J. Appl. Prob. **55,** 157–178 (2018)
*doi:*10.1017/jpr.2018.11

UNIFORMLY EFFICIENT SIMULATION FOR EXTREMES OF GAUSSIAN RANDOM FIELDS

XIAOOU LI,* *University of Minnesota*
GONGJUN XU,** *University of Michigan*

Abstract

In this paper we consider 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 problems. 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 $T \subset \mathbb{R}^d$; that is, for every finite subset of $\{t_1, \ldots, t_n\} \subset T$, $(f(t_1), \ldots, f(t_n))$ is a multivariate Gaussian random vector with $\mathbb{E} f(t_i) = 0$ and $\operatorname{var}(f(t_i)) = 1$ for $i = 1, \ldots, n$. We are interested in estimating the tail probability

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

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 known only to be in certain ranges.

The extremes of Gaussian random fields have wide applications in finance, spatial analysis, physical oceanography, and many other disciplines; see [4] and [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 suprema; see, e.g. [1], [7], [9]–[13], [18], [19], and [29]–[33]. Tail probabilities of other convex functions of Gaussian random fields have also been studied; see [21], [23], [25], and [28],

Most of the sharp theoretical approximations developed in the literature require the evaluation of certain constants that are hard to estimate, such as 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

Received 23 January 2017; revision received 29 August 2017.
* Postal address: School of Statistics, University of Minnesota, 224 Church ST SE, Minneapolis, MN 55455, USA. Email address: lixx1766@umn.edu
** Postal address: Department of Statistics, University of Michigan, 1085 South University, Ann Arbor, MI 48109, USA. Email address: gongjun@umich.edu

it is often unclear how large the tail values need to be to ensure the approximations are 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 [2] and [3]; see also [8], [20], [24], [26]–[28], and [34] for related studies.

To design an asymptotically efficient importance sampling estimator, one needs to construct a change of measure that is tailored to a specific event. Such a construction usually requires detailed information of the Gaussian random fields, such as $\mu(t)$ and $\sigma(t)$ where computation is 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 overheads, especially at the exploratory stage when one often needs to tune different model parameters. This motivates us to find 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 may help to 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 the 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 $\mu_l \le \mu_u \in \mathbb{R}$ and $\sigma_l \le \sigma_u \in (0, \infty)$ are constants that are prespecified. We propose a mixture type change of measure that yields a 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. In Section 3 we provide the construction of our importance sampling estimator and present 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 0, 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 (see, e.g. [3], [6], and [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 order* q in estimating $w_{\sigma,\mu}(b)$ if $\mathbb{E}L(b) = w_{\sigma,\mu}(b)$ and there exist constants $q \geq 0$ and $b_0 \geq 0$ such that

$$\sup_{b \geq b_0} \frac{\operatorname{var}(L(b))}{|\log w_{\sigma,\mu}(b)|^q w_{\sigma,\mu}^2(b)} < \infty.$$

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 $\varepsilon, \delta > 0$,

$$\mathbb{P}\bigg(\bigg|\frac{Z(b)}{w_{\sigma,\mu}(b)} - 1\bigg| > \varepsilon\bigg) < \delta. \tag{1}$$

If a standard Monte Carlo simulation method is used, then it requires at least $n = O(\varepsilon^{-2}\delta^{-1} w_{\sigma,\mu}^{-1}(b))$ independent and identically distributed (i.i.d.) replicates, according to the central limit theorem. By the Borell–Tsirelson–Ibragimov–Sudakov (Borell–TIS) lemma (Lemma 3), we know that $w_{\sigma,\mu}(b) \leq \exp\{-(1 + o(1))b^2/(2\sup_{t\in T}\sigma^2(t))\}$. 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 $\{L^{(j)}(b) \colon j = 1, \ldots, n\}$ be n i.i.d. copies of $L(b)$. Then the averaged estimator $Z(b) = (1/n)\sum_{j=1}^{n} L^{(j)}(b)$ has a mean-squared error (MSE) $\mathbb{E}[Z(b) - w_{\sigma,\mu}(b)]^2 = \operatorname{var}(L(b))/n$. A direct application of Chebyshev's inequality yields

$$\mathbb{P}\bigg(\bigg|\frac{Z(b)}{w_{\sigma,\mu}(b)} - 1\bigg| \geq \varepsilon\bigg) \leq \frac{\operatorname{var}(L(b))}{n\varepsilon^2 w_{\sigma,\mu}^2(b)}. \tag{2}$$

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 (1). 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 that of *weak efficiency* [6]. An estimator $L(b)$ is said to be weakly efficient in estimating $w_{\sigma,\mu}(b)$, if $\mathbb{E}L(b) = w_{\sigma,\mu}(b)$ and, for all positive constants $\varepsilon > 0$,

$$\limsup_{b\to\infty} \frac{\operatorname{var}(L(b))}{w_{\sigma,\mu}^{2-\varepsilon}(b)} = 0.$$

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

$$\begin{aligned} w_{\sigma,\mu}(b) &= \mathbb{E}[\mathbf{1}(\sup\nolimits_{t\in T}\{\sigma(t)f(t) + \mu(t)\} > b)] \\ &= \mathbb{E}^{\mathbb{Q}}\bigg[\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\mathbf{1}(\sup\nolimits_{t\in T}\{\sigma(t)f(t) + \mu(t)\} > b)\bigg], \end{aligned}$$

where $\mathbf{1}(\cdot)$ denotes the indicator function, $\mathbb{Q}$ is a probability measure that is absolutely continuous with respect to $\mathbb{P}$ on the set $\{\sup_{t\in T}\{\sigma(t)f(t) + \mu(t)\} > b\}$, and we use $\mathbb{E}$ and $\mathbb{E}^{\mathbb{Q}}$to

https://doi.org/10.1017/jpr.2018.11 Published online by Cambridge University Press

denote the expectations under the measures $\mathbb{P}$ and $\mathbb{Q}$, respectively. Then the random variable defined by

$$L_{\sigma,\mu}(b) = \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} \mathbf{1}(\sup_{t\in T}\{\sigma(t)f(t)+\mu(t)\} > b)$$

is an unbiased estimator of $w_{\sigma,\mu}(b)$ under the measure $\mathbb{Q}$. To have an efficient estimator, we want to choose $\mathbb{Q}$ such that the variance $\operatorname{var}^{\mathbb{Q}}(L_{\sigma,\mu}(b))$ is small. It is straightforward to show that the optimal change of measure is the conditional probability $\mathbb{Q}^*(\cdot) := \mathbb{P}(\cdot \mid \sup_{t\in T}\{\sigma(t)f(t)+\mu(t)\} > b) = \mathbb{P}(\cdot \cap \{\sup_{t\in T}\{\sigma(t)f(t)+\mu(t)\} > b\})/w_{\sigma,\mu}(b)$, for which the corresponding importance sampling estimator has a zero variance. However, $\mathbb{Q}^*$ 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 $\mathbb{Q}^*$.

2.2. Nonuniformly efficient issue and an example

Various importance sampling estimators for rare-event analysis of the suprema of Gaussian random fields have been studied in [2], [3], [8], and [20]. As the measure $\mathbb{Q}^*$ 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 $\mathbb{Q}$ that gives an efficient estimator $L_{\sigma,\mu}(b) = (\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q})\,\mathbf{1}(\sup_{t\in T}\{\sigma(t)f(t)+\mu(t)\} > b)$ for $w_{\sigma,\mu}(b)$ may no longer be efficient 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 $\mathbb{Q}$,

$$L_{\sigma',\mu'}(b) := \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} \mathbf{1}(\sup_{t\in T}\{\sigma'(t)f(t)+\mu'(t)\} > b),$$

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

To illustrate the nonuniform efficiency issue, we take the estimator proposed in [3] as an example. For simplicity, we consider the case when T contains finite points and write $T := \{t_1,\dots,t_M\}$.

For known μ and σ, Adler *et al.* [3] proposed the following simulation procedure.

Algorithm 1. (*Sampling procedure proposed by Adler et al. [3].*) The algorithm proceeds as follows.

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

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

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

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 $\{f(t)\colon t \neq \tau,\ t \in T\}$ from the original conditional distribution under $\mathbb{P}$.

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

Let $\mathbb{Q}^\dagger$ be the corresponding change of measure. We have

$$\frac{\mathrm{d}\mathbb{Q}^\dagger}{\mathrm{d}\mathbb{P}} = \frac{\sum_{i=1}^{M}\mathbf{1}(\sigma(t_i)f(t_i)+\mu(t_i) > b)}{\sum_{i=1}^{M}\mathbb{P}(\sigma(t_i)f(t_i)+\mu(t_i) > b)}.$$

https://doi.org/10.1017/jpr.2018.11 Published online by Cambridge University Press

Adler *et al.* [3] showed that $L_{\sigma,\mu}(b) = (\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}^{\dagger})\,\mathbf{1}(\sup_{t\in T}\{\sigma(t)f(t)+\mu(t)\} > b)$ is a polynomially efficient estimator for $w_{\sigma,\mu}(b)$ with order $q = 0$. We explain intuitively why this estimator is efficient. First, Algorithm 1 samples a 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 $\mathbb{P}(f(t^*) \in \cdot \mid f(t^*) > b)$. Finally, Algorithm 1 simulates the $f(t)$ at $t \neq \tau$ according to the original conditional distribution given $(f(\tau), \tau)$. Combining these three steps, the entire sample path $\{f(t)\colon t \in T\}$ generated from Algorithm 1 approximately follows the conditional distribution $\{f(t)\colon t \in T \mid \max_{t_i}(\sigma(t_i)f(t_i)+\mu(t_i)) > b\}$. According to the earlier discussion, this conditional probability measure is the optimal change of measure. See [3] for rigorous justifications of the above statements.

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

$$L_{\sigma',\mu'}(b) := \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}^{\dagger}}\,\mathbf{1}(\sup_{t\in T}\{\sigma'(t)f(t)+\mu'(t)\} > b).$$

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

(i) *If $\sigma'(t) \le \sigma(t)$ 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 $\varepsilon > 0$,*

$$\lim_{b\to\infty}\frac{\mathbb{E}^{\mathbb{Q}^{\dagger}}[(\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}^{\dagger})^2;\ \max_{t_i\in T}\sigma'(t_i)f(t_i) > b]}{w^{2-\varepsilon}_{\sigma',\mu}(b)} = \infty.$$

(ii) *If $\max_{t_i\in T}\sigma'(t_i) > \max_{t_i\in T}\sigma(t_i)$ then $\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}^{\dagger}$ is not well defined on the event*

$$\{\max_{t_i\in T}\sigma'(t_i)f(t_i) > b\}.$$

According to the definition of the weakly efficient estimator in Remark 1, Proposition 1(i) 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. Proposition 1(ii) 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 $\{f(t),\ t = 1,\dots,100\}$. For simplicity, we take $\mu(t) = 0$ and $\sigma(t) = \sigma$ for all t. We are interested in the probability $\mathbb{P}(\sigma\max_t f(t) > b)$ for $\sigma \in [0.3, 1.0]$ and $b = 3$. This is equivalent to simulating $\mathbb{P}(\max_t f(t) > b)$ for all $b \in [3, 10]$. In Table 1 we present the simulation results for $\sigma = 0.3$, 0.6, and 1.0, 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 state the theoretical values of the tail probabilities, that is, $\mathbb{P}(\max_i f(t_i) > b/\sigma) = 1 - \Phi(b/\sigma)^{100}$, where $\Phi(x) = \int_{-\infty}^{x}(1/\sqrt{2\pi})\mathrm{e}^{-t^2/2}\,\mathrm{d}t$ denotes the left tail probability of the standard Gaussian distribution. We can see that the estimator is more efficient when the value of σ is equal to the designed value 1 and less for other values of σ. In particular, when $\sigma = 0.3$, it yields the estimated value 0.

The above nonuniform efficiency result can be extended, using similar techniques, to the importance sampling estimators of [3] 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].

https://doi.org/10.1017/jpr.2018.11 Published online by Cambridge University Press

TABLE 1: Estimates based on Algorithm 1.

σ	EST	SD	CV	Theoretical value
0.3	0	0	–	7.62×10^{-22}
0.6	1.35×10^{-5}	1.35×10^{-3}	1.00×10^{-2}	2.87×10^{-5}
1.0	1.26×10^{-1}	2.32×10^{-2}	1.84×10^{-1}	1.26×10^{-1}

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] and [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, in this paper we consider 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 exist positive constants κ_{H} and $\beta > 0$ such that for all $s, t \in T$, $|\sigma(t) - \sigma(s)| + |\mu(t) - \mu(s)| \le \kappa_{\mathrm{H}}|s - t|^{\beta}$.

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

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

$$L_{\sigma,\mu}(b) = \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} \mathbf{1}(\sup_{t\in T}\{\sigma(t) f(t) + \mu(t)\} > b)$$

satisfies

$$\sup_{b\ge b_0,\ \mu,\sigma\in\mathcal{C}(\mu_l,\mu_u,\sigma_l,\sigma_u,\beta,\kappa_{\mathrm{H}})} \frac{\operatorname{var}(L_{\sigma,\mu}(b))}{|\log w_{\sigma,\mu}(b)|^q w^2_{\sigma,\mu}(b)} < \infty.$$

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

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

Remark 2. Although in this paper we focus 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 (see, e.g. [27] and [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 necessary 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 was proposed by Glasserman and Juneja [16] in order 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. Unlike in this study, the random variable parameters were 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 yields a uniformly efficient estimator. We assume that $T := \{t_1, \ldots, t_M\}$. We describe the new measure $\mathbb{Q}$ in two ways. First, we specify the sampling scheme of f under $\mathbb{Q}$ and then provide its Radon–Nikodym derivative with respect to $\mathbb{P}$. Under the measure $\mathbb{Q}$, $f(t)$ is generated according to the following algorithm.

Algorithm 2. (*Simulating* $f(\cdot)$ *under* $\mathbb{Q}$.) The algorithm proceeds as follows.

Input: $T = \{t_1, \ldots, t_M\}$, $\delta_b = ab^{-1}$ 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 $T = \{t_1, \ldots, t_M\}$.

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 $\{f(t) \colon t \neq \tau,\ t \in T\}$ from the original conditional distribution under $\mathbb{P}$.

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

For the measure $\mathbb{Q}$ defined above, it is not difficult to verify that $\mathbb{P}$ and $\mathbb{Q}$ are mutually absolutely continuous with the Radon–Nikodym derivative being

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

This yields the importance sampling estimator

$$L_{\sigma,\mu}(b) = \left(\int_{\mu_l}^{\mu_u+\delta_b} \int_{\sigma_l}^{\sigma_u+\delta_b^2} \frac{\sum_{i=1}^{M} \mathbf{1}(\varsigma f(t_i) + \nu > b)}{M\mathbb{P}(\varsigma f(t_1) + \nu > b)} g(\varsigma) h(\nu)\, \mathrm{d}\varsigma\, \mathrm{d}\nu \right)^{-1} \times \mathbf{1}(\sup_{i \colon t_i \in T} \sigma(t_i) f(t_i) + \mu(t_i) > b). \tag{3}$$

Note that under $\mathbb{Q}$, 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 $\mathbb{Q}$ by comparing it with the measure $\mathbb{Q}^\dagger$ discussed in Section 2.2. We can see that steps 1 and 2 of Algorithm 1 requires knowledge of the mean and variance functions 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 value of the 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. In the next proposition we state the uniform efficiency of the proposed change of measure.

Proposition 2. *Let $L_{\sigma,\mu}(b)$ be defined as in (3). Then there exist constants b_0 and κ_p, independent of $\sigma(\cdot)$, $\mu(\cdot)$, and b, and, for $b \geq b_0$,*

$$\frac{\mathbb{E}^{\mathbb{Q}}[L^2_{\sigma,\mu}(b)]}{M^2 b^6 w^2_{\sigma,\mu}(b)} \leq \kappa_p$$

for all μ and σ satisfying (C1).

Note that $|\log(w_{\sigma,\mu}(b))| = O(b^2)$. Therefore, from the above proposition we obtain the uniformly polynomial efficiency of $\mathbb{Q}$ with 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

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

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

Remark 5. To evaluate the Radon–Nikodym derivative in (3), we need to calculate the integral

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

Define

$$l(z) = \int_{\mu_l}^{\mu_u+\delta_b}\int_{\sigma_l}^{\sigma_u+\delta_b^2} \frac{\mathbf{1}(\varsigma z + \nu > b)}{\bar{\Phi}((b-\nu)/\varsigma)} g(\varsigma)h(\nu)\,\mathrm{d}\varsigma\,\mathrm{d}\nu,$$

https://doi.org/10.1017/jpr.2018.11 Published online by Cambridge University Press

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

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

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

$$\begin{aligned} l(z) &= \int\int_{b-\varsigma s\in I_1,\,\varsigma\in I_2,\,s<z}\frac{\varsigma}{\bar{\Phi}(s)g(\varsigma)h(b-s\varsigma)\,\mathrm{d}\varsigma\,\mathrm{d}s}, \\ &= \int_{s<z}\frac{1}{\bar{\Phi}(s)}\int_{\varsigma\in(b/s-(1/s)I_1)\cap I_2}\varsigma h(b-s\varsigma)g(\varsigma)\,\mathrm{d}\varsigma\,\mathrm{d}s, \end{aligned} \tag{4}$$

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 (4) 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 $r(s)=\frac{1}{2}(\sigma_u+\delta_b^2-\sigma_l)^{-1}(\mu_u+\delta_b-\mu_l)^{-1}\int_{\varsigma\in(b/s-(1/s)I_1)\cap I_2}\mathrm{d}\varsigma^2$. Then $l(z)$ can be further simplified as

$$l(z)=\int_{-\infty}^{z}\frac{r(s)}{\bar{\Phi}(s)}\,\mathrm{d}s,$$

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 $\mathbb{R}^d$: $G_{N,d}=\{(i_1/N,i_2/N,\dots,i_d/N)\colon i_1,\dots,i_d\in\mathbb{Z}\}$. That is, $G_{N,d}$ is a regular lattice on $\mathbb{R}^d$. Furthermore, let

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

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

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

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 $\mathbb{Q}_N$ and $\mathbb{P}_N$ as the discrete versions (on T_N) of $\mathbb{Q}$ and $\mathbb{P}$, respectively. Then $\mathrm{d}\mathbb{Q}_N/\mathrm{d}\mathbb{P}_N$ takes the form

$$\frac{\mathrm{d}\mathbb{Q}_N}{\mathrm{d}\mathbb{P}_N}=\int_{\mu_l}^{\mu_u+\delta_b}\int_{\sigma_l}^{\sigma_u+\delta_b^2}\frac{\sum_{i=1}^{M}\mathbf{1}(\varsigma f(t_i)+\nu>b)}{M\mathbb{P}(\varsigma f(t_1)+\nu>b)}g(\varsigma)h(\nu)\,\mathrm{d}\varsigma\,\mathrm{d}\nu.$$

Note that here M depends on N and goes to ∞ as $N\to\infty$. This yields the importance sampling estimator

$$\begin{aligned} L_{\sigma,\mu,N}(b) &:= \bigg(\int_{\mu_l}^{\mu_u+\delta_b}\int_{\sigma_l}^{\sigma_u+\delta_b^2}\frac{\sum_{i=1}^{M}\mathbf{1}(\varsigma f(t_i)+\nu>b)}{M\mathbb{P}(\varsigma f(t_1)+\nu>b)}g(\varsigma)h(\nu)\,\mathrm{d}\varsigma\,\mathrm{d}\nu\bigg)^{-1} \\ &\quad\times\mathbf{1}(\sup\nolimits_{i\colon t_i\in T_N}\sigma(t_i)f(t_i)+\mu(t_i)>b). \end{aligned}$$

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

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

(C3) The Gaussian random field f is almost surely continuous.

(C4) Define the correlation function $r(s,t)=\mathbb{E}[f(s)f(t)]$. There exists $\beta'>0$ and $\kappa'_{\mathrm{H}}>0$ such that

$$|r(t,s)-r(t',s')|\le \kappa'_{\mathrm{H}}[|t-t'|^{\beta'}+|s-s'|^{\beta'}] \quad \text{for all } s,t,s',t'\in T.$$

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

$$\frac{|w_{\sigma,\mu,N}(b)-w_{\sigma,\mu}(b)|}{w_{\sigma,\mu}(b)}<\varepsilon \quad \textit{uniformly for } \mu,\sigma\in\mathcal{C}(\mu_l,\mu_u,\sigma_l,\sigma_u,\beta,\kappa_{\mathrm{H}}).$$

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

$$\sup_{b\ge b_0,\,\varepsilon\in(0,1)}\frac{\mathbb{E}^{\mathbb{Q}_N}L^2_{\sigma,\mu,N}(b)}{b^q w^2_{\sigma,\mu}(b)\varepsilon^{-q_1}}<\kappa_c \quad \textit{uniformly for } \mu,\sigma\in\mathcal{C}(\mu_l,\mu_u,\sigma_l,\sigma_u,\beta,\kappa_{\mathrm{H}})$$

with $q=4d/\beta^*(3d/\beta^*+2+\varepsilon_0)+6$ *and* $q_1=4d/\beta^*+2d\varepsilon_0$.

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

$$Z_{\sigma,\mu,N}(b)=\frac{1}{n}\sum_{i=1}^{n}L^{(i)}_{\sigma,\mu,N}(b). \tag{5}$$

With the aid of Chebyshev's inequality, we have

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

The MSE $\mathbb{E}[Z_{\sigma,\mu,N}(b)-w_{\sigma,\mu}(b)]^2$ can be written as

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

The first and second terms on the right-hand side of the above display are 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 $n=2\kappa_c b^q\varepsilon^{-q_1-2}\delta^{-q_1/2-1}$, 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 conditions (C1)–(C4), let* $Z_{\sigma,\mu,N}(b)$ *be defined as in (5). If we choose* $n=2\kappa_c b^q\varepsilon^{-q_1-2}\delta^{-q_1/2-1}$ *and* $N=N_0(\varepsilon\delta^{1/2},b)$, *then*

$$\mathbb{P}\bigg(\bigg|\frac{Z_{\sigma,\mu,N}(b)}{w_{\sigma,\mu}(b)}-1\bigg|>\varepsilon\bigg)<\delta. \tag{6}$$

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 (step 5 of Algorithm 2). The overall computational cost is also a polynomial in ε, δ, and b. An algorithm with such a computational cost to achieve (6) is sometimes referred to as a fully polynomial randomized approximation scheme; see [3] for more details.

4. Simulation studies

In this section we present numerical examples to demonstrate 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 Examples 2–5. In each numerical example, we report the estimated tail probabilities (EST, as before) along with the estimated standard deviations (SD, as before), that is, $\mathrm{SD}^{\mathbb{Q}}\{L_{\sigma,\mu}(b)\} = \sqrt{\mathrm{var}^{\mathbb{Q}}\{L_{\sigma,\mu}(b)\}}$. The standard error of the estimator with 10^4 Monte Carlo samples is SD /100. We also report the coefficient of variation (CV, as before) of the estimators, which is the ratio SD/EST of the estimators.

We start with the discrete setting in Example 1, where $T = \{1, \dots, 100\}$ and $\{f(t), t = 1, \dots, 100\}$ are i.i.d. standard normal random variables. We take $\mu(t) = 0$ and $\sigma(t) = \sigma$ with $\sigma \in [0.3, 1.0]$ for all $t \in T$, and the probability of interest is $\mathbb{P}(\sigma \max_t f(t) > b)$ for $b = 3$. In Table 2 we present the simulation results for $\sigma = 0.3$, 0.6, and 1.0 using the proposed method. For different values of σ, the estimates are close to the true values. Compared with the results of Algorithm 1 (Table 1), the proposed method yields a better overall performance.

We proceed to an example of a continuous Gaussian random field, whose tail probability of the supremum is in 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 $T = [0, \frac{3}{4}]$. We let $b = 4$ and consider the class of constant variance and mean functions: $\sigma(t) = \sigma$ and $\mu(t) = \mu$, with $\sigma \in [\frac{1}{2}, 1]$ and $\mu \in [-\frac{1}{2}, \frac{1}{2}]$.

For constant mean and variance functions considered in this example, the probability

$$\mathbb{P}(\sup_{t\in T}(\sigma f(t) + \mu) > b)$$

is known to be in closed form [1]:

$$\mathbb{P}\Big(\sup_{0\le t\le 3/4}(\sigma f(t) + \mu) > b\Big) = \bar{\Phi}\Big(\frac{b-\mu}{\sigma}\Big) + \frac{3}{8\pi}\exp\Big(-\frac{(b-\mu)^2}{2\sigma^2}\Big). \tag{7}$$

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.

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

σ	EST	SD	CV	Theoretical value
0.3	7.55×10^{-22}	5.33×10^{-21}	7.05	7.62×10^{-22}
0.6	2.93×10^{-5}	1.33×10^{-4}	4.52	2.87×10^{-5}
1.0	1.26×10^{-1}	5.92×10^{-1}	4.69	1.26×10^{-1}

TABLE 3: Simulation result for Example 2 with $b = 4$ and $\delta_b = 1/b$. Theoretical values are computed according to (7).

σ	μ	EST	SD	CV	Theoretical value
0.5	0.5	4.18×10^{-12}	2.59×10^{-11}	6.2	4.01×10^{-12}
0.6	0.3	1.03×10^{-9}	4.38×10^{-9}	4.2	1.01×10^{-9}
0.7	0.1	3.34×10^{-8}	1.18×10^{-7}	3.5	3.43×10^{-8}
0.8	−0.1	3.68×10^{-7}	1.19×10^{-6}	3.2	3.85×10^{-7}
0.9	−0.3	2.10×10^{-6}	5.97×10^{-6}	2.8	2.20×10^{-6}
1.0	−0.5	8.11×10^{-6}	2.20×10^{-5}	2.7	8.18×10^{-6}

The theoretical value is computed according to (7). 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 decreases 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 $\{f(t) \colon 0 \le t \le 1\}$, whose covariance function is

$$r(s,t) = \mathbb{E}[f(s)f(t)] = \mathrm{e}^{-|s-t|}. \tag{8}$$

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.

Example 3. Consider the Gaussian random field $f(t)$ defined in (8), and the class of variance and mean functions $\sigma(t) = 1$ and $\mu(t) = \beta_1 t$ for $\beta_1 \in [-\frac{1}{2}, \frac{1}{2}]$. The probability of interest is $\mathbb{P}(\sup_{t\in[0,1]} f(t) + \beta_1 t > b)$ for $b = 7$.

We summarize the simulation results for Example 3 in Figure 1. In Figure 1(a) we present the scatter plot of the estimated probability (EST) against β_1 and in Figure 1(b) we present the scatter plot of the CV against β_1. We see that the probability of interest is an increasing function in β_1. Moreover, when the estimated probability is within the range 1×10^{-11} to 2×10^{-10}, the CV of the estimator is always controlled within the range 1.8 to 3.2, showing good performance of the proposed estimation method.

Example 4. Consider the Gaussian random field $\{f(t),\ t \in T\}$ defined in (8) and the class of variance and mean functions $\sigma(t) = 1 - \frac{1}{2}(t-\beta_2)^2$ and $\mu(t) = 0$, where $\beta_2 \in [0,1]$. The probability of interest is $\mathbb{P}(\sup_{t\in[0,1]}[1 - \frac{1}{2}(t-\beta_2)^2]f(t) > b)$ for $b = 7$.

For Example 4, the scatter plot of the estimated probability and the CV of the estimator are presented in Figure 2. Note that in Example 4, the maximum variance $\max_{t\in T} \operatorname{var}(\sigma(t)f(t)) = \operatorname{var}(\sigma(\beta_2)f(\beta_2)) = 1$. Therefore, for all $\beta_2 \in [0,1]$, the probability of interest has the same exponential decay rate

$$\begin{aligned}\mathbb{P}\Big(\sup_{t\in[0,1]} \sigma(t)f(t) > b\Big) &= \exp\bigg(-(1+o(1))\frac{b^2}{2\max_{t\in T}\operatorname{var}(\sigma(t)f(t))}\bigg)\\ &= \exp\bigg(-\frac{(1+o(1))b^2}{2}\bigg) \quad \text{as } b \to \infty.\end{aligned}$$

https://doi.org/10.1017/jpr.2018.11 Published online by Cambridge University Press

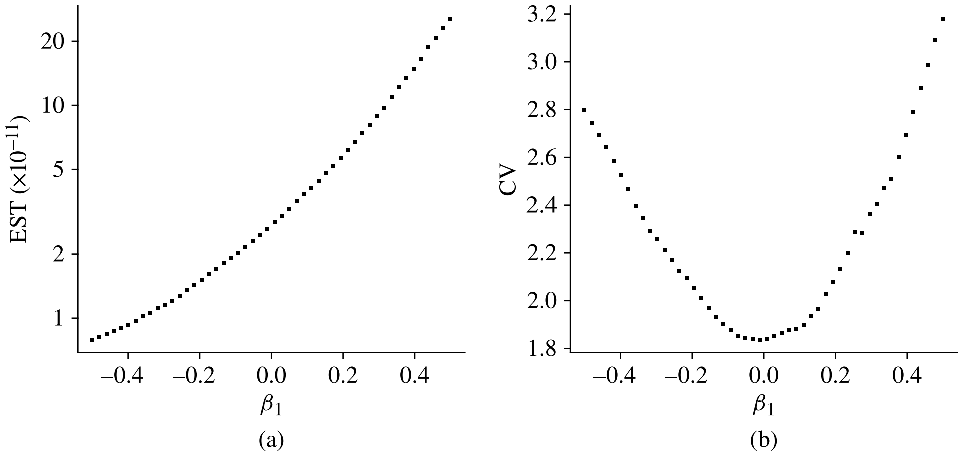

FIGURE 1: Simulation results for Example 3, where $b = 7$ and $\delta_b = 1/b$. (a) Estimates as a function of β_1, (b) CV as a function of β_1.

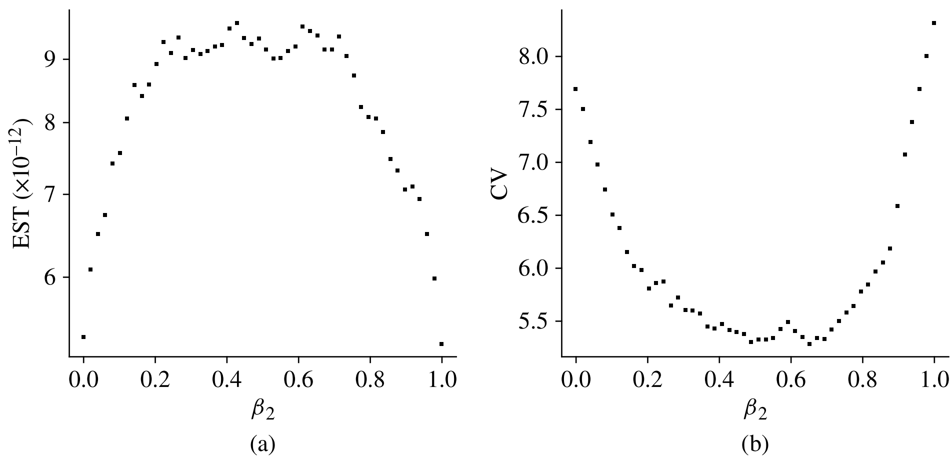

FIGURE 2: Simulation results for Example 4, where $b = 7$ and $\delta_b = 1/b$. (a) Estimates as a function of β_2, (b) CV as a function of β_2.

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 $\beta_2 \in [0.2, 0.8]$, when it 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 the range 5 to 10. This is again consistent with our theoretical results.

Example 5. Consider the Gaussian random field $\{f(t),\ t \in T\}$ defined in (8), and the class of variance and mean functions $\sigma(t) = 1 - \frac{1}{2}(x - \beta_2)^2$ and $\mu(t) = \beta_1 t$, where $\beta_1 \in [-\frac{1}{2}, \frac{1}{2}]$ and $\beta_2 \in [0, 1]$. The probability of interest is

$$\mathbb{P}\Big(\sup_{t\in[0,1]} \big\{\big[1 - \tfrac{1}{2}(t - \beta_2)^2\big] f(t) + \beta_1 t\big\} > b\Big) \quad \text{for } b = 7.$$

https://doi.org/10.1017/jpr.2018.11 Published online by Cambridge University Press

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

β_1	β_2	EST	SD	CV
−0.50	0.00	4.20×10^{-12}	4.03×10^{-11}	9.6
−0.33	0.17	5.60×10^{-12}	3.69×10^{-11}	6.6
−0.17	0.33	5.69×10^{-12}	3.29×10^{-11}	5.8
0.00	0.50	8.78×10^{-12}	5.09×10^{-11}	5.8
0.17	0.67	2.09×10^{-11}	1.27×10^{-10}	6.1
0.33	0.83	5.82×10^{-11}	4.04×10^{-10}	6.9
0.50	1.00	1.16×10^{-10}	1.15×10^{-9}	9.9

In Table 4 we present 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 $9.9\% \times \mathbb{E}^{\mathbb{Q}} L_{\sigma,\mu}(b)$.

5. Proofs of the main results

Throughout the proofs, we write $a(b) = O(c(b))$ if there exists a positive constant κ, independent of b, $\sigma(\cdot)$, and $\mu(\cdot)$, 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 condition (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 $\tilde{\varepsilon}$ as a generic notation for small positive constants.

Proof of Proposition 1. (i) We 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 have

$$\begin{aligned}
&\mathbb{E}^{\mathbb{Q}^\dagger}\bigg[\bigg(\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}^\dagger}\bigg)^2;\ \max_{t_i \in T} \sigma'(t_i) f(t_i) + \mu(t_i) > b\bigg] \\
&\quad = \mathbb{E}\bigg[\frac{\mathrm{d}\mathbb{Q}^\dagger}{\mathrm{d}\mathbb{P}}\bigg(\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}^\dagger}\bigg)^2;\ \max_{t_i \in T} \sigma'(t_i) f(t_i) + \mu(t_i) > b\bigg] \\
&\quad = \mathbb{E}\bigg[\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}^\dagger};\ \max_{t_i \in T} \sigma'(t_i) f(t_i) + \mu(t_i) > b\bigg] \\
&\quad = \mathbb{E}\bigg[\frac{\sum_{i=1}^M \mathbb{P}(\sigma(t_i) f(t_i) + \mu(t_i) > b)}{\sum_{i=1}^M \mathbf{1}(\sigma(t_i) f(t_i) + \mu(t_i) > b)};\ \max_{t_i \in T} \sigma'(t_i) f(t_i) + \mu(t_i) > b\bigg].
\end{aligned}$$

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

$$\begin{aligned}
&\frac{1}{M}\bigg(\sum_{i=1}^M \mathbb{P}(\sigma(t_i) f(t_i) + \mu(t_i) > b)\bigg) w_{\sigma',\mu}(b) \\
&\quad \ge \frac{1}{M} \max_{t_i \in T} \mathbb{P}(\sigma(t_i) f(t_i) + \mu(t_i) > b) w_{\sigma',\mu}(b)
\end{aligned}$$

$$= \exp\bigg\{-(1+o(1))\frac{b^2}{2\max_{t_i\in T}\sigma(t_i)^2} - (1+o(1))\frac{b^2}{2\max_{t_i\in T}\sigma'(t_i)^2}\bigg\},$$

where we used the following lemma, whose proof can be found in Section 5.1, to obtain

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

Lemma 1. *Let $\{f(t)\colon t\in T\}$ 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 exist positive $\tilde{\varepsilon}$ such that*

$$\mathbb{P}\Big(\sup_{t\in T}\sigma(t)f(t)+\mu(t)>b\Big) = \exp\bigg(-(1+o(1))\frac{b^2}{2\max_{t\in T}\sigma^2(t)}\bigg),$$

or

$$\mathbb{P}\Big(\sup_{t\in T}\sigma(t)f(t)+\mu(t)>b\Big) \geq \tilde{\varepsilon}b^{-1}\max_{t\in T}\exp\bigg(-\frac{(b-\mu(t))^2}{2\sigma^2(t)}\bigg).$$

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

$$\frac{\mathbb{E}^{\mathbb{Q}^\dagger}[(\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}^\dagger)^2;\ \max_{t_i\in T}\sigma'(t_i)f(t_i)+\mu(t_i)>b]}{w^{2-\varepsilon}_{\sigma',\mu}(b)} \geq w^{-\varepsilon}_{\sigma',\mu}(b),$$

which tends to ∞ as $b\to\infty$.

We return to the proof of Proposition 1. (ii) Let $t'_{\max} = \arg\max_{t\in T}\sigma'(t)$. We consider the event

$$F = \bigg\{\frac{b}{\sigma'(t'_{\max})} < f(t'_{\max}) < \min_{t_i\in T}\bigg[\frac{b}{\sigma(t_i)}\bigg]\bigg\}.$$

Since $\max_{t_i\in T}\sigma'(t_i) > \max_{t_i\in T}\sigma(t_i)$, F is nonempty 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 $\mathbb{Q}^\dagger(F)>0$. On the other hand, when the event F happens, $\sum_{i=1}^M \mathbf{1}(\sigma(t_i)f(t_i)>b)=0$, therefore $\mathbb{Q}^\dagger(\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}^\dagger=\infty)\geq\mathbb{Q}^\dagger(F)>0$. In other words, $\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}^\dagger$ is not well defined. □

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

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}[L^2_{\sigma,\mu}(b)] &= \mathbb{E}^{\mathbb{Q}}\bigg[\bigg(\int_{\mu_l}^{\mu_u+\delta_b}\int_{\sigma_l}^{\sigma_u+\delta_b^2}\frac{\sum_{i=1}^M\mathbf{1}(\varsigma f(t_i)+\nu>b)}{M\mathbb{P}(\varsigma f(t_1)+\nu>b)}g(\varsigma)\,\mathrm{d}\varsigma\,\mathrm{d}\nu\bigg)^{-2};\\ &\qquad \max_{t_i\in T}\sigma(t_i)f(t_i)+\mu(t_i)>b\bigg]\\ &\leq \mathbb{E}^{\mathbb{Q}}\bigg[\bigg(\int_{\mu(t^*)}^{\mu(t^*)+\delta_b}\int_{\sigma(t^*)}^{\sigma(t^*)+\delta_b^2}\frac{\sum_{i=1}^M\mathbf{1}(\varsigma f(t_i)+\nu>b)}{M\mathbb{P}(\varsigma f(t_1)+\nu>b)}g(\varsigma)h(\nu)\,\mathrm{d}\varsigma\,\mathrm{d}\nu\bigg)^{-2};\\ &\qquad \max_{t_i\in T}\sigma(t_i)f(t_i)+\mu(t_i)>b\bigg]\end{aligned}$$

https://doi.org/10.1017/jpr.2018.11 Published online by Cambridge University Press

$$= \mathbb{E}^{\mathbb{Q}}\bigg[\bigg(\int_{\mu(t^*)}^{\mu(t^*)+\delta_b}\int_{\sigma(t^*)}^{\sigma(t^*)+\delta_b^2}\frac{\sum_{i=1}^{M}\mathbf{1}(\varsigma f(t_i)+\nu>b)}{M\bar{\Phi}((b-\nu)/\varsigma)}g(\varsigma)h(\nu)\,\mathrm{d}\varsigma\,\mathrm{d}\nu\bigg)^{-2};\ \max_{t_i\in T}\sigma(t_i)f(t_i)+\mu(t_i)>b\bigg]. \tag{9}$$

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 that $\varsigma f(t^*)+\nu\geq b$. Consequently, $\sum_{i=1}^{M}\mathbf{1}(\varsigma f(t_i)+\nu>b)\geq 1$ on the event $\max_{t_i\in T}\sigma(t_i)f(t_i)+\mu(t_i)>b$. Therefore, (9) is further bounded from above by

$$\begin{aligned}
&\mathbb{E}^{\mathbb{Q}}[L^2_{\sigma,\mu}(b)]\\
&\quad\leq M^2\mathbb{E}^{\mathbb{Q}}\bigg[\bigg(\int_{\mu(t^*)}^{\mu(t^*)+\delta_b}\int_{\sigma(t^*)}^{\sigma(t^*)+\delta_b^2}\frac{g(\varsigma)h(\nu)}{\bar{\Phi}((b-\nu)/\varsigma)}\,\mathrm{d}\varsigma\,\mathrm{d}\nu\bigg)^{-2};\ \max_{t_i\in T}\sigma(t_i)f(t_i)+\mu(t_i)>b\bigg]\\
&\quad\leq O(1)M^2\mathbb{E}^{\mathbb{Q}}\bigg[\bigg(\int_{\mu(t^*)}^{\mu(t^*)+\delta_b}\int_{\sigma(t^*)}^{\sigma(t^*)+\delta_b^2}g(\varsigma)h(\nu)b\exp\bigg(\frac{(b-\nu)^2}{2\varsigma^2}\bigg)\,\mathrm{d}\varsigma\,\mathrm{d}\nu\bigg)^{-2};\\
&\qquad\qquad \max_{t_i\in T}\sigma(t_i)f(t_i)+\mu(t_i)>b\bigg].
\end{aligned}\tag{10}$$

Note that, for all $(\varsigma,\nu)\in[\mu(t^*),\mu(t^*)+\delta_b]\times[\sigma(t^*),\sigma(t^*)+\delta_b^2]$, we have

$$b\exp\bigg(\frac{(b-\nu)^2}{2\varsigma^2}\bigg)=O(1)b\exp\bigg(\frac{(b-\mu(t^*))^2}{2\sigma^2(t^*)}\bigg).$$

Therefore, (10) is bounded from above by

$$\begin{aligned}
&\mathbb{E}^{\mathbb{Q}}[L^2_{\sigma,\mu}(b)]\\
&\quad\leq O(1)M^2\mathbb{E}^{\mathbb{Q}}\bigg[\bigg(\int_{\mu(t^*)}^{\mu(t^*)+\delta_b}\int_{\sigma(t^*)}^{\sigma(t^*)+\delta_b^2}g(\varsigma)h(\nu)b\exp\bigg(\frac{(b-\mu(t^*))^2}{2\sigma^2(t^*)}\bigg)\,\mathrm{d}\varsigma\,\mathrm{d}\nu\bigg)^{-2};\\
&\qquad\qquad \max_{t_i\in T}\sigma(t_i)f(t_i)+\mu(t_i)>b\bigg]\\
&\quad= O(1)M^2\delta_b^{-6}b^{-2}\mathbb{E}^{\mathbb{Q}}\bigg[\exp\bigg(-\frac{(b-\mu(t^*))^2}{\sigma^2(t^*)}\bigg);\ \max_{t_i\in T}\sigma(t_i)f(t_i)+\mu(t_i)>b\bigg]\\
&\quad\leq O(1)M^2\delta_b^{-6}b^{-2}\max_{t_i\in T}\exp\bigg(-\frac{(b-\mu(t_i))^2}{\sigma^2(t_i)}\bigg).
\end{aligned}\tag{11}$$

On the other hand, according to Lemma 1, we have

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

Combining this and (11), it follows that there exists b_0 sufficiently large such that, for $b\geq b_0$,

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

https://doi.org/10.1017/jpr.2018.11 Published online by Cambridge University Press

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

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

We split the above probability into two parts:

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

which is further bounded from above by

$$\begin{aligned}&\mathbb{P}\bigg(b<\sup_{t\in T}\sigma(t)f(t)+\mu(t)\leq b+\frac{\gamma}{b}\bigg)\\&\quad+\mathbb{P}\bigg(\sup_{t\in T}\sigma(t)f(t)+\mu(t)>b+\frac{\gamma}{b},\ \sup_{t\in T_N}\sigma(t)f(t)+\mu(t)\leq b\bigg),\end{aligned}\tag{12}$$

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

Lemma 2. (Proposition 6.5 of [3].) *Under conditions (C1), (C3), and (C4), for any $v>0$, let $\beta^*=\min(\beta,\beta')$ and $\rho=2d/\beta^*+dv+1$, where d is the dimension of T. There exist constants $b_0,\lambda\in(0,\infty)$ such that, for all $b\geq b_0\geq 1$,*

$$\mathbb{P}\bigg(\max_{t\in T}\sigma(t)f(t)+\mu(t)\leq b+\frac{\gamma}{b}\ \bigg|\ \max_{t\in T}\sigma(t)f(t)+\mu(t)>b\bigg)\leq\lambda ab^{\rho}.$$

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

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

with $\rho=3d/\beta^*+1$. We choose $\gamma:=2^{-1}\lambda^{-1}b^{-\rho}\varepsilon$. The above display yields the following upper bound for the first term in (12):

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

We proceed to the second term in (12). According to condition (C2), we have

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

which is further bounded from above by

$$\mathbb{P}\Big(\sup_{t,s\in T,\,|t-s|\le\kappa_m/N}|\sigma(t)f(t)-\sigma(s)f(s)|+\sup_{t,s\in T,\,|t-s|\le\kappa_m/N}|\mu(t)-\mu(s)|>\frac{\gamma}{b}\Big). \qquad (13)$$

According to condition (C1), we have

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

Substituting this into (13), we have

$$\mathbb{P}\Big(\sup_{t,s\in T,\,|t-s|\le\kappa_m/N}|\sigma(t)f(t)+\mu(t)-\sigma(s)f(s)+\mu(s)|>\frac{\gamma}{b}\Big)$$
$$\le \mathbb{P}\Big(\sup_{t,s\in T,\,|t-s|\le\kappa_m/N}|\sigma(t)f(t)-\sigma(s)f(s)|>\frac{\gamma}{b}-\frac{\kappa_m^{\beta^*}}{N^{\beta^*}}\Big).$$

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

$$\mathbb{P}\Big(\sup_{t,s\in T,\,|t-s|\le\kappa_m/N}|\sigma(t)f(t)+\mu(t)-\sigma(s)f(s)+\mu(s)|>\frac{\gamma}{b}\Big)$$
$$\le \mathbb{P}\Big(\sup_{t,s\in T,\,|t-s|\le\kappa_m/N}|\sigma(t)f(t)-\sigma(s)f(s)|>\frac{\gamma}{2b}\Big).$$

To control the above probability, we use the following lemma, known as the Borell–TIS lemma, which was proved independently by Borell [10] and Tsirelson *et al.* [33].

Lemma 3. (The Borell–TIS lemma.) *Let $\{f(t);\ t\in\mathcal{U}\}$, where $\mathcal{U}$ is a compact set, be a zero-mean Gaussian random field with f almost surely bounded on $\mathcal{U}$. Then $\mathbb{E}[\sup_{\mathcal{U}}f(t)]<\infty$ and $\mathbb{P}(\sup_{t\in\mathcal{U}}f(t)-\mathbb{E}[\sup_{t\in\mathcal{U}}f(t)]\ge b)\le\exp(-b^2/2\sigma_{\mathcal{U}^2})$, where $\sigma_{\mathcal{U}}^2=\sup_{t\in\mathcal{U}}\operatorname{var}[f(t)]$.*

We define a new Gaussian random field

$$\xi(s,t)=\sigma(s)f(s)-\sigma(t)f(t).$$

We use the next lemma to characterize $\mathbb{E}[\sup_{t,s\in T,\,|t-s|\le\kappa_m/N}\xi(s,t)]$ (see Section 5.1 for the proof).

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

$$\mathbb{E}\Big[\sup_{t,s\in T,\,|t-s|\le\kappa_m/N}|\xi(s,t)|\Big]<\kappa_\xi N^{-\beta^*/2}\log N.$$

https://doi.org/10.1017/jpr.2018.11 Published online by Cambridge University Press

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

$$\begin{aligned}\operatorname{var}(\xi(s,t)) &= (\sigma(s)-\sigma(t))^2 + 2\sigma(s)\sigma(t)(1-r(s,t))\\ &\le \kappa_{\mathrm{H}}^2|s-t|^{2\beta^*} + 2\sigma_u^2|s-t|^{\beta^*}\\ &\le O(|s-t|^{\beta^*}). \end{aligned} \tag{14}$$

According to conditions (C1) and (C4), the above display is further bounded from above by

$$\operatorname{var}(\xi(s,t)) \le O(N^{-\beta^*}).$$

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

$$\mathbb{P}\bigg(\sup_{|t-s|\le\kappa_m/N}|\xi(s,t)| > \frac{\gamma}{4b}\bigg) \le \exp\bigg(-\tilde{\varepsilon}\frac{\gamma^2}{N^{-\beta^*}b^2}\bigg).$$

The above display is of order $o(\varepsilon w_{\sigma,\mu}(b))$ if $\gamma^2/N^{-\beta^* b^2} \ge \tilde{\kappa}^{\beta^*}\max(-\log\varepsilon, b^2)$ for a large enough and possibly different constant $\tilde{\kappa}$. Therefore, it is sufficient to choose

$$N \ge \tilde{\kappa}\max(-\log\varepsilon, b^2)^{1/\beta^*} b^{2/\beta^*}\gamma^{-2/\beta^*}(\log b)^{\tilde{\kappa}}.$$

Combining this with our choice of γ, and recalling our choice of ρ in Lemma 2, it is sufficient to choose

$$N \ge \tilde{\kappa}\max(-\log\varepsilon, b^2)^{1/\beta^*} b^{2/\beta^*+2/\beta^*(3d/\beta^*+1)}\varepsilon^{-2/\beta^*}(\log b)^{\tilde{\kappa}},$$

which is bounded by $N_0 = b^{2/\beta^*(3d/\beta^*+2+\varepsilon_0)}\varepsilon^{-2/\beta^*-\varepsilon_0}$ for any $\varepsilon_0 > 0$ and sufficiently large b. This completes the proof. □

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

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

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

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

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

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

$$\begin{aligned}\mathbb{E}[Z_{\sigma,\mu,N}(b) - w_{\sigma,\mu}(b)]^2 &= [\mathbb{E}Z_{\sigma,\mu,N}(b) - w_{\sigma,\mu}(b)]^2 + \operatorname{var}(Z_{\sigma,\mu,N}(b))\\ &= [w_{\sigma,\mu,N}(b) - w_{\sigma,\mu}(b)]^2 + \frac{\operatorname{var}(L_{\sigma,\mu,N}(b))}{n}.\end{aligned}$$

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 $N \ge N(\varepsilon\delta^{1/2}, b)$. Furthermore, according to Theorem 2, we have $\operatorname{var}(L_{\sigma,\mu,N}(b))/n \le \varepsilon^2\delta w^2_{\sigma,\mu}(b)/2$ for $n \ge 2\kappa_c b^q\varepsilon^{-q_1-2}\delta^{-q_1/2-1}$. Consequently, for such N and n, we have $\mathbb{E}[Z_{\sigma,\mu,N}(b) - w_{\sigma,\mu}(b)]^2 \le \varepsilon^2\delta$. Thanks to Chebyshev's inequality, we have

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

Therefore, $Z_{\sigma,\mu,N}(b)$ satisfies (6). □

5.1. Proofs of supporting lemmas

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

$$\mathbb{P}\Big(\sup_{t\in T}\sigma(t)f(t)+\mu(t)>b\Big)\le\mathbb{P}\Big(\sup_{t\in T}\sigma(t)f(t)>b-\max_{t\in T}\mu(t)\Big)$$
$$\le\exp\bigg(-(1+o(1))\frac{b^2}{2\max_{t\in T}\sigma^2(t)}\bigg). \tag{15}$$

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

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

which is further bounded from below by

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

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

$$\mathbb{P}\Big(\sup_{t\in T}\sigma(t)f(t)+\mu(t)>b\Big)\ge\tilde{\varepsilon}b^{-1}\max_{t\in T}\exp\bigg(-\frac{(b-\mu(t))^2}{2\sigma^2(t)}\bigg).$$

Combining the above expression with (15), 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*

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

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

$$\mathbb{E}\Big[\sup_{t\in\mathcal{U}}f(t)\Big]\le K\int_0^{\operatorname{diam}(\mathcal{U})}(\log N(\varepsilon))^{1/2}\,\mathrm{d}\varepsilon.$$

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

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

We first investigate the metric d_ξ. We have

$$d_\xi^2((s,t),(s',t'))\le 2\operatorname{var}(\xi(s,s'))+2\operatorname{var}(\xi(t,t')).$$

Applying (14) to the above display, it follows that there is a $\tilde{\kappa}$ uniformly for all σ, μ satisfying condition (C1), such that

$$d_\xi((s,t),(s',t'))\le\tilde{\kappa}\sqrt{|s-s'|^{\beta^*}+|t-t'|^{\beta^*}}. \tag{16}$$

https://doi.org/10.1017/jpr.2018.11 Published online by Cambridge University Press

According to the relationship between the l_p-norms, we have $(|s-s'|^{\beta^*}+|t-t'|^{\beta^*})^{1/\beta^*} \le d^{1/2-1/\beta^*}\sqrt{|s-s'|^2+|t-t'|^2}$. The result, together with (16), implies that $B((s,t), \tilde{\varepsilon}\varepsilon^{2/\beta^*}) \subset B_{d_\xi}((s,t),\varepsilon)$ for some constant $\tilde{\varepsilon}$ that depends only 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 $\tilde{\kappa}\varepsilon^{-4d/\beta^*}$ many $B(\tilde{\varepsilon}\varepsilon^{2/\beta^*})$ 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 $d_\xi((s,t),(s',t')) \le 2\operatorname{var}(\xi(s,t))+2\operatorname{var}(\xi(s',t'))$. Also, according to (14), 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, $\operatorname{diam}(\mathcal{U}) \le \tilde{\kappa}N^{-\beta^*/2}$. According to Lemma 5, we have

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

Acknowledgements

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

References

[1] Adler, R. J. and Taylor, J. E. (2007). *Random Fields and Geometry*. Springer, New York.

[2] Adler, R. J., Blanchet, J. and Liu, J. (2008). Efficient simulation for tail probabilities of Gaussian random fields. In *Proceedings of the 2008 Winter Simulation Conference*, IEEE, pp. 328–336.

[3] Adler, R. J., Blanchet, J. H. and Liu, J. (2012). Efficient Monte Carlo for high excursions of Gaussian random fields. *Ann. Appl. Prob.* **22,** 1167–1214,

[4] Adler, R. J., Müller, P. and Rozovskii, B. (1996). *Stochastic Modelling in Physical Oceanography*. Birkhaüser, Boston, MA.

[5] Adler, R. J., Taylor, J. E. and Worsley, K. J. (2017). Applications of random fields and geometry: foundations and case studies. Available at http://webee.technion.ac.il/Sites/People/adler/.

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

[7] Azaïs, J.-M. and Wschebor, M. (2008). A general expression for the distribution of the maximum of a Gaussian field and the approximation of the tail. *Stoch. Process. Appl.* **118,** 1190–1218. (Erratum: **120** (2010), 2100–2101.)

[8] Azaïs, J.-M. and Wschebor, M. (2009). *Level Sets and Extrema of Random Processes and Fields*. John Wiley, Hoboken, NJ.

[9] Berman, S. M. (1985). An asymptotic formula for the distribution of the maximum of a Gaussian process with stationary increments. *J. Appl. Prob.* **22,** 454–460.

[10] Borell, C. (1975). The Brunn-Minkowski inequality in Gauss space. *Invent. Math.* **30,** 207–216.

[11] Borell, C. (2003). The Ehrhard inequality. *C. R. Math. Acad. Sci. Paris* **337,** 663–666.

[12] Cheng, D. and Xiao, Y. (2016). The mean Euler characteristic and excursion probability of Gaussian random fields with stationary increments. *Ann. Appl. Prob.* **26,** 722–759.

[13] Dębicki, K., Hashorva, E. and Ji, L. (2016). Extremes of a class of nonhomogeneous Gaussian random fields. *Ann. Prob.* **44,** 984–1012.

[14] Dudley, R. M. (1973). Sample functions of the Gaussian process. *Ann. Prob.* **1,** 66–103.

[15] Glasserman, P. (2004). *Monte Carlo Methods in Financial Engineering*. Springer, New York.

[16] Glasserman, P. and Juneja, S. (2008). Uniformly efficient importance sampling for the tail distribution of sums of random variables. *Math. Operat. Res.* **33,** 36–50.

[17] Juneja, S. and Shahabuddin, P. (2006). Rare-event simulation techniques: an introduction and recent advances. In *Handbooks in Operations Research and Management Science*, Vol. 13, Elsevier, pp. 291–350.
[18] Landau, H. J. and Shepp, L. A. (1970). On the supremum of a Gaussian process. *Sankhyā A* **32,** 369–378.
[19] Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes*. Springer, Berlin.
[20] Li, X. and Liu, J. (2015). Rare-event simulation and efficient discretization for the supremum of Gaussian random fields. *Adv. Appl. Prob.* **47,** 787–816.
[21] Li, X., Liu, J. and Xu, G. (2016). On the tail probabilities of aggregated lognormal random fields with small noise. *Math. Operat. Res.* **41,** 236–246.
[22] Li, X., Liu, J. and Ying, Z. (2016). Chernoff index for Cox test of separate parametric families. Preprint. Available at https://arxiv.org/abs/1606.08248.
[23] Liu, J. (2012). Tail approximations of integrals of Gaussian random fields. *Ann. Prob.* **40,** 1069–1104.
[24] Liu, J. and Xu, G. (2012). Rare-event simulations for exponential integrals of smooth Gaussian processes. In *Proceedings of the 2012 Winter Simulation Conference*, IEEE, pp. 1–10.
[25] Liu, J. and Xu, G. (2012). Some asymptotic results of Gaussian random fields with varying mean functions and the associated processes. *Ann. Statist.* **40,** 262–293.
[26] Liu, J. and Xu, G. (2013). On the density functions of integrals of Gaussian random fields. *Adv. Appl. Prob.* **45,** 398–424.
[27] Liu, J. and Xu, G. (2014). Efficient simulations for the exponential integrals of Hölder continuous Gaussian random fields. *ACM Trans. Model. Comput. Simul.* **24,** 9.
[28] Liu, J. and Xu, G. (2014). On the conditional distributions and the efficient simulations of exponential integrals of Gaussian random fields. *Ann. Appl. Prob.* **24,** 1691–1738.
[29] Marcus, M. B. and Shepp, L. A. (1970). Continuity of Gaussian processes. *Trans. Amer. Math. Soc.* **151,** 377–391.
[30] Sudakov, V. N. and Tsirelson, B. S. (1974). Extremal properties of half-spaces for spherically invariant measures. *Zap. Naučn. Sem. LOMI* **41,** 14–24.
[31] Sun, J. (1993). Tail probabilities of the maxima of Gaussian random fields. *Ann. Prob.* **21,** 34–71.
[32] Talagrand, M. (1996). Majorizing measures: the generic chaining. *Ann. Prob.* **24,** 1049–1103.
[33] Tsirelson, B. S., Ibragimov, I. A. and Sudakov, V. N. (1976). Norms of Gaussian sample functions. In *Proceedings of the Third Japan-USSR Symposium on Probability Theory*, Springer, Berlin, pp. 20–41.
[34] Xu, G. (2014). Uniformly efficient simulation for tail probabilities of Gaussian random fields. In *Proceedings of the 2014 Winter Simulation Conference*, IEEE, pp. 533–542.