\sigma(t)f(t)-\sigma(s)f(s)|>\frac{\gamma}{b}-\kappa_m^{\beta^*}/N^{\beta^*}\Big). \end{aligned} \tag{23}$$

We choose $N\geq\tilde{\kappa}\lambda^{1/\beta^*}b^{(\rho+1)/\beta^*}\varepsilon^{-1/\beta^*}$ for $\tilde{\kappa}$ sufficiently large, then $\frac{\gamma}{b}-\kappa_m^{\beta^*}\frac{1}{N^{\beta^*}}>\frac{\gamma}{2b}$. Therefore, we further have

$$\begin{aligned} & P\Big(\sup_{t,s\in T,|t-s|\leq\kappa_m/N}|\sigma(t)f(t)+\mu(t)-\sigma(s)f(s)+\mu(s)|>\frac{\gamma}{b}\Big)\\ \leq & P\Big(\sup_{t,s\in T,|t-s|\leq\kappa_m/N}|\sigma(t)f(t)-\sigma(s)f(s)|>\frac{\gamma}{2b}\Big). \end{aligned} \tag{24}$$

To control the above probability, we use the following lemma known as the Borell-TIS lemma, which is proved independently by [10] and [33].

Lemma 3. (Borell-TIS.) *Let $\{f(t);t\in\mathcal{U}\}$, where $\mathcal{U}$ is a compact set, be a mean zero Gaussian random field. f is almost surely bounded on $\mathcal{U}$. Then, $E[\sup_{\mathcal{U}}f(t)]<\infty$, and $P\left(\sup_{t\in\mathcal{U}}f(t)-E[\sup_{t\in\mathcal{U}}f(t)]\geq b\right)\leq\exp\left(-\frac{b^2}{2\sigma_{\mathcal{U}}^2}\right)$, where $\sigma_{\mathcal{U}}^2=\sup_{t\in\mathcal{U}}Var[f(t)]$.*

We define a new Gaussian random field

$$\xi(s,t)=\sigma(s)f(s)-\sigma(t)f(t). \tag{25}$$

The next lemma, whose proof will be provided in Section 5.1, characterizes $E\sup_{t,s\in T,|t-s|\leq\kappa_m/N}\xi(s,t)$.

Lemma 4. *For all σ,μ and f satisfying Assumptions C1, C3 and C4, there is a uniform constant $\kappa_\xi>0$ such that*

$$E\sup_{t,s\in T,|t-s|\leq\kappa_m/N}|\xi(s,t)|<\kappa_\xi N^{-\beta^*/2}\log N$$

Furthermore, the variance of $\xi(s,t)$ is bounded from above by

$$Var(\xi(s,t)) = (\sigma(s)-\sigma(t))^2 + 2\sigma(s)\sigma(t)(1-r(s,t)) \leq \kappa_H^2|s-t|^{2\beta^*} + 2\sigma_u^2|s-t|^{\beta^*} \leq O(|s-t|^{\beta^*}). \quad (26)$$

According to Assumption C1 and C4, the above display is further bounded from above by

$$Var(\xi(s,t)) \leq O(N^{-\beta^*}) \quad (27)$$

We choose N such that $\kappa_\xi N^{-\frac{\beta^*}{2}} \log N \leq \frac{\gamma}{4b}$. Then according to the Borell-TIS lemma and Lemma 4, we have

$$P\Big(\sup_{|t-s|\leq \frac{\kappa_m}{N}} |\xi(s,t)| > \frac{\gamma}{4b}\Big) \leq \exp\left(-\tilde{\epsilon}\frac{\gamma^2}{N^{-\beta^*}b^2}\right). \quad (28)$$

The above display is of order $o(\varepsilon w_{\sigma,\mu}(b))$ if $\frac{\gamma^2}{N^{-\beta^*}b^2} \geq \tilde{\kappa}^{\beta^*}\max(-\log\varepsilon, b^2)$, for a large enough and possibly different constant $\tilde{\kappa}$. Therefore, it is sufficient to choose $N \geq \tilde{\kappa}\max(-\log\varepsilon, b^2)^{1/\beta^*} b^{2/\beta^*}\gamma^{-2/\beta^*}(\log b)^{\tilde{\kappa}}$. Combining this with our choice of γ, and recall our choice of ρ in Lemma 2 it is sufficient to choose $N \geq \tilde{\kappa}\max(-\log\varepsilon, b^2)^{1/\beta^*} b^{2/\beta^*+2/\beta^*(\frac{3d}{\beta^*}+1)}\varepsilon^{-2/\beta^*}(\log b)^{\tilde{\kappa}}$, which is bounded by $N_0 = b^{2/\beta^*(\frac{3d}{\beta^*}+2+\varepsilon_0)}\varepsilon^{-2/\beta^*-\varepsilon_0}$ for any $\varepsilon_0 > 0$ and b sufficiently large. This completes our proof. $\square$

Proof of Theorem 2. According to Proposition 2 with $M = O(N^d)$, we have

$$E^Q[L^2_{\sigma,\mu,N}(b)] = O(1)N^{2d}\delta_b^{-6}w^2_{\sigma,\mu,N}(b).$$

According to the choice of N_0 in Theorem 1, we have

$$E^Q[L^2_{\sigma,\mu,N}(b)] = O(1)b^{4d/\beta^*(\frac{3d}{\beta^*}+2+\varepsilon_0)+6}\varepsilon^{-4d/\beta^*-2d\varepsilon_0}w^2_{\sigma,\mu,N}(b)$$

uniformly for $\mu,\sigma \in \mathcal{C}(\mu_l,\mu_u,\sigma_l,\sigma_u,\beta,\kappa_H)$. This completes our proof. $\square$

Proof of Corollary 1. The mean squared error of $Z_{\sigma,\mu,N}(b)$ is decomposed as the sum of its bias and variance,

$$E[Z_{\sigma,\mu,N}(b)-w_{\sigma,\mu}(b)]^2 = [EZ_{\sigma,\mu,N}(b)-w_{\sigma,\mu}(b)]^2+Var(Z_{\sigma,\mu,N}(b)) = [w_{\sigma,\mu,N}(b)-w_{\sigma,\mu}(b)]^2+Var(L_{\sigma,\mu,N}(b))/n.$$

Setting $\varepsilon := \varepsilon\delta^{1/2}$ in Theorem 1, we have $[w_{\sigma,\mu,N}(b) - w_{\sigma,\mu}(b)]^2 < \varepsilon^2\delta w^2_{\sigma,\mu}(b)/2$ for $N \geq N(\varepsilon\delta^{1/2}, b)$. Furthermore, according to Theorem 2, we have $Var(L_{\sigma,\mu,N}(b))/n \leq \varepsilon^2\delta w^2_{\sigma,\mu}(b)/2$ for $n \geq 2\kappa_c b^q\varepsilon^{-q_1-2}\delta^{-\frac{q_1}{2}-1}$. Consequently, for such N and n we have $E[Z_{\sigma,\mu,N}(b) - w_{\sigma,\mu}(b)]^2 \leq \varepsilon^2\delta$. Thanks to Chebyshev's inequality, we have

$$P(|Z_{\sigma,\mu,N}(b) - w_{\sigma,\mu}(b)| > \varepsilon) < \frac{E[Z_{\sigma,\mu,N}(b) - w_{\sigma,\mu}(b)]^2}{\varepsilon^2} \leq \delta.$$

Therefore, $Z_{\sigma,\mu,N}(b)$ satisfies (14). $\square$

5.1. Proofs of supporting lemmas

Proof of Lemma 1. First, according to Lemma 3, we have

$$P\Big(\sup_{t\in T}\sigma(t)f(t)+\mu(t)>b\Big)\le P\left(\sup_{t\in T}\sigma(t)f(t)>b-\max_{t\in T}\mu(t)\right)\le e^{-(1+o(1))\frac{b^2}{2\max_{t\in T}\sigma^2(t)}}. \tag{29}$$

On the other hand, for each $t\in T$ we have

$$P\Big(\sup_{t\in T}\sigma(t)f(t)+\mu(t)>b\Big)\ge P\Big(\sigma(t)f(t)+\mu(t)>b\Big)=P\Big(f(t)>\frac{b-\mu(t)}{\sigma(t)}\Big),$$

which is further bounded from below by

$$P\Big(\sup_{t\in T}\sigma(t)f(t)+\mu(t)>b\Big)\ge\frac{1}{\sqrt{2\pi}\sigma(t)}\Big(\frac{\sigma(t)}{b-\mu(t)}-\frac{\sigma^3(t)}{(b-\mu(t))^3}\Big)e^{-\frac{(b-\mu(t))^2}{2\sigma^2(t)}}=\tilde{\epsilon}b^{-1}e^{-\frac{b^2}{2\sigma^2(t)}}.$$

To obtain the last equation in the above display, we used the fact that $\mu(t)\in[\mu_l,\mu_u]$ and $\sigma(t)\in[\sigma_l,\sigma_u]$ with $\sigma_l>0$. Taking the maximum of the right-hand side of the above display, we have

$$P\Big(\sup_{t\in T}\sigma(t)f(t)+\mu(t)>b\Big)\ge\tilde{\epsilon}b^{-1}\max_{t\in T}e^{-\frac{(b-\mu(t))^2}{2\sigma^2(t)}}. \tag{30}$$

Combining the above expression with (29), we complete the proof. □

Proof of Lemma 4. To prove this lemma, we will need the following entropy bound ([14]).

Lemma 5. *Let f be a centered Gaussian field living on a metric space $\mathcal{U}$. Define the pseudo-metric*

$$d_f(s,t)=\sqrt{E(f(s)-f(t))^2}.$$

Assume that $\mathcal{U}$ is a compact space under the metric d_f and for each $\varepsilon>0$. Denote by $N(\varepsilon)$ the smallest number of balls with radius ε under the metric d_f. Then there exists a universal constant K such that

$$E\Big[\sup_{t\in\mathcal{U}}f(t)\Big]\le K\int_0^{diam(\mathcal{U})}(\log N(\varepsilon))^{\frac{1}{2}}d\varepsilon. \tag{31}$$

Let $\mathcal{U}=\{(s,t):s,t\in T,|s-t|\le\kappa_m\frac{1}{N}\}$ and

$$d_\xi^2((s,t),(s',t'))=E[\xi(s,t)-\xi(s',t')]^2=E[\xi(s,s')-\xi(t,t')]^2.$$

We first investigate the metric d_ξ. We have

$$d_\xi^2((s,t),(s',t'))\le 2Var(\xi(s,s'))+2Var(\xi(t,t')). \tag{32}$$

Applying (26) to the above display, we have that there is a $\tilde{\kappa}$ uniformly for all σ,μ satisfying Assumption C1, such that

$$d_\xi((s,t),(s',t'))\le\tilde{\kappa}\sqrt{|s-s'|^{\beta^*}+|t-t'|^{\beta^*}}. \tag{33}$$

According to the relationship between the l_p norms, we have $(|s-s'|^{\beta^*}+|t-t'|^{\beta^*})^{\frac{1}{\beta^*}}\le d^{\frac{1}{2}-\frac{1}{\beta^*}}\sqrt{|s-s'|^2+|t-t'|^2}$. The result, together with (33), implies that $B((s,t),\tilde{\epsilon}\varepsilon^{\frac{2}{\beta^*}})\subset B_{d_\xi}((s,t),\varepsilon)$ for some constant $\tilde{\epsilon}$ that only depends on d, β^* and $\tilde{\kappa}$, where B and B_ξ denote balls under the Euclidean norm and d_ξ metrics respectively. Note that the set $T \times T$ can be covered by $\tilde{\kappa}\varepsilon^{-\frac{4d}{\beta^*}}$ many $B(\tilde{\epsilon}\varepsilon^{\frac{2}{\beta^*}})$ balls with a possibly different $\tilde{\kappa}$. Consequently, the set $\mathcal{U}$ can be covered by the same number of $B_{d_\xi}(\varepsilon)$ balls. Therefore, we have

$$\log(N(\varepsilon)) \le \log \tilde{\kappa} + \frac{4d}{\beta^*} \log \varepsilon^{-1}$$

On the other hand, we have $d_\xi((s,t),(s',t')) \le 2Var(\xi(s,t)) + 2Var(\xi(s',t'))$. Also according to (26), we have $d_\xi^2((s,t),(s',t')) = O(|s-t|^{\beta^*} + |s'-t'|^{\beta^*})$. Therefore, for $|s-t| \le \kappa_m/N$ we have $d_\xi((s,t),(s',t')) = O(N^{-\beta^*/2})$. Consequently, $diam(\mathcal{U}) \le \tilde{\kappa}N^{-\beta^*/2}$. According to Lemma 5, we have

$$E \sup_{t,s\in T, |t-s|\le \kappa_m \frac{1}{N}} \le \tilde{\kappa}(\frac{4d}{\beta^*})^{1/2} \int_0^{\tilde{\kappa}N^{-\beta^*/2}} (\log \varepsilon^{-1})^{1/2} d\varepsilon = O(N^{-\beta^*/2} \log N).$$

This completes our proof. $\square$

Acknowledgement

The authors thank the Editor, an associate Editor and two referees for their constructive comments. Li's research is partially supported by the National Science Foundation grant DMS-1712657. Xu's research is partially supported by the National Science Foundation grants DMS-1712717 and SES-1659328, and National Security Agency grant H98230-17-1-0308.

References

[1] R. Adler, J. Blanchet, and J. Liu. Efficient simulation for tail probabilities of Gaussian random fields. In *Proceedings of the 2008 Winter Simulation Conference*, 2008.

[2] R. Adler, J. Blanchet, and J. Liu. Efficient Monte Carlo for large excursions of Gaussian random fields. *The Annals of Applied Probability*, 22(3):1167–1214, 2012.

[3] R. Adler and J. Taylor. *Random fields and geometry.* Springer, New York, 2007.

[4] R. J. Adler, P. Müller, and B. Rozovskii. *Stochastic modelling in physical oceanography.* Birkhaüser, Boston, 1996.

[5] R. J. Adler, J. E. Taylor, and K. J. Worsley. *Applications of random fields and geometry: Foundations and case studies.* available at http://webee.technion.ac.il/people/adler/hrf.pdf, 2010.

[6] S. Asmussen and P. Glynn. *Stochastic Simulation: Algorithms and Analysis.* Springer, New York, NY, USA, 2007.

[7] J. M. Azais and M. Wschebor. A general expression for the distribution of the maximum of a Gaussian field and the approximation of the tail. *Stochastic Processes and Their Applications*, 118(7):1190–1218, 2008.

[8] J.-M. Azaïs and M. Wschebor. *Level sets and extrema of random processes and fields*. John Wiley & Sons, 2009.

[9] S. M. Berman. An asymptotic formula for the distribution of the maximum of a Gaussian process with stationary increments. *Journal of Applied Probability*, 22(2):454–460, 1985.

[10] C. Borell. The Brunn-Minkowski inequality in Gauss space. *Inventiones Mathematicae*, 30(2):207–216, 1975.

[11] C. Borell. The Ehrhard inequality. *Comptes Rendus Mathematique*, 337(10):663–666, 2003.

[12] D. Cheng and Y. Xiao. The mean Euler characteristic and excursion probability of Gaussian random fields with stationary increments. *The Annals of Applied Probability*, 26(2):722–759, 2016.

[13] K. Dębicki, E. Hashorva, and L. Ji. Extremes of a class of non-homogeneous Gaussian random fields. *The Annals of Probability*, 44(2):984–1012, 2016.

[14] R. M. Dudley. Sample functions of the Gaussian process. *The Annals of Probability*, 1:66–103, 1973.

[15] P. Glasserman. *Monte Carlo methods in financial engineering*, volume 53. Springer Science & Business Media, 2013.

[16] P. Glasserman and S. Juneja. Uniformly efficient importance sampling for the tail distribution of sums of random variables. *Mathematics of Operations Research*, 33(1):36–50, 2008.

[17] S. Juneja and P. Shahabuddin. Rare-event simulation techniques: an introduction and recent advances. *Handbooks in operations research and management science*, 13:291–350, 2006.

[18] H. J. Landau and L. A. Shepp. Supremum of a Gaussian process. *Sankhya-the Indian Journal of Statistics Series A*, 32(Dec):369–378, 1970.

[19] M. Ledoux and M. Talagrand. *Probability in Banach spaces : isoperimetry and processes*. Springer-Verlag, 1991.

[20] X. Li and J. Liu. Rare-event simulation and efficient discretization for the supremum of Gaussian random fields. *Advances in Applied Probability*, 47(3):787–816, 2015.

[21] X. Li, J. Liu, and G. Xu. On the tail asymptotics of exponential integrals of Gaussian random fields with small noise. *Mathematics of Operations Research*, 41:236–246, 2016.

[22] X. Li, J. Liu, and Z. Ying. Chernoff index for cox test of separate parametric families. *arXiv preprint arXiv:1606.08248*, 2016.

[23] J. Liu. Tail approximations of integrals of Gaussian random fields. *The Annals of Probability*, 40(3):1069–1104, 2012.

[24] J. Liu and G. Xu. Rare-event simulations for exponential integrals of smooth Gaussian processes. In *Proceedings of the Winter Simulation Conference*. Winter Simulation Conference, 2012.

[25] J. Liu and G. Xu. Some asymptotic results of Gaussian random fields with varying mean functions and the associated processes. *The Annals of Statistics*, 40:262–293, 2012.

[26] J. Liu and G. Xu. On the density functions of integrals of Gaussian random fields. *Advances in Applied Probability*, 45:398–424, 2013.

[27] J. Liu and G. Xu. Efficient simulations for the exponential integrals of Hölder continuous Gaussian random fields. *The ACM Transactions on Modeling and Computer Simulation*, 24:9:1–9:24, 2014.

[28] J. Liu and G. Xu. On the conditional distributions and the efficient simulations of exponential integrals of Gaussian random fields. *The Annals of Applied Probability*, 24:1691–1738, 2014.

[29] M. B. Marcus and L. A. Shepp. Continuity of Gaussian processes. *Transactions of the American Mathematical Society*, 151(2), 1970.

[30] V. Sudakov and B. Tsirelson. Extremal properties of half spaces for spherically invariant measures. *Zap. Nauchn. Sem. LOMI*, 45:75–82, 1974.

[31] J. Y. Sun. Tail probabilities of the maxima of Gaussian random-fields. *The Annals of Probability*, 21(1):34–71, 1993.

[32] M. Talagrand. Majorizing measures: The generic chaining. *The Annals of Probability*, 24(3):1049–1103, 1996.

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

[34] G. Xu. Uniformly efficient simulation for tail probabilities of Gaussian random fields. In *Proceedings of the 2014 Winter Simulation Conference*, 2014.