Adv. Appl. Prob. **47,** 787–816 (2015)
Printed in Northern Ireland

RARE-EVENT SIMULATION AND EFFICIENT DISCRETIZATION FOR THE SUPREMUM OF GAUSSIAN RANDOM FIELDS

XIAOOU LI * AND
JINGCHEN LIU,* ** *Columbia University*

Abstract

In this paper we consider a classic problem concerning the high excursion probabilities of a Gaussian random field f living on a compact set T. We develop efficient computational methods for the tail probabilities $\mathbb{P}\{\sup_T f(t) > b\}$. For each positive ε, we present Monte Carlo algorithms that run in *constant* time and compute the probabilities with relative error ε for arbitrarily large b. The efficiency results are applicable to a large class of Hölder continuous Gaussian random fields. Besides computations, the change of measure and its analysis techniques have several theoretical and practical indications in the asymptotic analysis of Gaussian random fields.

Keywords: Gaussian random field; high-level excursion; Monte Carlo; tail distribution; efficiency

2010 Mathematics Subject Classification: Primary 60G15; 65C05
Secondary 60G60; 62G32

1. Introduction

In this paper we consider the design and the analysis of efficient Monte Carlo methods for the high excursion events of Gaussian random fields. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a Gaussian random field

$$f : T \times \Omega \to R$$

living on a d-dimensional compact subset $T \subset R^d$. Most of the time, we omit the second argument and write $f(t)$. Let $M = \sup_{t\in T} f(t)$. In this paper, we are interested in the efficient computation of the high excursion probabilities, i.e.

$$w(b) \triangleq \mathbb{P}\{M > b\} \quad \text{as } b \to \infty.$$

On computing small probabilities converging to 0, it is sensible to consider the relative accuracy that is defined as follows.

Definition 1. For some positive ε and δ, a Monte Carlo estimator Z of w is said to admit $\varepsilon - \delta$ relative accuracy if

$$\mathbb{P}\{|Z - w| < \varepsilon w\} > 1 - \delta. \tag{1}$$

We propose a Monte Carlo estimator admitting $\varepsilon - \delta$ relative accuracy for computing the tail probabilities $w(b)$. One notable feature of this estimator is that the total computational

Received 15 May 2013; revision received 23 August 2014.
* Postal address: Department of Statistics, Columbia University, 1255 Amsterdam Avenue, New York, NY 10027, USA.
** Email address: jcliu@stat.columbia.edu

https://doi.org/10.1239/aap/1444308882 Published online by Cambridge University Press

complexity to generate one such estimator is bounded by a constant $C(\varepsilon, \delta)$ that is independent of the excursion level b. Thus, to compute $w(b)$ with any prescribed relative accuracy as in (1), the total computational complexity remains bounded as the event becomes arbitrarily rare. With such an algorithm, the computation of rare-event probabilities is at the same level of complexity as the computation of regular probabilities. This efficiency result is applicable to a large class of Hölder continuous Gaussian random fields and, thus, is very generally applicable.

The analysis mainly consists of two components. First, we consider a change of measure on the continuous sample path space (denoted by Q_b). The corresponding importance sampling estimator given in (9) is unbiased. The first step of the analysis is to show that this estimator admits a standard deviation on the order $O(w(b))$. Such estimators are said to be *strongly efficient*, which is a common efficiency concept in the rare-event simulation literature; see [6], [13].

The second part of the analysis concerns the implementation. The simulation of the estimators in the previous paragraph requires the generation of the entire sample path of f. In that context, the process f is a continuous function. A computer can only generate finite-dimensional objects, so we need to seek an appropriate discretization scheme in order to perform the simulations. For instance, a natural approach is to choose a subset

$$T_m = (t_1, \dots, t_m) \subset T$$

and to use the discrete field on T_m to approximate the continuous field. Thanks to continuity and under certain regularity conditions of T_m, one can show that $\mathbb{P}\{\sup_{T_m} f(t) > b\}/w(b) \to 1$ as $m \to \infty$, i.e. the bias vanishes as the size of the discretization increases. However, it is well understood that this convergence is not uniform in b. The smaller $w(b)$ is, the slower it converges. Thus, the set T_m needs to grow in order to maintain a prefixed relative bias. In fact, as discussed in [3], for any deterministic subset T_m, the size m *must increase* at least polynomially with b to ensure a given relative accuracy. In this paper we introduce a random discretization scheme adapted to (correlated with) the random field f. This adaptive scheme substantially reduces the computation complexity to a constant level.

The high-level excursion of Gaussian random fields is a classic topic in probability. There is a wealth of literature that contains general bounds on $\mathbb{P}\{\sup f(t) > b\}$ as well as sharp asymptotic approximations as $b \to \infty$. For an incomplete list of references; see [10], [11], [12], [18], [19], [25], [28], and [30]. Several methods have been introduced to obtain bounds and asymptotic approximations, each of which imposes different regularity conditions on the random fields. A general upper bound for the tail of max $f(t)$ was developed in [11] and [14], which is known as the Borel–TIS (Borel–Tsirelson–Ibragimov–Sudakov) lemma. For asymptotic results, there are several methods. The double sum method [27] requires an expansion of the covariance function around its global maximum and also locally stationary structure. The Euler–Poincaré characteristics of the excursion set approximation (denoted by $\chi(A_b)$, where A_b is the excursion set) uses the fact that $\mathbb{P}\{M > b\} \approx \mathbb{E}\{\chi(A_b)\}$ and requires the random field to be at least twice differentiable; see [1], [2], [31], [32]. The tube method [29] uses the Karhunen–Loève expansion and imposes differentiability assumptions on the covariance function (fast decaying eigenvalues) and regularity conditions on the random field. The Rice method [7], [8] represents the distribution of M (density function) in an implicit form. For other convex functionals, the exact tail approximation of integrals of exponential functions of Gaussian random fields was developed in [21], [22], and [23]. Recently, Adler *et al*. [4] studied the geometric properties of high-level excursion sets for infinitely divisible non-Gaussian fields as well as the conditional distributions of such properties given the high excursion. In their recent paper, Alder *et al*. [3] studied numerical methods and proposed importance sampling estimators of $w(b)$. In particular, the authors show that the proposed estimator is a fully polynomial randomized approximation scheme, i.e. to achieve the $\varepsilon - \delta$ relative accuracy, the total computation complexity is of order $O(\varepsilon^{-q_1}\delta^{-q_2}|\log w(b)|^q)$; see [26], [33], and [35]. When $w(b)$ is very small, the complexity $O(|\log w(b)|^q)$ could be computationally heavy.

The algorithm in this paper is built upon a change of measure initially introduced in [3]. Nevertheless, the results are nontrivial and are substantial generalizations of [3]. The contributions are as follows. Firstly, we show that the continuous importance sampling estimator proposed in [3] given as in (9) is strongly efficient to compute $w(b)$ for Hölder continuous fields and under mild regularity conditions. This generalizes the results in [3] who established that their relative error grows polynomially fast with b unless the process is twice differentiable for which the exact Slepian model is available. Second, we introduce an adaptive discretization scheme that reduces the overall computational cost to a constant level. This is a substantial improvement on [3] who require the discretization size grow polynomially in b for both differentiable and non-differentiable fields.

The rest of this paper is organized as follows. In Section 2, we present the problem settings and some existing results that we will refer to in the later analysis. In Section 3, we present the Monte Carlo methods and the computed efficiency results. Numerical implementations are included in Section 4. Sections 5 and 6 include the proofs of the theorems.

2. Preliminaries: Gaussian random fields and rare-event simulation

2.1. Gaussian random fields

Throughout this paper, we consider a Gaussian random field living on a d-dimensional compact subset $T \subset R^d$, i.e. for any finite subset $(t_1, \ldots, t_n) \subset T$, $(f(t_1), \ldots, f(t_n))$ is a multivariate Gaussian random vector. For each $s, t \in T$, we define the following functions:

$$\mu(t) = \mathbb{E}\{f(t)\}, \qquad C(s,t) = \operatorname{cov}(f(s), f(t)), \qquad \mu_T = \sup_{t\in T}|\mu(t)|,$$
$$\sigma^2(t) = C(t,t), \qquad \sigma_T^2 = \sup_{t\in T}\sigma^2(t), \qquad r(s,t) = \frac{C(s,t)}{\sigma(s)\sigma(t)}.$$

Let A_γ be the excursion set over the level γ,

$$A_\gamma = \{t \in T : f(t) > \gamma\}$$

and, thus, $w(b) = \mathbb{P}\{A_b \neq \varnothing\}$. Furthermore, we define the concept of a slowly varying function.

Definition 2. A function L is said to be slowly varying at 0 if $\lim_{x\to 0}(L(tx)/L(x)) = 1$ for all $t \in (0, 1)$.

Throughout this paper, we impose the following technical conditions.

(A1) The process $f(t)$ is almost surely continuous in t.

(A2) For some $\alpha_1 \in (0, 2]$, the correlation function satisfies the following local expansion:

$$1 - r(s,t) \sim \Delta_s L_1(|t-s|)|t-s|^{\alpha_1} \quad \text{as } t \to s,$$

where $\Delta_s \in (0,\infty)$ is continuous in s and L_1 is a slowly varying function at 0. Furthermore, there exist nonnegative constants κ_r, β_0, and positive constant $\beta_1 > 0$

https://doi.org/10.1239/aap/1444308882 Published online by Cambridge University Press

satisfying $\beta_0 + \beta_1 \geq \alpha_1$ such that

$$|r(t, t+s_1) - r(t, t+s_2)| \leq \kappa_r L_1(|s_1|)|s_1|^{\beta_0}|s_1 - s_2|^{\beta_1} \quad \text{for } |s_1| \leq |s_2|. \tag{2}$$

(A3) The correlation function is nondegenerate, i.e. $r(s, t) < 1$ for all $s \neq t$.

(A4) The standard deviation $\sigma(t)$ belongs to either of the following two types.

Type 1. Where $\sigma(t) = 1$ for all $t \in T$.

Type 2. Where $\sigma(t)$ has a unique maximum attained at t^* and satisfies the following conditions:

$$|\sigma(t) - \sigma(s)| \leq \kappa_\sigma L_2(|t-s|)|t-s|^{\alpha_2} \quad \text{for all } s, t \in T;$$
$$\sigma(t^*) - \sigma(t) \sim \Lambda L_2(|t^* - t|)|t^* - t|^{\alpha_2} \quad \text{as } t \to t^*,$$

where $\alpha_2 \in (0, 1]$, $\Lambda > 0$, and L_2 is a slowly varying function at 0 such that the limit $\lim_{x\to 0+}(L_1(x)/L_2(x))$ exists.

(A5) There exists $\kappa_\mu > 0$ such that if $\sigma(t)$ is of Type 1 then $|\mu(s) - \mu(s+t)| \leq \kappa_\mu \sqrt{L_1(|t|)}|t|^{\alpha_1/2}$; if $\sigma(t)$ is of Type 2 then $|\mu(s) - \mu(s+t)| \leq \kappa_\mu \sqrt{L_2(|t|)}|t|^{\alpha_2/2}$.

(A6) There exist κ_m and ε small enough, such that $\text{mes}(B(t, \varepsilon) \cap T) \geq \kappa_m \varepsilon^d \omega_d$ for any $t \in T$, where $B(t, \varepsilon)$ is the ε-ball centered around t and ω_d is the volume of the d-dimensional unit ball.

Condition (A2) ensures that the normalized process $(f(t) - \mu(t))/\sigma(t)$ is Hölder continuous with coefficient $\alpha_1/2$. The bound in (2) imposes slightly more conditions on the normalized process. For instance, in the $1 - r(s, t) = |t - s|^{\alpha_1}$ case, we can choose $\beta_0 = \alpha_1 - 1$ and $\beta_1 = 1$ if $\alpha_1 \geq 1$; $\beta_0 = 0$ and $\beta_1 = \alpha_1$ if $0 < \alpha_1 < 1$. Condition (A3) excludes the degenerated case as it is not essential and makes the technical development more concise. Conditions (A4) and (A5) require that the mean and the standard deviation functions are also Hölder continuous. In Condition (A4), we can adjust the constant Λ such that the limit $\lim_{x\to 0+} L_1(x)/L_2(x)$ belongs to the set $\{0, 1, \infty\}$. Condition (A5) ensures that the variation of the mean function is bounded by those of $f(t)$ and $\sigma(t)$. In the later technical development, the analysis is divided into two cases: $\alpha_1 < \alpha_2$ and $\alpha_1 \geq \alpha_2$.

Throughout this paper, we use the following notation for the asymptotics. We write $h(b) = o(g(b))$ if $h(b)/g(b) \to 0$ as $b \to \infty$; $h(b) = O(g(b))$ if $h(b) \leq \kappa g(b)$ for some $\kappa > 0$; $h(b) = \Theta(g(b))$ if $h(b) = O(g(b))$ and $g(b) = O(h(b))$; $h(b) \sim g(b)$ if $h(b)/g(b) \to 1$ as $b \to \infty$.

2.2. Rare-event simulation and importance sampling

2.2.1. *Rare-event simulation.* The research focus of rare-event simulation is on estimating $w = \mathbb{P}\{B\}$, where $\mathbb{P}\{B\} \approx 0$. It is customary to introduce a parameter, say $b > 0$, with a meaningful interpretation from an applied standpoint such that $w(b) \to 0$ as $b \to \infty$. Consider an estimator Z_b such that $\mathbb{E}Z_b = w(b)$. A popular efficiency concept in the rare-event simulation literature is the so-called strong efficiency, see [6], [13], [17], that is defined as follows.

Definition 3. A Monte Carlo estimator Z_b is said to be *strongly efficient* in estimating $w(b)$ if $\mathbb{E}\{Z_b\} = w(b)$ and there exists a $\kappa_0 \in (0, \infty)$ such that

$$\sup_{b>0} \frac{\text{var}(Z_b)}{w^2(b)} < \kappa_0.$$

Strong efficiency measures mean squared error in relative terms for an unbiased estimator. Suppose that a strongly efficient estimator of $w(b)$ has been constructed, denoted by Z_b, and n independent and identically distributed (i.i.d.) replicates of Z_b are generated $Z_b^{(1)}, \dots, Z_b^{(n)}$. Let $\bar{Z}_{b,n} \triangleq (1/n) \sum_{i=1}^n Z_b^{(i)}$ be the averaged estimator that has variance $\mathrm{var}(Z_b)/n$. By means of the Chebyshev inequality, we obtain

$$\mathbb{P}\{|\bar{Z}_{b,n} - w(b)| > \varepsilon w(b)\} \leq \frac{\mathrm{var}(Z_b)}{n\varepsilon^2 w^2(b)}.$$

For any $\delta > 0$, to achieve the $\varepsilon - \delta$ accuracy, we need to generate

$$n = \frac{\mathrm{var}(Z_b)}{\delta\varepsilon^2 w^2(b)} \leq \frac{\kappa_0}{\delta\varepsilon^2}$$

replicates of Z_b. This choice of n is uniform in the rarity parameter b. We will later show that the proposed continuous importance sampling estimator is strongly efficient.

2.2.2. *Importance sampling and variance reduction.* Importance sampling is based on the basic identity,

$$\mathbb{P}\{B\} = \int \mathbf{1}_{\{\omega \in B\}}\, \mathrm{d}\mathbb{P}\{\omega\} = \int \mathbf{1}_{\{\omega \in B\}} \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}Q}(\omega)\, \mathrm{d}Q(\omega) \tag{3}$$

for a measurable set B, where we assume that the probability measure Q is such that $Q\{\cdot \cap B\}$ is absolutely continuous with respect to the measure $\mathbb{P}\{\cdot \cap B\}$. We denote the indicator function by $\mathbf{1}_{\{\cdot\}}$. If we use $\mathbb{E}^Q$ to denote expectation under Q then (3) trivially yields the random variable $Z(\omega) = \mathbf{1}_{\{\omega \in B\}}(\mathrm{d}\mathbb{P}/\mathrm{d}Q)(\omega)$ is an unbiased estimator of $\mathbb{P}\{B\} > 0$ under the measure Q, or symbolically, $\mathbb{E}^Q Z = \mathbb{P}\{B\}$.

A central component lies in the selection of Q in order to minimize the variance of Z. It is easy to verify that if we choose $\mathcal{Q}^*\{\cdot\} = \mathbb{P}\{\cdot \mid B\} = \mathbb{P}\{\cdot \cap B\}/\mathbb{P}\{B\}$ then the corresponding estimator has 0 variance and, thus, it is usually referred to as the *zero-variance change of measure.* However, $\mathcal{Q}^*$ is clearly a change of measure that is of no practical value, since $\mathbb{P}\{B\}$ – the quantity that we are attempting to evaluate in the first place – is unknown. Nevertheless, when constructing a good importance sampling distribution for a family of sets $\{B_b \colon b \geq b_0\}$ for which $0 < \mathbb{P}\{B_b\} \to 0$ as $b \to \infty$, it is often useful to analyze the asymptotic behavior of $\mathcal{Q}^*$ as $\mathbb{P}\{B_b\} \to 0$ in order to guide the construction of a useful Q.

2.2.3. *The change of measure.* We now present a change of measure defined on the continuous sample path space denoted by Q_b. This measure was initially proposed by Alder *et al.* [3]. We should be able to compute the Radon–Nikodym derivative and also be able to simulate the process f under Q_b. We describe the measure Q_b from two aspects. First, we present its Radon–Nikokym derivative with respect to $\mathbb{P}$,

$$\frac{\mathrm{d}Q_b}{\mathrm{d}\mathbb{P}}(f) = \int_T h_b(t) \frac{q_{b,t}(f(t))}{\varphi_t(f(t))}\, \mathrm{d}t, \tag{4}$$

where $h_b(t)$ is a density function on the set T, $q_{b,t}(x)$ is a density function on the real line, and $\varphi_t(x)$ is the density function of $f(t)$ under the measure $\mathbb{P}$ evaluated at $f(t) = x$. We will need to choose $h_b(t)$ and $q_{b,t}(x)$ such that the measure Q_b satisfies the absolute continuity condition to guarantee the unbiasedness.

We will present the specific forms of $h_b(t)$ and $q_{b,t}(x)$ momentarily. Before that, we will complete the description of Q_b by presenting the simulation method of f under Q_b.

https://doi.org/10.1239/aap/1444308882 Published online by Cambridge University Press

Algorithm 1. (*Continuous simulation.*) To generate a random sample path under the measure Q_b, we need a three-step procedure.

Step 1. Generate a random index $\tau \in T$ following the density $h_b(t)$.

Step 2. Conditional on the realization of τ, sample $f(\tau)$ from the density $q_{b,\tau}(x)$.

Step 3. Conditional on the realization of $(\tau, f(\tau))$, generate $\{f(t) : t \neq \tau\}$ from the original conditional distribution $\mathbb{P}\{f \in \cdot \mid f(\tau)\}$.

It is not difficult to verify that the above three-step procedure is consistent with the Radon–Nikodym derivative given as in (4). The process $f(t)$ mostly follows the distribution under $\mathbb{P}$ except at one random location τ where $f(\tau)$ follows an alternative distribution $q_{b,\tau}(x)$. The overall Radon–Nikodym derivative is an average of the likelihood ratio $q_{b,t}(f(t))/\varphi_t(f(t))$ with respect to the density $h_b(t)$.

Now, we present the specific forms of $h_b(t)$ and $q_{b,t}(x)$ for the computation of $w(b)$. For some positive constant a, let γ be

$$\gamma = b - \frac{a}{b}. \tag{5}$$

We choose

$$q_{b,t}(x) = \varphi_t(x)\frac{\mathbf{1}_{\{f(t)>\gamma\}}}{\mathbb{P}\{f(t) > \gamma\}}, \tag{6}$$

i.e. the conditional distribution of $f(t)$ given that $f(t) > \gamma$. The distribution of τ is chosen as

$$h_b(t) = \frac{\mathbb{P}\{f(t) > \gamma\}}{\int_T \mathbb{P}\{f(s) > \gamma\}\, \mathrm{d}s}. \tag{7}$$

The choice of a in (5) does not affect the efficiency results, nor the complexity analysis. To simplify the discussion, we fix a to be unity, i.e.

$$\gamma = b - \frac{1}{b}.$$

The random index τ indicates the location where the distribution of the random field is changed. Furthermore, $q_{b,t}(x)$ is chosen to be the conditional distribution given a high excursion. The index τ basically localizes the maximum of $f(t)$. Thus, as an approximation of the zero-variance change of measure, the distribution $h_b(t)$ should be chosen close to the conditional distribution of the maximum $t_* \triangleq \arg\sup_t f(t)$ given that $f(t_*) > b$. This is our guideline to choose $h_b(t)$. For each $t \in T$, the conditional probability that $f(t) > b$ given $M > b$ is

$$\mathbb{P}\{f(t) > b \mid M > b\} = \frac{\mathbb{P}\{f(t) > b\}}{\mathbb{P}\{M > b\}}.$$

The denominator $\mathbb{P}\{M > b\}$ is free of t and, thus, $\mathbb{P}\{f(t) > b \mid M > b\} \propto \mathbb{P}\{f(t) > b\}$. Our choice of $h_b(t) \propto \mathbb{P}\{f(t) > \gamma\}$ approximates $\mathbb{P}\{f(t) > b \mid M > b\}$ by replacing b with γ mostly for technical convenience. With such choices of $h_b(t)$ and $q_{b,t}(x)$, the Radon–Nikodym takes the following form:

$$\frac{\mathrm{d}Q_b}{\mathrm{d}\mathbb{P}} = \frac{\int_T \mathbf{1}_{\{f(t)>\gamma\}}\, \mathrm{d}t}{\int_T \mathbb{P}\{f(t) > \gamma\}\, \mathrm{d}t} = \frac{\mathrm{mes}(A_\gamma)}{\int_T \mathbb{P}\{f(t) > \gamma\}\, \mathrm{d}t}, \tag{8}$$

https://doi.org/10.1239/aap/1444308882 Published online by Cambridge University Press

where

$$\text{mes}(A_\gamma) = \int \mathbf{1}_{\{t\in A_\gamma\}}\, \mathrm{d}t$$

is the Lebesgue measure of A_γ. According to Fubini's theorem, the denominator of (8) is

$$\int_T \mathbb{P}\{f(t) > \gamma\}\, \mathrm{d}t = \mathbb{E}[\text{mes}(A_\gamma)].$$

Remark 1. For different problems, we may choose different $h_b(t)$ and $q_{b,t}(x)$ to approximate various conditional distributions. For instance, $q_{b,t}(x)$ was chosen to be in the exponential family of $\varphi_t(x)$ in [24] for the derivation of tail approximations of $\int \mathrm{e}^{f(t)}\, \mathrm{d}t$.

2.3. The bias control

In addition to the variance control, we also need to account for the computational effort required to generate Z_b. This issue is especially important in this paper. The random objects in this analysis are continuous processes. For the implementation, we need to use a discrete object to approximate the continuous process. Inevitably, discretization induces bias, though it vanishes as the discretization mesh increases. To ensure the $\varepsilon - \delta$ relative accuracy, the bias needs to be controlled to a level less than $\varepsilon w(b)$.

In [3], it was established that, to ensure a bias of order $\varepsilon w(b)$, the size of the discretization must grow at a polynomial rate of b for both differentiable and non-differentiable fields. The authors also provided an optimality result. For twice differentiable and homogeneous fields, the size of a prefixed/deterministic set T_m must be at least of order $O(b^d)$ so that the bias can be controlled to the level $\varepsilon w(b)$. In this paper, we adopt an adaptive discretization scheme that substantially reduces the necessary size of T_m to a constant.

3. Main results

The main results of this paper consist of a random discretization scheme of T associated with the change of measure Q_b and the efficiency results of the importance sampling estimators and the overall complexity.

3.1. An adaptive discretization scheme and the algorithms

3.1.1. *The continuous estimator and the challenges.* Based on the change of measure Q_b, an unbiased estimator for $w(b)$ is given by

$$Z_b \triangleq \mathbf{1}_{\{M>b\}} \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}Q_b} = \mathbf{1}_{\{M>b\}} \frac{\int_T \mathbb{P}\{f(t) > \gamma\}\, \mathrm{d}t}{\text{mes}(A_\gamma)}. \tag{9}$$

We call Z_b the *continuous estimator*. It is straightforward to obtain $\mathbb{E}_b\{Z_b\} = w(b)$, where we use $\mathbb{E}_b\{\cdot\}$ to denote the expectation under the measure Q_b. The second moment of Z_b is

$$\mathbb{E}_b\{Z_b^2\} = \mathbb{E}_b\left\{\frac{\{\int_T \mathbb{P}\{f(t) > \gamma\}\, \mathrm{d}t\}^2}{\text{mes}^2(A_\gamma)}; M > b\right\},$$

where $f(t)$ is generated from Algorithm 1. We will later show that Z_b (under regularity conditions) is strongly efficient, i.e. $\mathbb{E}_b\{Z_b^2\} = O(w^2(b))$.

For the implementation, we are not able to simulate the continuous field f and, therefore, have to adopt a simulatable estimator, $\hat{Z}_b$, that approximates the continuous estimator Z_b.

https://doi.org/10.1239/aap/1444308882 Published online by Cambridge University Press

A natural approach is to consider the random field on a finite set $T_m = \{t_1, \dots, t_m\} \subset T$ and to use $\mathbb{P}\{\max_{T_m} f(t_i) > b\}$ as an approximation of $w(b) = \mathbb{P}\{\sup_T f(t) > b\}$. The bias is given by

$$\mathbb{P}\Big\{\sup_T f(t) > b\Big\} - \mathbb{P}\Big\{\max_{T_m} f(t) > b\Big\} = \mathbb{P}\{T_m \cap A_b = \varnothing, M > b\}.$$

We explain without rigorous derivation that the above scheme usually induces a heavy computational overhead. To simplify the discussion, we consider that f is a stationary process and its covariance function satisfies the local expansion (slightly abusing the notation)

$$C(t) \triangleq \operatorname{cov}(f(s), f(s+t)) = 1 - |t|^\alpha + o(|t|^\alpha). \tag{10}$$

Then, the process is Hölder continuous with coefficient $\alpha/2$. Under this setting, standard results yield an estimate of the excursion set $\mathbb{E}\{\operatorname{mes}(A_b) \mid M > b\} = \Theta(b^{-2d/\alpha})$. Thanks to stationarity, A_b is approximately uniformly distributed over the domain T.

Note that the bias term $\mathbb{P}\{T_m \cap A_b = \varnothing, M > b\}$ is the probability that T_m does not intersect with A_b. Therefore, if $m \ll b^{2d/\alpha}$, T_m is too sparse such that it is not able to catch the set A_b no matter how T_m is distributed over T. It is necessary to have a lattice of size at least of order $O(b^{2d/\alpha})$. This heuristic calculation was made rigorous for smooth fields in [3]. Thus, the computational complexity to generate the process f on the set T_m grows at a polynomial rate with b. In this paper, we aim to further reduce the discretization size to a constant level while still maintaining the ε-relative bias. For this reason, we propose to randomly sample an appropriate discrete set that is correlated with f.

3.1.2. *A closer look at the excursion set* A_γ. The proposed adaptive discretization scheme is closely associated with the three-step simulation procedure. Of the three steps in Algorithm 1, Step 1 and Step 2 are implementable. It is Step 3, generating $\{f(t) \colon t \neq \tau\}$ conditional on $(\tau, f(\tau))$, that requires discretization. In order to estimate $w(b)$ and to generate the estimator Z_b, we need only to simulate the random indicator $\mathbf{1}_{\{M>b\}}$ and the volume of the excursion set $\operatorname{mes}(A_\gamma)$ conditional on $(\tau, f(\tau))$. The term $\int_T \mathbb{P}\{f(t) > \gamma\}\,\mathrm{d}t$ is a deterministic number that can be computed via routine numerical methods.

In what follows, we focus on the simulation and approximation of $\mathbf{1}_{\{M>b\}}$ and $\operatorname{mes}(A_\gamma)$. For the purposes of illustration, we discuss the stationary case with covariance function satisfying the expansion in (10). We define $\zeta = b^{2/\alpha}$ and the normalized process

$$g(t) = b\bigg(f\bigg(\tau + \frac{t}{\zeta}\bigg) - b\bigg). \tag{11}$$

Note that $b(f(\tau) - \gamma)$ asymptotically follows an exponential distribution. Conditional on $f(\tau) = \gamma + z/b$ the g process has expectation $\mathbb{E}_b\{g(t) \mid f(\tau) = \gamma + z/b\} = z - 1 - (1 + o(1))|t/\zeta|^\alpha[b^2 + (z-1)]$. For all $z = o(b^2)$, we have

$$\mathbb{E}_b\bigg\{g(t) \mid f(\tau) = \gamma + \frac{z}{b}\bigg\} = z - 1 - (1 + o(1))|t|^\alpha \quad \text{as } b \to \infty.$$

In addition, the covariance of $g(t)$ is $\operatorname{cov}(g(s), g(t)) = (|s|^\alpha + |t|^\alpha - |s-t|^\alpha) + o(1)$ where $o(1) \to 0$ as $b \to \infty$. Therefore, $g(t)$ converges in distribution to a Gaussian process with the above mean and covariance function. In addition, $f(\tau + t/\zeta) \geq \gamma$ if and only if $g(t) > -1$. The excursion set A_γ can be written as

$$A_\gamma = \tau + \zeta^{-1} A_{-1}^g \triangleq \{\tau + \zeta^{-1} t \colon t \in A_{-1}^g\},$$

where $A^g_{-1} = \{t : g(t) > -1\}$. Note that the process $g(t)$ is a Gaussian process with standard deviation $O(|t|^{\alpha/2})$ and a negative drift of order $O(-|t|^{\alpha})$. Therefore, in expectation, $g(t)$ goes below -1 when $z \ll |t|^{\alpha}$, where z is asymptotically an exponential random variable. Thus, the excursion set A^g_{-1} is of order $O(1)$. Furthermore, A_γ is a random set within $O(\zeta^{-1})$ distance from the random index τ. The volume $\mathrm{mes}(A_\gamma)$ is of order $O(\zeta^{-d})$. This heuristic calculation is well understood; see [5], [9]. The above discussion quantifies the intuition that τ localizes the global maximum of f. It also localizes the excursion set A_γ. Therefore, upon considering approximating/computing $\mathrm{mes}(A_\gamma)$ and $\mathbf{1}_{\{M>b\}}$, we should focus on the region around τ.

Conditional on a specific realization of the process f, we formulate the approximation of $\mathrm{mes}(A_\gamma)$ as an estimation problem. The ratio $\mathrm{mes}(A_\gamma)/\mathrm{mes}(T) \in [0, 1]$ corresponds to the following probability:

$$\frac{\mathrm{mes}(A_\gamma)}{\mathrm{mes}(T)} = \mathbb{P}\{U \in A_\gamma\},$$

where U is a uniform random variable on the set T with respect to the Lebesgue measure. Estimating $\mathrm{mes}(A_\gamma)$ constitutes another rare-event simulation problem.

3.1.3. *An adaptive discretization scheme.* Based on the understanding of the excursion set A_γ, we construct a discretization scheme adaptive to the realization of τ. To proceed, we provide the general form of ζ in the presence of slowly varying functions

$$\zeta \triangleq \max\{|s|^{-1} : L_1(|s|)|s|^{\alpha_1} \geq b^{-2} \text{ or } L_2(|s|)|s|^{\alpha_2} \geq b^{-2}\}.$$

In the case of constant variance, we formally define $\alpha_2 = \infty$ and, thus, ζ is defined as $\zeta \triangleq \max\{|s|^{-1} : L_1(|s|)|s|^{\alpha_1} \geq b^{-2}\}$. We further define two other scale factors,

$$\zeta_i \triangleq \max\{|s|^{-1} : L_i(|s|)|s|^{\alpha_i} \geq b^{-2}\}, \qquad i = 1, 2. \tag{12}$$

It is straightforward to verify that $\zeta = \max(\zeta_1, \zeta_2)$. Consider an isotropic distribution (centered around 0) with density $k(t)$, i.e. $k(t) = k(s)$ if $|s| = |t|$. We choose $k(t)$ to be reasonably heavy-tailed such that for some $\varepsilon_1 > 0$,

$$k(t) \sim |t|^{-d-\varepsilon_1} \quad \text{as } t \to \infty.$$

In addition there exists a $\kappa_1 > 0$ such that $k(t) \leq \kappa_1$ for all t. For instance, we can choose $k(t)$ to be, but not necessarily restricted to, the multivariate t-distribution. Furthermore, conditional on τ, we define the rescaled density

$$k_{\tau,\zeta}(t) = \zeta^d k(\zeta(t - \tau)) \tag{13}$$

that centers around τ and has scale ζ^{-1}. We construct a τ-adapted random subset of T by generating i.i.d. random variables from the density $k_{\tau,\zeta}(t)$, denoted by $t_1, \ldots, t_m$. Then, define

$$\widehat{\mathrm{mes}}(A_\gamma) \triangleq \frac{1}{m} \sum_{i=1}^{m} \frac{\mathbf{1}_{\{f(t_i)>\gamma\}}}{k_{\tau,\zeta}(t_i)} \tag{14}$$

that is an unbiased estimator of $\mathrm{mes}(A_\gamma)$ in the sense that for each realization of f,

$$\mathbb{E}_{\tau,\zeta}\{\widehat{\mathrm{mes}}(A_\gamma) \mid f\} = \mathrm{mes}(A_\gamma),$$

where $\mathbb{E}_{\tau,\zeta}\{\cdot \mid f\}$ is the expectation with respect to $t_1, \ldots, t_m$ under the density $k_{\tau,\zeta}$ for a particular realization of f. Notationally, if $t_i \notin T$ then $\mathbf{1}_{\{f(t_i)>\gamma\}} = 0$.

https://doi.org/10.1239/aap/1444308882 Published online by Cambridge University Press

Similar to the approximation of $\mathrm{mes}(A_\gamma)$, we use the same τ-adapted random subset to approximate $\mathbf{1}_{\{M>b\}}$, i.e.

$$\mathbf{1}_{\{\max_{i=1}^m f(t_i)>b\}} \approx \mathbf{1}_{\{M>b\}}\,.$$

Based on the above discussions, we present the final algorithm.

Algorithm 2. The algorithm consists of the following steps.

Step 1. Generate a random index $\tau \in T$ following the density $h_b(t)$ in (7).

Step 2. Conditional on the realization of τ, sample $f(\tau)$ from $q_{b,t}(x)$ in (6).

Step 3. Conditional on the realization of τ, generate i.i.d. random indices $t_1, \ldots, t_m$ following density $k_{\tau,\zeta}(t)$.

Step 4. Conditional on the realization of $(\tau, f(\tau))$, generate multivariate normal random vector $(f(t_1), \ldots, f(t_m))$ from the original/nominal conditional distribution of $\mathbb{P}\{\cdot \mid f(\tau)\}$.

Step 5. Output

$$\hat{Z}_b = \frac{\mathbf{1}_{\{\max_{i=1}^m f(t_i)>b\}}}{\widehat{\mathrm{mes}}(A_\gamma)} \int_T \mathbb{P}\{f(t) > \gamma\}\,\mathrm{d}t,$$

where $\widehat{\mathrm{mes}}(A_\gamma)$ is given as in (14).

We will call $\hat{Z}_b$ the *discrete estimator*.

3.2. The main results

We present the efficiency results of the proposed algorithms.

Theorem 1. *Consider a Gaussian random field f that satisfies conditions (A1)–(A6) of Definition 2. Let Z_b be given as in (9) and Algorithm 1. Then, Z_b is strongly efficient in estimating $w(b)$, i.e. there exists κ_0 such that*

$$\mathbb{E}_b\{Z_b^2\} \le \kappa_0 w^2(b) \quad \textit{for all } b > 0.$$

Theorem 2. *Consider a Gaussian random field f that satisfies conditions (A1)–(A6) of Definition 2. Let $\hat{Z}_b$ be the estimator given by Algorithm 2. There exists $\lambda > 0$ such that for any $\varepsilon > 0$ if we choose $m = \lambda\varepsilon^{-d(2/\min(\alpha_1,\alpha_2)+2/\beta_1)}$ then*

$$|\mathbb{E}_b\{\hat{Z}_b\} - w(b)| \le \varepsilon w(b) \quad \textit{for all } b > 0.$$

Furthermore, there exists κ_0 such that

$$\mathbb{E}_b\{\hat{Z}_b^2\} \le \kappa_0 w^2(b).$$

With the above results, we generate n i.i.d. replicates of $\hat{Z}_b$, denoted by $\hat{Z}_b^{(1)}, \ldots, \hat{Z}_b^{(n)}$, with m chosen as in Theorem 2 such that the averaged estimator, $(1/n)\sum_{i=1}^n \hat{Z}_b^{(i)}$, has its bias bounded by $\varepsilon w(b)/2$ and its variance bounded by $\kappa_0 w^2(b)/n$. To achieve ε relative error with $(1-\delta)$ confidence, we need to choose $n = 4\kappa_0/\varepsilon^2\delta$, i.e.

$$\mathbb{P}\left\{\left|\frac{1}{m}\sum_{i=1}^n \hat{Z}_b^{(i)} - w(b)\right| > \varepsilon w(b)\right\} < \delta.$$

The total computational complexity is of order $O(m^3\varepsilon^{-2}\delta^{-1})$, where m^3 is the complexity of the Cholesky decomposition of the covariance matrix for the generation of an m-dimensional Gaussian random vector.

4. Numerical analysis

We present four numerical examples in order to demonstrate the performance of our algorithm. First, we consider a one-dimensional Gaussian field whose tail probability is known in a closed form. For the discretization, we deploy $m = 20$ points when $d = 1$, and 40 points when $d = 2$. To make sure that the bias is small enough, we have performed the simulations with ten times more points and the results did not change substantially. We only report the results for the cases with fewer points to illustrate the efficiency.

Example 1. Consider $f(t) = X\cos t + Y\sin t$, $T = [0, \frac{3}{4}]$, where X and Y are independent standard Gaussian variables. The probability $\mathbb{P}\{\sup_{t\in T} f(t) > b\}$ is known in closed form (see [1]),

$$\mathbb{P}\Big\{\sup_{0\le t\le 3/4} f(t) > b\Big\} = 1 - \Phi(b) + \frac{3}{8\pi}\mathrm{e}^{-b^2/2}. \tag{15}$$

In Table 1, we list the simulation results.

The following three examples consider random fields over a two-dimensional square.

Example 2. Consider a mean 0, unit variance, stationary, and smooth Gaussian field over $T = [0, 1]^2$ with covariance function

$$C(t) = \mathrm{e}^{-|t|^2}.$$

In Table 2, we list the simulation results.

Example 3. Consider a continuous inhomogenous Gaussian field on $T = [0, 1]^2$ with mean and covariance function

$$\mu(t) = 0.1t_1 + 0.1t_2, \qquad C(s, t) = \mathrm{e}^{-|t-s|^2}.$$

In Table 3, we list the simulation results.

Example 4. Consider the continuous Gaussian field living on $T = [0, 1]^2$ with mean and covariance function

$$\mu(t) = 0.1t_1 + 0.1t_2, \qquad C(s, t) = \mathrm{e}^{-|t-s|/4}.$$

In Table 4, we list the simulation results.

Table 1: Simulation results for the cosine process in Example 1, where $n = 1000$ and $m = 20$, and where $k(t)$ is chosen to be the density function of the t-distribution with degrees of freedom 3. The 'True value' is calculated from (15), 'Est.' is the estimated value, 'Std. dev.' is the standard deviation of the averaged Monte Carlo estimator over n i.i.d. samples, and the 'Coefficient of variation' is the ratio between the standard deviation of a single Monte Carlo estimator and its expectation.

b	True value	Est.	Std. dev.	Coefficient of variation
3	2.68×10^{-3}	2.55×10^{-3}	1.09×10^{-4}	1.35
4	7.17×10^{-5}	7.17×10^{-5}	3.22×10^{-6}	1.42
5	7.31×10^{-7}	7.33×10^{-7}	3.41×10^{-8}	1.47
6	2.80×10^{-9}	2.84×10^{-9}	1.35×10^{-10}	1.51
7	4.01×10^{-12}	4.07×10^{-12}	1.98×10^{-13}	1.54

https://doi.org/10.1239/aap/1444308882 Published online by Cambridge University Press

TABLE 2: Simulation results for Example 2, where $n = 1000$, $m = 40$, and $k(t) = (25/32\pi)(1 + 0.64|t|^2)^{-3}$.

b	Est.	Std. dev.	Coefficient of variation
3	9.32×10^{-3}	3.63×10^{-4}	1.23
4	3.39×10^{-4}	1.51×10^{-5}	1.41
5	4.20×10^{-6}	1.71×10^{-7}	1.28
6	1.93×10^{-8}	8.15×10^{-10}	1.33
7	3.25×10^{-11}	1.27×10^{-12}	1.23
8	1.87×10^{-14}	7.11×10^{-16}	1.20

TABLE 3: Simulation results for Example 3, where $n = 1000$, $m = 40$, and $k(t)$ is the same as that of Example 2.

b	Est.	Std. dev.	Coefficient of variation
3	1.25×10^{-2}	5.61×10^{-4}	1.42
4	4.95×10^{-4}	1.95×10^{-5}	1.24
5	7.16×10^{-6}	2.80×10^{-7}	1.24
6	3.51×10^{-8}	1.36×10^{-9}	1.22
7	6.69×10^{-11}	2.72×10^{-12}	1.29
8	4.50×10^{-14}	1.91×10^{-15}	1.34

TABLE 4: Simulation results for Example 4, where $n = 1000$, $m = 40$, and $k(t) = (1/8\pi)(1 + |t|^2)^{-3}$.

b	Est.	Std. dev.	Coefficient of variation
3	1.35×10^{-2}	6.63×10^{-4}	1.55
4	7.40×10^{-4}	4.36×10^{-5}	1.86
5	1.54×10^{-5}	7.53×10^{-7}	1.55
6	9.93×10^{-8}	5.23×10^{-9}	1.66
7	2.87×10^{-10}	1.33×10^{-11}	1.47
8	2.60×10^{-13}	1.41×10^{-14}	1.71

For all the examples, the ratios of standard error over the estimated value do not increase as b increases. This is consistent with our theoretical analysis. Also note that m does not increase as the level increases, which reduces the computational complexity significantly. Overall, the numerical estimates are very accurate.

5. Proof of Theorem 1

Throughout the proof, we will use κ as a generic notation to denote large and not-so-important constants whose value may vary from place to place. Similarly, we use ε_0 as a generic notation for small positive constants.

The first result we cite is the Borel–TIS inequality (see [2], [11], [14]) that will be used very often in our technical development.

Proposition 1. *Let $f(t)$ be a centered Gaussian process almost surely bounded in T. Then, $\mathbb{E}\{\sup_{t\in T} f(t)\} < \infty$ and*

$$P\Big(\sup_{t\in T} f(t) - \mathbb{E}\{\sup_{t\in T} f(t)\} \ge b\Big) \le \exp\Big(-\frac{b^2}{2\sigma_T^2}\Big).$$

In order to show strong efficiency, we need to establish a lower bound of the probability

$$w(b) = \mathbb{E}_b\bigg\{\frac{1}{\operatorname{mes}(A_\gamma)}; M > b\bigg\} \int_T \mathbb{P}\{f(t) > \gamma\}\,\mathrm{d}t$$

and an upper bound of the second moment

$$E_b(Z_b^2) = \mathbb{E}_b\bigg\{\frac{1}{\operatorname{mes}^2(A_\gamma)}; M > b\bigg\}\bigg[\int_T \mathbb{P}\{f(t) > \gamma\}\,\mathrm{d}t\bigg]^2.$$

The central analysis lies in the following two quantities:

$$I_1 = \mathbb{E}_b\bigg\{\frac{1}{\operatorname{mes}(A_\gamma)}; M > b\bigg\}, \qquad I_2 = \mathbb{E}_b\bigg\{\frac{1}{\operatorname{mes}^2(A_\gamma)}; M > b\bigg\}.$$

We will show that there exist constants κ and ε_0 such that

$$I_1 \ge \varepsilon_0 \zeta^d, \qquad I_2 \le \kappa \zeta^{2d}.$$

If these inequalities are proved then $\limsup_{b\to\infty} I_2/I_1^2 < \infty$ is in place and we conclude our proof for Theorem 1. For the rest of the proof, we establish these two inequalities.

To proceed, we describe the conditional Gaussian random field given $f(\tau)$. First, if we write $f(\tau) = \gamma + z/b$ then (conditional on τ) z asymptotically follows an exponential distribution with expectation $\sigma^2(\tau)$. Conditional on $f(\tau) = \gamma + z/b$, let

$$f(t+\tau) = \mathbb{E}\bigg\{f(t+\tau) \mid f(\tau) = \gamma + \frac{z}{b}\bigg\} + f_0(t), \tag{16}$$

where $f_0(t)$ is a zero-mean Gaussian process. By means of conditional Gaussian calculation, the conditional mean and conditional covariance function are given by

$$\begin{aligned} \mu_\tau(t) &= \mathbb{E}\bigg\{f(t+\tau) \mid f(\tau) = \gamma + \frac{z}{b}\bigg\} \\ &= \mu(t+\tau) + \frac{\sigma(\tau+t)}{\sigma(\tau)} r(\tau+t,\tau)\Big(\gamma + \frac{z}{b} - \mu(\tau)\Big), \qquad (17) \\ C_0(s,t) &= \operatorname{cov}(f_0(s), f_0(t)) \\ &= \sigma(\tau+s)\sigma(\tau+t)[r(s+\tau,t+\tau) - r(\tau+t,\tau) r(\tau+s,\tau)]. \end{aligned}$$

The next lemma controls the conditional variance.

Lemma 1. *Under conditions (A1)–(A6) of Definition 2, there exists a constant $\lambda_1 > 0$ such that, for all $\tau \in T$ and large enough b, the following statements hold.*

(i) *For all $t+\tau \in T$,*

$$C_0(t,t) \le \lambda_1 L_1(|t|)|t|^{\alpha_1}.$$

https://doi.org/10.1239/aap/1444308882 Published online by Cambridge University Press

(ii) *For* $s, t \in T$,

$$\mathrm{var}(f_0(s) - f_0(t)) \le \lambda_1 \max(L_1(|t-s|)|t-s|^{\alpha_1}, L_2(|t-s|)|t-s|^{\alpha_2}).$$

(iii) *For any* $\varepsilon > 0$, *there exists* $\delta > 0$ *(independent of* b*) such that for each* t,

$$\mathbb{E}\Big\{\sup_{|s-t|\le\delta\zeta^{-1}} f_0(s)\Big\} \le \frac{\varepsilon}{b}.$$

The proofs for (i) and (ii) are an application of conditions (A2), (A3), and (A6) of Definition 2 and elementary calculations. Part (iii) is a direct corollary of (ii) and Dudley's entropy bound [16, Theorem 1.1]. We omit the detailed derivations. We proceed to the analysis of I_1 and I_2 by considering the Type 1 and Type 2 standard deviation functions (condition (A4)) separately.

Here, we provide the proof only when $\sigma(t)$ is of Type 1 in condition (A4), i.e. a constant variance. The proof of the non-constant case is similar; see [20]. The constant variance case corresponds to $\alpha_2 = \infty$. The scaling factor is given by $\zeta = \zeta_1$. We aim to show that $I_2 \le \kappa\zeta_1^{2d}$ and $I_1 \ge \varepsilon_0\zeta_1^d$.

5.1. The I_2 term

For some $y_0 > 0$ chosen to be sufficiently small (independent of b) and to be determined in the later analysis, the I_2 term is bounded by

$$\mathbb{E}_b\Big\{\frac{1}{\mathrm{mes}^2(A_\gamma)}; M > b\Big\} \le y_0^{-2d}\zeta_1^{2d} + \mathbb{E}_b\Big\{\frac{1}{\mathrm{mes}^2(A_\gamma)}; \mathrm{mes}(A_\gamma) < y_0^d\zeta_1^d, M > b\Big\}. \quad (18)$$

To control the second term of the above inequality, we need to provide a bound on the following tail probability for $0 < y < y_0$:

$$Q_b\{\mathrm{mes}(A_\gamma) < y^d\zeta_1^{-d}, M > b\} = \int \mathbb{P}\Big\{\mathrm{mes}(A_\gamma) < y^d\zeta_1^{-d}, M > b \mid f(\tau) = \gamma + \frac{z}{b}\Big\} \times h_b(\tau)\frac{q_{b,\tau}(\gamma + z/b)}{b}\,\mathrm{d}\tau\,\mathrm{d}z. \quad (19)$$

The probability inside the integral is with respect to the original measure $\mathbb{P}$ because, conditional on $f(\tau)$, $f(t)$ follows the original conditional distribution. We develop bounds for $\mathbb{P}\{\mathrm{mes}(A_\gamma) < y^d\zeta_1^{-d}, M > b \mid f(\tau) = \gamma + z/b\}$ under two situations: $z > 1$ and $0 < z \le 1$.

Situation 1. Where $z > 1$. Define a constant $c_d = \omega_d^{-1/d}$, where ω_d is the volume of the d-dimensional unit ball. The event $\{\mathrm{mes}(A_\gamma) < y^d\zeta_1^{-d}\}$ implies the event $\{\inf_{|t-\tau|\le c_d y\zeta_1^{-1}} f(t) \le \gamma\}$. Otherwise, if $\{\inf_{|t-\tau|\le c_d y\zeta_1^{-1}} f(t) > \gamma\}$ then $\{|t-\tau| \le c_d y\zeta_1^{-1}\} \subseteq A_\gamma$ and $\mathrm{mes}(A_\gamma) \ge y^d\zeta_1^{-d}$. Thus, we have the bound

$$\mathbb{P}\Big\{\mathrm{mes}(A_\gamma) \le y^d\zeta_1^{-d}, M > b \mid f(\tau) = \gamma + \frac{z}{b}\Big\} \le \mathbb{P}\Big\{\inf_{|t-\tau|\le c_d y\zeta_1^{-1}} f(t) \le \gamma \mid f(\tau) = \gamma + \frac{z}{b}\Big\}.$$

Using the representation in (16), the right-hand side of the above probability can be expressed as

$$\mathbb{P}\Big\{\inf_{|t|\le c_d y\zeta_1^{-1}} f_0(t) + \mu_\tau(t) \le \gamma\Big\}. \quad (20)$$

https://doi.org/10.1239/aap/1444308882 Published online by Cambridge University Press

Note that $\mu_\tau(0) = \gamma + z/b > \gamma + 1/b$. For the constant variance case, (17) can be written as

$$\mu_\tau(t) = \mu(t+\tau) + r(\tau + t, \tau)\biggl(\gamma + \frac{z}{b} - \mu(\tau)\biggr). \tag{21}$$

According to condition (A5), we have $|\mu_\tau(t) - \mu_\tau(0)| = O(bL_1(t)|t|^{\alpha_1}) + O(\sqrt{L_t(t)|t|^{\alpha_1}})$. According to the choice of ζ_1 in (12), we have

$$bL_1(t)|t|^{\alpha_1} \le \kappa b L_1(c_d y \zeta_1^{-1}) y^{\alpha_1} \zeta_1^{-\alpha_1} = \kappa b^{-1} \frac{L_1(c_d y \zeta_1^{-1})}{L_1(\zeta_1^{-1})} y^{\alpha_1} \quad \text{for } |t| \le c_d y \zeta_1^{-1},$$

According to [20, Lemma 5(i)], the ratio $L_1(c_d y \zeta_1^{-1})/L_1(\zeta_1^{-1})$ varies slower than any polynomial of y. Thus, we have

$$|\mu_\tau(t) - \mu_\tau(0)| \le y^{\alpha_1/2} b^{-1}. \tag{22}$$

By choosing y small, we have

$$\mu_\tau(t) \ge \gamma + \frac{1}{2b} \quad \text{for } |t| \le c_d y \zeta_1^{-1}. \tag{23}$$

Furthermore, by Lemma 1(i) the conditional variance is $C_0(t,t) \le \lambda_1 L_1(c_d y \zeta_1^{-1}) c_d^{\alpha_1} y^{\alpha_1} \zeta_1^{-\alpha_1}$. Following the same argument as that of (22), we obtain

$$C_0(t,t) = O(y^{\alpha_1/2} b^{-2}) \quad \text{for } |t| \le c_d y \zeta_1^{-1}. \tag{24}$$

By Lemma 1(iii), $\mathbb{E}\{\sup_{|t| \le c_d y_0 \zeta_1^{-1}} b f_0(t)\} = o(1)$ as $y_0 \to 0$. So we can choose y_0 small enough such that

$$\mathbb{E}\Bigl\{\sup_{|t| \le c_d y_0 \zeta_1^{-1}} f_0(t)\Bigr\} \le \frac{1}{4b}. \tag{25}$$

By the Borel–TIS inequality (see Proposition 1), (20), (23), (24), and (25), there exists a positive constant ε_0, such that

$$\begin{aligned}\mathbb{P}\Bigl\{\mathrm{mes}(A_\gamma) \le y^d \zeta_1^{-d}, M > b \mid f(\tau) = \gamma + \frac{z}{b}\Bigr\} &\le \mathbb{P}\Bigl\{\inf_{|t| \le c_d y \zeta_1^{-1}} |f_0(t)| > \frac{1}{2b}\Bigr\} \\ &\le \exp(-\varepsilon_0 y^{-\alpha_1/2}).\end{aligned}$$

Situation 2. Where $0 < z \le 1$. With y_0 defined to satisfy (23) and (25), we let $c = c_d y_0$ and define a finite subset $\tilde{T} = \{t_1, \dots, t_N\} \subset T$ such that

1. For $i \ne j$, $|t_i - t_j| \ge c/2\zeta_1$.
2. For any $t \in T$, there exists i, such that $|t - t_i| \le c/\zeta_1$.

Furthermore, let

$$B_i = \{t \in T : |t - t_i| \le c\zeta_1^{-1}\} \quad \text{for } i \in \{1, 2, \dots, N\}$$

and, thus, $\bigcup_i B_i = T$. Note that

$$\begin{aligned}&\mathbb{P}\Bigl\{\mathrm{mes}(A_\gamma) \le y^d \zeta_1^{-d}, M > b \mid f(\tau) = \gamma + \frac{z}{b}\Bigr\} \\ &\quad \le \sum_{i=1}^N \mathbb{P}\Bigl\{\mathrm{mes}(A_\gamma) \le y^d \zeta_1^{-d}, \sup_{t \in B_i} f(t) > b \mid f(\tau) = \gamma + \frac{z}{b}\Bigr\}.\end{aligned}$$

https://doi.org/10.1239/aap/1444308882 Published online by Cambridge University Press

With c_d as previously chosen, each of the summands in the above equation is bounded by

$$\mathbb{P}\left\{\operatorname{mes}(A_\gamma) \le y^d \zeta_1^{-d}, \sup_{t\in B_i} f(t) > b \mid f(\tau) = \gamma + \frac{z}{b}\right\}$$
$$\le \mathbb{P}\left\{\sup_{t\in B_i, |s-t|\le c_d y \zeta_1^{-1}} |f(t) - f(s)| > \frac{1}{b}, \sup_{t\in B_i} f(t) > b \mid f(\tau) = \gamma + \frac{z}{b}\right\}. \tag{26}$$

The above inequality is derived from the following argument. Suppose that $f(t_0) > b$. In order to have $\operatorname{mes}(A_\gamma) \le y^d \zeta_1^{-d}$, with the same argument as that of (20), we must have $\inf_{|s-t_0|\le c_d y \zeta_1^{-1}} f(s) \le b - 1/b$. Thus, there exist $|s_0 - t_0| \le c_d y \zeta_1^{-1}$ and $|f(s_0) - f(t_0)| > 1/b$. Therefore, the event $\{\operatorname{mes}(A_\gamma) > y^d \zeta_1^{-d}, \sup_{t\in B_i} f(t) > b\}$ is a subset of $\{\sup_{t\in B_i, |s-t|\le c_d y \zeta_1^{-1}} |f(t) - f(s)| > 1/b, \sup_{t\in B_i} f(t) > b\}$, which yields (26).

Selecting $\delta_0, \delta_1 > 0$ small enough, and λ large enough, we provide a bound for (26) under the following four cases.

Case 1. When $0 < |t_i - \tau| < y^{-\delta_0} \zeta_1^{-1}$.

Case 2. When $y^{-\delta_0} \zeta_1^{-1} < |t_i - \tau| < \delta_1$.

Case 3. When $|t_i - \tau| \ge \delta_1$, $y < b^{-\lambda}$.

Case 4. When $|t_i - \tau| \ge \delta_1$, $y \ge b^{-\lambda}$.

To facilitate the discussion, define $x_i \triangleq \zeta_1 |t_i - \tau|$.

Case 1. When $0 < |t_i - \tau| < y^{-\delta_0} \zeta_1^{-1}$. We provide a bound for (26) via the conditional representation (16) and the calculation in (17). According to conditions (A2) and (A5) of Definition 2, for $|t - s| \le c_d y \zeta_1^{-1}$ and $t \in B_i$, we have

$$|\mu_\tau(t) - \mu_\tau(s)| \le \kappa_\mu \zeta_1^{-\alpha_1/2} \sqrt{L_1\left(\frac{y}{\zeta_1}\right)} y^{\alpha_1/2} + \kappa_r (x_i + 1)^{\beta_0} L_1((x_i + 1)\zeta_1^{-1}) y^{\beta_1} \zeta_1^{-\alpha_1} b.$$

According to the definition of ζ_1 in (12) and [20, Lemma 5(i)], the above equation can be bounded by

$$|\mu_\tau(t) - \mu_\tau(s)| \le \frac{2\kappa_\mu y^{\alpha_1/4} + 2\kappa_r y^{-\delta_0\beta_0+\beta_1-\varepsilon_0}}{b}.$$

We choose δ_0 small such that it is further bounded by $|\mu_\tau(t) - \mu_\tau(s)| \le \kappa y^{\varepsilon_0} b^{-1}$ for some possibly different $\varepsilon_0 > 0$. Furthermore, we choose $y_0 > 0$ small enough such that for $0 < y < y_0$ and $|s - t| < c_d y \zeta_1^{-1}$,

$$|\mu_\tau(s) - \mu_\tau(t)| \le \frac{1}{2b}.$$

The above inequality provides a bound on the variation of the mean function over the set B_i when t_i is within $y^{-\delta_0} \zeta_1^{-1}$ distance close to τ. The probability in (26) can be bounded by

$$(26) \le \mathbb{P}\left\{\sup_{t\in B_i, |t-s|\le c_d y \zeta_1^{-1}} |f_0(t) - f_0(s)| > \frac{1}{2b}\right\}.$$

Note that, by Lemma 1(ii), for $|s - t| < c_d y \zeta_1^{-1}$ and for $y < y_0$, we have

$$\operatorname{var}(f_0(s) - f_0(t)) \le \lambda_1 \frac{L_1(c_d y \zeta_1^{-1})}{L_1(\zeta_1^{-1})} y^{\alpha_1} b^{-2} = O(y^{\alpha_1/2} b^{-2}).$$

We apply the Borel–TIS inequality (see Proposition 1) to the double-indexed Gaussian field $\xi(s,t) \triangleq f_0(s) - f_0(t)$ and show that there exists a positive constant ε_0 such that

$$\begin{aligned}
&\mathbb{P}\bigg\{\frac{1}{\operatorname{mes}(A_\gamma)} > y^{-d}\zeta_1^d, \sup_{t\in B_i} f(t) > b \mid f(\tau) = \gamma + \frac{z}{b}\bigg\} \\
&\qquad \le \mathbb{P}\bigg\{\sup_{t\in B_i, |t-s|\le c_d y\zeta_1^{-1}} |f_0(t) - f_0(s)| > \frac{1}{2b}\bigg\} \\
&\qquad \le \exp(-\varepsilon_0 y^{-\alpha_1/2}).
\end{aligned}$$

Combining all the B_i such that $|t_i - \tau| < y^{-\delta_0}\zeta_1^{-1}$, we obtain

$$\begin{aligned}
&\mathbb{P}\bigg\{\frac{1}{\operatorname{mes}(A_\gamma)} > y^{-d}\zeta_1^d, \sup_{|t-\tau|\le y^{-\delta_0}\zeta_1^{-1}} f(t) > b \mid f(\tau) = \gamma + \frac{z}{b}\bigg\} \\
&\qquad = O(y^{-\delta_0 d}\exp(-\varepsilon_0 y^{-\alpha_1/2})) \\
&\qquad \le \exp(-y^{-\varepsilon_0}),
\end{aligned}$$

possibly redefining ε_0.

Case 2. When $y^{-\delta_0}\zeta_1^{-1} < |t_i - \tau| < \delta_1$. For this case, we implicitly require that $y^{-\delta_0}\zeta_1^{-1} < \delta_1$. For $t \in B_i$ and y small enough, we have

$$\begin{aligned}
&\mathbb{P}\bigg\{\sup_{t\in B_i, |s-t|\le c_d y\zeta_1^{-1}} |f(t) - f(s)| > \frac{1}{b}, \sup_{t\in B_i} f(t) > b \mid f(\tau) = \gamma + \frac{z}{b}\bigg\} \\
&\qquad \le \mathbb{P}\bigg\{\sup_{t\in B_i} f(t) > b \mid f(\tau) = \gamma + \frac{z}{b}\bigg\}.
\end{aligned}$$

According to condition (A2) and (21), we have the bound

$$\mu_\tau(t) \le b - \frac{\Delta_\tau}{2}\frac{L_1(x_i\zeta_1^{-1})}{L_1(\zeta_1^{-1})} x_i^{\alpha_1} b^{-1} \quad \text{for } \tau + t \in B_i.$$

According to Lemma 1 and the definition of ζ_1, the variance of $f_0(t)$ is controlled by

$$C_0(t,t) \le 2\lambda_1 \frac{L_1(x_i\zeta_1^{-1})}{L_1(\zeta_1^{-1})} x_i^{\alpha_1} b^{-2}.$$

According to Proposition 1 and [20, Lemma 5(ii)], we have $L_1(x_i\zeta_1^{-1})/L_1(\zeta_1^{-1})x_i^{\alpha_1} > x_i^{\alpha_1/2}$ for $y^{-\delta_0} < x_i < \delta_1\zeta_1$. It follows that

$$\begin{aligned}
\mathbb{P}\bigg\{\sup_{t\in B_i} f(t) > b \mid f(\tau) = \gamma + \frac{z}{b}\bigg\} &\le \mathbb{P}\bigg\{\sup_{t+\tau\in B_i} f_0(t) > \frac{\Delta_\tau}{2}\frac{L_1(x_i\zeta_1^{-1})}{L_1(\zeta_1^{-1})} x_i^{\alpha_1} b^{-1}\bigg\} \\
&\le \exp\bigg(-\frac{\Delta_\tau^2}{8\lambda_1}\frac{L_1(x_i\zeta_1^{-1})}{L_1(\zeta_1^{-1})} x_i^{\alpha_1}\bigg) \\
&\le \exp\bigg(-\frac{\Delta_\tau^2}{8\lambda_1} x_i^{\alpha_1/2}\bigg).
\end{aligned}$$

Combining all the B_i such that $y^{-\delta_0} < x_i < \delta_1\zeta_1$, we have

$$\begin{aligned}
&\mathbb{P}\bigg\{\frac{1}{\mathrm{mes}(A_\gamma)} > y^{-d}\zeta_1^d, \sup_{y^{-\delta_0}\zeta_1^{-1}<|t-\tau|<\delta_1} f(t) > b \mid f(\tau) = \gamma + \frac{z}{b}\bigg\}\\
&\quad\le \sum_{k=0}^{\infty}\kappa(y^{-\delta_0}+k)^{d-1}\exp\bigg[-\frac{\Delta_\tau^2}{8\lambda_1}(y^{-\delta_0}+k)^{\alpha_1/2}\bigg]\\
&\quad\le \exp(-y^{-\varepsilon_0})
\end{aligned}$$

for some constant $\varepsilon_0 > 0$.

Case 3. When $|t_i - \tau| \ge \delta_1$ and $y < b^{-\lambda}$. Since $C(s,t)$ is uniformly Hölder continuous, we can always choose λ large such that for $|s-t| \le c_d y\zeta_1^{-1} \le c_d b^{-\lambda}\zeta_1^{-1}$,

$$|\mu_\tau(t) - \mu_\tau(s)| \le \frac{1}{4b}.$$

By Lemma 1(ii) and [20, Lemma 5(i)], for $|s-t| \le c_d y\zeta_1^{-1}$, the conditional variance $\mathrm{var}(f_0(s) - f_0(t))$ is bounded by

$$\mathrm{var}(f_0(s) - f_0(t)) \le \lambda_1\frac{L_1(c_d y\zeta_1^{-1})}{L_1(\zeta_1^{-1})}y^{\alpha_1}b^{-2} = O(y^{\alpha_1/2}b^{-2}).$$

Thus, there exists a constant $\varepsilon_0 > 0$ such that

$$\begin{aligned}
&\mathbb{P}\bigg\{\sup_{t\in B_i, |s-t|\le c_d y\zeta_1^{-1}} |f(t) - f(s)| > \frac{1}{b}, \sup_{t\in B_i} f(t) > b \mid f(\tau) = \gamma + \frac{z}{b}\bigg\}\\
&\quad\le \mathbb{P}\bigg\{\sup_{t\in B_i, |s-t|\le c_d y\zeta_1^{-1}} |f_0(t) - f_0(s)| > \frac{1}{2b}\bigg\}\\
&\quad\le 2\exp(-\varepsilon_0 y^{-\alpha_1}).
\end{aligned}$$

Note that $\zeta_1 \ll b^{4/\alpha_1}$, so for $y < b^{-\lambda}$, we have

$$\begin{aligned}
&\mathbb{P}\bigg\{\frac{1}{\mathrm{mes}(A_\gamma)} > y^{-d}\zeta_1^d, \sup_{|t-\tau|>\delta_1} f(t) > b \mid f(\tau) = \gamma + \frac{z}{b}\bigg\}\\
&\quad\le O(\zeta_1^d)\sup_i \mathbb{P}\bigg\{\sup_{t\in B_i, |s-t|\le c_d y\zeta_1^{-1}} |f(t) - f(s)| > \frac{1}{b}, \sup_{t\in B_i} f(t) > b \mid f(\tau) = \gamma + \frac{z}{b}\bigg\}\\
&\quad\le O(b^{4d/\alpha_1})\exp(-\varepsilon_0 y^{-\alpha_1/2})\\
&\quad\le O(y^{-4d/\alpha_1\lambda})\exp(-\varepsilon_0 y^{-\alpha_1/2})\\
&\quad\le \exp(-y^{-\varepsilon_0})
\end{aligned}$$

for some possibly different constant ε_0.

Case 4. Where $|t_i - \tau| \ge \delta_1$ and $y \ge b^{-\lambda}$. Note that condition (A3) implies that for any $\delta_1 > 0$, there exists $\varepsilon > 0$ such that for $|s-t| > \delta_1$, we have $r(s,t) < 1 - \varepsilon$. Thus, according to (21), there exists $\varepsilon > 0$ such that $\mu_\tau(t) \le (1-\varepsilon)b$. According to Proposition 1, we have for large enough b and some $\varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\bigg\{\frac{1}{\mathrm{mes}(A_\gamma)} > y^{-d}\zeta_1^d, \sup_{|t-\tau|\geq\delta_1} f(t) > b \mid f(\tau) = \gamma + \frac{z}{b}\bigg\} \\
&\quad \leq \mathbb{P}\Big\{\sup_{|t|\geq\delta_1} f_0(t) + \mu_\tau(t) > b\Big\} \\
&\quad \leq \mathbb{P}\Big\{\sup_{|t|\geq\delta_1} f_0(t) > \varepsilon b\Big\} \\
&\quad \leq \exp\bigg(-\frac{\varepsilon^2 b^2}{2\sigma_T^2}\bigg) \\
&\quad \leq \exp(-y^{-\varepsilon_0}).
\end{aligned}$$

Combining Cases 1–4, for some constants ε_0 and y_0 chosen to be small, we have for $y \in (0, y_0]$,

$$\mathbb{P}\bigg\{\frac{1}{\mathrm{mes}(A_\gamma)} > y^{-d}\zeta_1^d, M > b \mid f(\tau) = \gamma + \frac{z}{b}\bigg\} \leq \exp(-y^{-\varepsilon_0}). \tag{27}$$

Together with (19), we have

$$Q_b\bigg\{\frac{1}{\mathrm{mes}(A_\gamma)} > y^{-d}\zeta_1^d, M > b\bigg\} \leq \exp(-y^{-\varepsilon_0}). \tag{28}$$

Thus, according to (18), for some $\kappa > 0$, we have

$$\mathbb{E}^Q\bigg\{\frac{1}{\mathrm{mes}(A_\gamma)^2}; M > b\bigg\} \leq (\kappa + y_0^{-2d})\zeta_1^{2d}. \tag{29}$$

5.2. The I_1 term

To provide a lower bound of

$$I_1 = \mathbb{E}^{Q_b}\bigg\{\frac{1}{\mathrm{mes}(A_\gamma)}; M > b\bigg\},$$

we basically need to prove that $\mathrm{mes}(A_\gamma)$ cannot always be very large. Thus, it is sufficient to show that $f(t)$ drops below γ when t is reasonably far away from τ. In the next lemma we show that for any $\delta > 0$, the process $f(t)$ drops below γ almost all the time when $|t - \tau| > \delta$.

Lemma 2. *Under conditions (A1)–(A6), for a standard deviation of Type 1, we have*

$$Q_b\Big\{\sup_{|t-\tau|>\delta} f(t) \geq \gamma\Big\} \leq \mathrm{e}^{-\varepsilon_0 b^2} \quad \textit{for some } \varepsilon_0 > 0.$$

Lemma 3. *Under conditions (A1)–(A6), there exist δ small and κ large (independent of b), such that for $x > \kappa$, we have*

$$Q_b\Big\{\sup_{x\zeta^{-1}\leq|t-\tau|\leq\delta} f(t) \geq \gamma\Big\} < \mathrm{e}^{-\varepsilon_0 x^{\alpha_1/4}}.$$

https://doi.org/10.1239/aap/1444308882 Published online by Cambridge University Press

For the proof of Lemma 2 and Lemma 3; see [20]. We proceed to developing a lower bound for I_1. First, note that the event $\{M > b\}$ is a regular event under Q_b, i.e.

$$Q_b\{M > b\} \geq Q_b\{f(\tau) > b\} > \tfrac{1}{2}\mathrm{e}^{-1}.$$

The final step is based on an asymptotic calculation of the overshoot distribution of a standard Gaussian random variable. According to Lemmas 2 and 3, we choose x such that

$$Q_b\Big\{\sup_{|t-\tau|>x\zeta_1^{-1}} f(t) \geq \gamma\Big\} < \tfrac{1}{2}\mathrm{e}^{-2}.$$

Let ω_d be the volume of the d-dimensional unit ball. Thus, we have

$$\begin{aligned} I_1 &\geq \mathbb{E}^{Q_b}\Big\{\frac{1}{\mathrm{mes}(A_\gamma)}; M > b, \mathrm{mes}(A_\gamma) < \omega_d x^d \zeta_1^{-d}\Big\} \\ &\geq \omega_d^{-1} x^{-d} \zeta_1^d Q_b\{\mathrm{mes}(A_\gamma) < \omega_d x^d \zeta_1^{-d}, M > b\} \\ &\geq \omega_d^{-1} x^{-d} \zeta_1^d [Q_b\{M > b\} - Q_b\{\mathrm{mes}(A_\gamma) \geq \omega_d x^d \zeta_1^{-d}\}] \\ &\geq \omega_d^{-1} x^{-d} \zeta_1^d \Big[Q_b\{M > b\} - Q_b\Big\{\sup_{|t-\tau|>x\zeta_1^{-1}} f(t) \geq \gamma\Big\}\Big] \\ &\geq \tfrac{1}{2}\omega_d^{-1} x^{-d} \zeta_1^d (\mathrm{e}^{-1} - \mathrm{e}^{-2}). \end{aligned} \tag{30}$$

Summarizing (29) and (30), we have

$$\mathbb{E}_b\{Z_b^2\} \leq \kappa \zeta_1^{2d}\Big(\int \mathbb{P}\{f(t) > \gamma\}\,\mathrm{d}t\Big)^2, \qquad \mathbb{P}\{M > b\} > \varepsilon_0 \zeta_1^d \int \mathbb{P}\{f(t) > \gamma\}\,\mathrm{d}t,$$

and, therefore,

$$\sup_b \frac{\mathbb{E}^{Q_b} Z_b^2}{\mathbb{P}^2\{M > b\}} < \infty.$$

6. Proof of Theorem 2

Let $T_m = \{t_1, \ldots, t_m\}$ be generated in Step 3 of Algorithm 2. We start the analysis with the following decomposition:

$$\begin{aligned} \hat{Z}_b - Z_b &= \bigg[\frac{\mathbf{1}_{\{\sup f(t) > b\}}}{\mathrm{mes}(A_\gamma)} - \frac{\mathbf{1}_{\{\max_{i=1}^m f(t_i) > b\}}}{\widehat{\mathrm{mes}}(A_\gamma)}\bigg]\mathbb{E}\{\mathrm{mes}(A_\gamma)\} \\ &= \mathbb{E}\{\mathrm{mes}(A_\gamma)\}\bigg[\frac{\mathbf{1}_{\{\sup f(t) > b\}}}{\mathrm{mes}(A_\gamma)} - \frac{\mathbf{1}_{\{\max_{i=1}^m f(t_i) > b\}}}{\mathrm{mes}(A_\gamma)} \\ &\qquad + \frac{\mathbf{1}_{\{\max_{i=1}^m f(t_i) > b\}}}{\mathrm{mes}(A_\gamma)} - \frac{\mathbf{1}_{\{\max_{i=1}^m f(t_i) > b\}}}{\widehat{\mathrm{mes}}(A_\gamma)}\bigg], \end{aligned}$$

where $\widehat{\mathrm{mes}}(A_\gamma)$ is defined as in (14). According to the result in Theorem 1, it is sufficient to show that $|\mathbb{E}^{Q_b}\{\hat{Z}_b - Z_b\}| \leq \varepsilon \mathbb{P}\{M > b\}$ and $\mathrm{var}(\hat{Z}_b - Z_b) = O(\mathbb{P}^2\{M > b\})$. We define the following notation:

$$J_1 = \frac{\mathbf{1}_{\{\sup f(t) > b\}}}{\mathrm{mes}(A_\gamma)} - \frac{\mathbf{1}_{\{\max_{i=1}^m f(t_i) > b\}}}{\mathrm{mes}(A_\gamma)}, \qquad J_2 = \frac{\mathbf{1}_{\{\max_{i=1}^m f(t_i) > b\}}}{\mathrm{mes}(A_\gamma)} - \frac{\mathbf{1}_{\{\max_{i=1}^m f(t_i) > b\}}}{\widehat{\mathrm{mes}}(A_\gamma)}.$$

We establish upper bounds for the first and second moments for each of the two terms, respectively.

https://doi.org/10.1239/aap/1444308882 Published online by Cambridge University Press

6.1. The J_1 term

Note that J_1 is nonnegative and

$$\mathbb{E}_b\{J_1\} = \mathbb{E}_b\bigg\{\frac{1}{\mathrm{mes}(A_\gamma)}; M > b; \max_{i=1}^{m} f(t_i) \le b\bigg\}.$$

From the proof of Theorem 1, in particular (28), it follows that $\mathbf{1}_{\{M>b\}}/\zeta^d \mathrm{mes}(A_\gamma)$ is uniformly integrable in the parameter b, where $\zeta = \max(\zeta_1, \zeta_2)$. Thus, for any δ small enough, we have

$$\sup_{Q_b(B)\le\delta} \mathbb{E}_b\bigg\{\frac{1}{\mathrm{mes}(A_\gamma)}; M > b; B\bigg\} \le (-\log\delta)^{1/\varepsilon_0}\delta\zeta^d. \tag{31}$$

Therefore, it is sufficient to derive a bound for

$$Q_b\Big\{M > b; \max_{i=1}^{m} f(t_i) \le\Big\}.$$

Let x be large and δ' be small, we have the following:

$$\begin{aligned}
&Q_b\Big\{M > b; \max_{i=1}^{m} f(t_i) \le b\Big\} \\
&\quad\le Q_b\Big\{\sup_{x\zeta^{-1}<|t-\tau|<\delta'} f(t) > b; \max_{i=1}^{m} f(t_i) \le b\Big\} \\
&\qquad + Q_b\Big\{\sup_{|t-\tau|<x\zeta^{-1}} f(t) > b, \sup_{|t-\tau|>x\zeta^{-1}} f(t) \le b; \max_{i=1}^{m} f(t_i) \le b\Big\} \\
&\qquad + Q_b\Big\{\sup_{|t-\tau|\ge\delta'} f(t) > b; \max_{i=1}^{m} f(t_i) \le b\Big\}.
\end{aligned} \tag{32}$$

We will provide a specific choice of m such that

$$Q_b\Big\{\sup f(t) > b; \max_{i=1}^{m} f(t_i) \le b\Big\} \le \delta \triangleq \varepsilon^{1+\varepsilon_0},$$

where ε is the relative bias preset in the statement of the theorem. We consider each of the three terms in (32).

6.1.1. *The first term in (32).* We choose $x = \min\{(-\log\delta)^{4/\alpha}, \delta'\zeta\}$, where $\alpha = \min\{\alpha_1, \alpha_2\}$. According to Lemma 3, the first term in (32) is bounded by

$$Q_b\Big\{\sup_{x\zeta^{-1}<|t-\tau|<\delta'} f(t) > b; \max_{i=1}^{m} f(t_i) \le b\Big\} \le Q_b\Big\{\sup_{x\zeta^{-1}<|t-\tau|<\delta'} f(t) > b\Big\} \le \delta.$$

Notationally, we define $\sup_{t\in\varnothing} f(t) = -\infty$. Thus, when $x = \delta'\zeta$, the above probability is 0.

6.1.2. *The second term in (32).* Simple derivations yield

$$\begin{aligned}
&Q_b\Big\{\sup_{|t-\tau|<x\zeta^{-1}} f(t) > b, \sup_{|t-\tau|>x\zeta^{-1}} f(t) \le b, \max_{i=1}^{m} f(t_i) \le b\Big\} \\
&\quad= \mathbb{E}_b\Big\{Q_b\Big\{\max_{i=1}^{m} f(t_i) \le b \mid f\Big\}; \sup_{|t-\tau|<x\zeta^{-1}} f(t) > b, \sup_{|t-\tau|>x\zeta^{-1}} f(t) \le b\Big\} \\
&\quad\le E_b\Big\{(1-\beta(A_b))^m; \sup_{|t-\tau|<x\zeta^{-1}} f(t) > b\Big\},
\end{aligned} \tag{33}$$

https://doi.org/10.1239/aap/1444308882 Published online by Cambridge University Press

where $\beta(A_b) = \zeta^d \mathrm{mes}(A_b \cap B(\tau, x/\zeta)) \inf_{|t|\le x} k(t)$ is a lower bound of the probability that $Q_b\{t_i \in A_b \mid f\}$ and $B(\tau, x)$ is the ball centered around τ with radius x. In what follows, we need to show that $\mathrm{mes}(A_b)$ cannot be too small on the set $\{\sup_{|t-\tau|<x\zeta^{-1}} f(t) > b\}$ and, therefore, $\beta(A_b)$ cannot be too small. We write $\mathcal{E}_1 = \{\sup_{|t-\tau|<x\zeta^{-1}} f(t) > b\}$ and write (33) as

$$\mathbb{E}_b\{(1-\beta(A_b))^m; \mathcal{E}_1\} = \mathbb{E}_b\{(1-\beta(A_b))^m; \mathcal{E}_1, D^c_{\lambda_3,\delta_1}\} + \mathbb{E}_b\{(1-\beta(A_b))^m; \mathcal{E}_1, D_{\lambda_3,\delta_1}\},$$

where for some λ_3 and δ_1 positive, we define

$$D_{\lambda_3,\delta_1} = \Big\{ \sup_{|s-t|\le \lambda_3\zeta^{-1} s, t\in B(\tau, x\zeta^{-1})} |f(s) - f(t)| \le \delta_1 b^{-1} \Big\}.$$

For some ε_0 small, we choose $\delta_1 = \varepsilon_0 \delta$ and $\lambda_3 = \varepsilon_0 \delta_1^{2/\alpha + 1/\beta_1 + \varepsilon_0}$. We apply the Borel–TIS lemma to the double-indexed process $\xi(s, t) = f(s) - f(t)$ whose variance is bounded by Lemma 1(ii). Thus, we obtain the following bound:

$$\mathbb{E}_b\{(1-\beta(A_b))^m; \mathcal{E}_1, D^c_{\lambda_3,\delta_1}\} \le Q_b\{D^c_{\lambda_3,\delta_1}\} \le \delta.$$

Therefore, (33) is bounded by

$$\delta + \mathbb{E}_b(1-\beta(A_b))^m; \mathcal{E}_1, D_{\lambda_3,\delta_1}\}.$$

We further split the expectation

$$\begin{aligned}
&\mathbb{E}_b\{(1-\beta(A_b))^m; \mathcal{E}_1, D_{\lambda_3,\delta_1}\} \\
&\quad \le \mathbb{E}_b\Big\{(1-\beta(A_b))^m; D_{\lambda_3,\delta_1}; \sup_{|t-\tau|<x\zeta^{-1}} f(t) > b + \delta_1 b^{-1}, \mathcal{E}_1\Big\} \\
&\qquad + Q_b\Big\{D_{\lambda_3,\delta_1}; b < \sup_{|t-\tau|<x\zeta^{-1}} f(t) \le b + \delta_1 b^{-1}, \mathcal{E}_1\Big\}.
\end{aligned}$$

We derive a bound of the second term by considering the standardized process $g(t) = b(f(\tau + t/\zeta) - b)$ conditional on $f(\tau) = \gamma + z/b$. $g(t)$ can be written as

$$g(t) = \frac{C(t/\zeta + \tau, \tau)}{C(\tau, \tau)} z + l(t), \tag{34}$$

where $l(t)$ is a random field whose distribution is independent of z. So, we have

$$Q_b\Big\{b < \sup_{|t-\tau|<x\zeta^{-1}} f(t) < b + \delta_1 b^{-1}\Big\} = Q_b\Big\{\sup_{|t|\le x} \frac{C(t/\zeta + \tau)}{C(\tau, \tau)} z + l(t) \in (0, \delta_1)\Big\} = O(\delta_1).$$

The last equality holds because z has a density bounded everywhere (asymptotically exponential), and $\frac{1}{2} < C(t/\zeta + \tau)/C(\tau, \tau) < \sigma_T^2/\sigma^2(\tau)$. Given a realization of $l(t)$,

$$\sup_{|t|\le x} \frac{C(t/\zeta + \tau)}{C(\tau, \tau)} z + l(t) \in (0, \delta_1)$$

implies that z has to fall in an interval with length less than $2\delta_1$. Thus, if we choose ε_0 small and $\delta_1 = \varepsilon_0 \delta$, then

$$Q_b\Big\{b < \sup_{|t-\tau|<x\zeta^{-1}} f(t) < b + \delta_1 \zeta^{-1}\Big\} < \delta.$$

Therefore, (33) is bounded by

$$2\delta + \mathbb{E}^{Q_b}\Big\{(1-\beta(A_b))^m; D_{\lambda_3,\delta_1}; \sup_{|t-\tau|<x\zeta^{-1}} f(t) > b + \delta_1 b^{-1}, \mathcal{E}_1\Big\}.$$

Note that, on the set D_{λ_3,δ_1}, $\mathrm{mes}(A_b \cap B(\tau, x\zeta^{-1}))$ is controlled by the overshoot

$$\sup_{|t-\tau|<x\zeta^{-1}} f(t) - b,$$

i.e. if $\sup_{|t-\tau|<x\zeta^{-1}} f(t) > b + \delta_1/b$ then $\mathrm{mes}(A_b \cap B(\tau, x\zeta^{-1})) \ge \varepsilon_0 \lambda_3^d \zeta^{-d}$. In addition, the density $k_{\tau,\zeta}(t)$ is bounded from below by $x^{-d-\varepsilon_1}$ for $t \in B(\tau, x\zeta^{-1})$. Thus, the probability $\beta(A_b)$ has a lower bound

$$\beta(A_b) \ge \varepsilon_0 x^{-d-\varepsilon_1}\lambda_3^d \ge \varepsilon_0 \delta^{2d/\alpha + d/\beta_1 + 2\varepsilon_0}.$$

The final step of the above inequality follows from the fact that $x = \min\{(-\log\delta)^{4/\alpha}, \delta'\zeta\}$. Thus, (33) is bounded by

$$2\delta + (1 - \varepsilon_0 \delta^{2d/\alpha + d/\beta_1 + 2\varepsilon_0})^m.$$

For some large κ, $m = \kappa\delta^{-2d/\alpha - d/\beta_1 - 3\varepsilon_0}$ and, therefore,

$$Q_b\Big\{\sup_{|t-\tau|<x\zeta^{-1}} f(t) > b, \sup_{|t-\tau|>x\zeta^{-1}} f(t) \le b; \max_{i=1}^m f(t_i) \le b\Big\} \le 4\delta.$$

6.1.3. *The last term in (32).* According to the result in Lemma 2, we can choose ε_0 and δ' such that

$$Q_b\Big\{\sup_{|t-\tau|\ge\delta'} f(t) \ge \gamma\Big\} \le \mathrm{e}^{-\varepsilon_0 b^2}.$$

There are two cases: $\delta > \mathrm{e}^{-\varepsilon_0 b^2}$ and $\delta \le \mathrm{e}^{-\varepsilon_0 b^2}$.

Case 1. Where $\delta > \mathrm{e}^{-\varepsilon_0 b^2}$. In this case, the last term in (32) is bounded trivially by

$$Q_b\Big\{\sup_{|t-\tau|\ge\delta'} f(t) > b; \max_{i=1}^m f(t_i) \le b\Big\} \le Q_b\Big\{\sup_{|t-\tau|\ge\delta'} f(t) \ge \gamma\Big\} \le \delta.$$

Case 2. Where $\delta < \mathrm{e}^{-\varepsilon_0 b^2}$. We need a similar analysis to that of the second term. We now split the probability for $\delta_2 = \delta^{1+\varepsilon_0}$,

$$\begin{aligned} Q_b\Big\{\sup_{|t-\tau|\ge\delta'} f(t) > b; \max_{i=1}^m f(t_i) \le b\Big\} &\le Q_b\Big\{\sup_{|t-\tau|\ge\delta'} f(t) \in [b, \delta_2 b^{-\lambda}]\Big\} \\ &\quad + Q_b\Big\{\sup_{|t-\tau|\ge\delta'} f(t) > b + \delta_2 b^{-\lambda}; \max_{i=1}^m f(t_i) \le b\Big\}. \end{aligned}$$

We now consider the first term and split the set $\{t : |t-\tau| > \delta'\}$ into two parts. Define the set

$$F = \Big\{t : \frac{C(t,\tau)}{C(\tau,\tau)} > \frac{1}{(-\log\delta_2)^2}\Big\},$$

We start with the small overshoot probability on the set F,

$$Q_b\Big\{b < \sup_{|t-\tau|>\delta', t\in F} f(t) \le b + \frac{\delta_2}{b}\Big\}.$$

Using (34) and applying a similar analysis to that of the second term, we have

$$\begin{aligned} Q_b\Big\{b < \sup_{|t-\tau|\ge\delta', t\in F} f(t) < b+\delta_2 b^{-1}\Big\} &\le Q_b\Big\{\sup_{|t/\zeta|>\delta', t/\zeta+\tau\in F} \frac{C(t/\zeta+\tau)}{C(\tau,\tau)} z + l(t) \in (0,\delta_2)\Big\} \\ &= O((-\log\delta_2)^2\delta_2) \\ &\le \delta. \end{aligned} \tag{35}$$

The last two steps are based on the fact that z is a random variable independent of $l(t)$ and has bounded density. Thus, the above probability is bounded by

$$\sup_x \mathbb{P}\{x < z < x + (\log\delta_2)^2\delta_2\} = O((\log\delta_2)^2\delta_2).$$

We will return to this estimate soon.

We now consider t in F^c. For some κ_0 large, we have $Q_b\{z > -\kappa_0\log\delta_2\} < \delta_2$. Thus, we only consider $z < -\kappa_0\log\delta_2$. Conditional on $f(\tau) = \gamma + z/b$, the conditional mean is $\sup_{t\in F^c}\mu_\tau(t-\tau) \le C > 0$. In addition, the conditional variance of $f(t)$ on the set F^c is almost $\sigma^2(t)$. Thus, we can apply classic results on the density estimation of the sup $f(t)$ (cf. [34, Theorem 2]). That is, conditional on $f(\tau) = \gamma + z/b$, $\sup_{|t-\tau|\ge\delta', F^c} f(t)$ has a bounded density over $[b, b+\delta_2 b^{-\lambda}]$ for some $\lambda \ge 1$ and, thus,

$$Q_b\Big\{\sup_{|t-\tau|\ge\delta', t\in F^c} f(t) \in [b, b+\delta_2 b^{-\lambda}] \mid f(\tau) = \gamma + \frac{z}{b}\Big\} = O(\delta_2).$$

Summarizing the above results, we have

$$\begin{aligned} &Q_b\Big\{\sup_{|t-\tau|\ge\delta'} f(t) \in [b, b+\delta_2 b^{-\lambda}]\Big\} \\ &\quad\le Q_b\Big\{\sup_{|t-\tau|\ge\delta', t\in F} f(t) \in [b, b+\delta_2 b^{-\lambda}]\Big\} + Q_b\{z \ge -\kappa_0\log\delta_2\} \\ &\qquad + Q_b\Big\{\sup_{|t-\tau|\ge\delta', t\in F^c} f(t) \in [b, b+\delta_2 b^{-\lambda}], z \le -\kappa_0\log\delta_2\Big\} \\ &\quad\le 3\delta. \end{aligned}$$

The last term in (32) is bounded by

$$\begin{aligned} &Q_b\Big\{\sup_{|t-\tau|\ge\delta'} f(t) > b; \max_{i=1}^m f(t_i) \le b\Big\} \\ &\qquad\le 3\delta + Q_b\Big\{\sup_{|t-\tau|\ge\delta'} f(t) > b+\delta_2 b^{-\lambda}; \max_{i=1}^m f(t_i) \le b\Big\}. \end{aligned}$$

For the second term, we apply the old trick of choosing $\lambda_4 = \delta_2^{2/\alpha+1/\beta_1+\varepsilon_0} b^{-2\lambda/\alpha-\lambda/\beta_1}$, and, thus,

$$Q_b\Big\{\sup_{|s-t|<\lambda_4} |f(s)-f(t)| > \delta_2 b^{-\lambda}\Big\} < \delta_2. \tag{36}$$

Note that $b^2 \le -\varepsilon_0^{-1}\log\delta_2$. We can choose a different ε_0 so that λ_4 can be simplified to

$$\lambda_4 = \delta_2^{2/\alpha+1/\beta_1+\varepsilon_0}.$$

If $\sup_{|s-t|<\lambda_4}|f(s)-f(t)| < \delta_2 b^{-\lambda}$ and $\sup_{|t-\tau|\geq\delta'} f(t) > b+\delta_2 b^{-\lambda}$, we have $\beta(A_b) \geq \varepsilon_0\lambda_4^d\zeta^{-d-\varepsilon_1}$. With a different choice of ε_0, we choose

$$m = -2\lambda_4^{-d}\zeta^{d+\varepsilon_1}\log\delta = O(\delta^{-d(2/\alpha+1/\beta_1)-\varepsilon_0}), \tag{37}$$

then, we have

$$\mathbb{E}_b\Big\{(1-\beta(A_b))^m;\ \sup_{|s-t|<\lambda_4} b < |f(s)-f(t)| < \delta_2 b^{-\lambda},\ f(t) > b+\delta_2 b^{-\lambda}\Big\} \leq \delta. \tag{38}$$

Therefore, combining the bounds in (35), (36), and (38), if $\varepsilon < \mathrm{e}^{-\varepsilon_0 b^2}$ and we choose m as in (37), then

$$Q_b\Big\{\sup_{|t-\tau|>\delta'} f(t) > b;\ \max_{i=1}^{m} f(t_i) \leq b\Big\} \leq 5\delta.$$

Combining the bounds for all the three terms in (32), we have

$$Q_b\Big\{M > b;\ \max_{i=1}^{m} f(t_i) \leq b\Big\} \leq 5\delta.$$

If we choose $\delta = \varepsilon^{1+\varepsilon_0}$ and

$$m = O(\delta^{-d(2/\alpha+1/\beta_1+\varepsilon_0)}) = O(\varepsilon^{-d(2/\alpha+1/\beta_1)-2d\varepsilon_0})$$

then according to the bound in (31), we have

$$\mathbb{E}^{Q_b} J_1 \leq \zeta^d\varepsilon.$$

Similarly, according to the uniform integrability of $\zeta^{-2d}/\mathrm{mes}^2(A_\gamma)$, by choosing the same m, there exists a κ_0 such that

$$\mathbb{E}^{Q_b}\{J_1^2\} \leq \kappa_0\zeta^{2d}.$$

6.2. The J_2 term

We now proceed to

$$J_2 = \mathbf{1}_{\{\max_{i=1}^m f(t_i)>b\}}\bigg[\frac{1}{\mathrm{mes}(A_\gamma)} - \frac{1}{\widehat{\mathrm{mes}}(A_\gamma)}\bigg].$$

We study the behavior of J_2 by means of the scaled process $g(t)$ defined as in (11). For the analysis of J_2, we translate everything to the scale of $g(t)$. Recall the process $g(t)$ given by (11) is

$$g(t) = b\bigg(f\bigg(\tau+\frac{t}{\zeta}\bigg)-b\bigg)$$

for each t, $f(\tau+t/\zeta) > \gamma$ if and only if $g(t) > -1$.

Conditional on $\tau, t_1, \ldots, t_m$ are i.i.d. with density $k_{\tau,\zeta}(t)$ defined as in (13). Let $s_i = (t_i-\tau)\zeta$ and, thus, $s_1, \ldots, s_m$ are i.i.d. following density $k(s)$. We can then rewrite the estimator in (14) as

$$\widehat{\mathrm{mes}}(A_\gamma) = \frac{\zeta^{-d}}{m}\sum_{i=1}^{m}\frac{\mathbf{1}_{\{g(s_i)>-1\}}}{k(s_i)}.$$

https://doi.org/10.1239/aap/1444308882 Published online by Cambridge University Press

Thus, $\widehat{\mathrm{mes}}(A_\gamma)$ is an unbiased estimator of $\mathrm{mes}(A_\gamma)$, i.e. $\mathbb{E}(\widehat{\mathrm{mes}}(A_\gamma) \mid f) = \mathrm{mes}(A_\gamma)$. Conditional on a particular realization of $f(t)$ (or equivalently, g(t)), the variance of $\widehat{\mathrm{mes}}(A_\gamma)$ is

$$\mathrm{var}(\widehat{\mathrm{mes}}(A_\gamma) \mid f) = \frac{\kappa_f \zeta^{-2d}}{m},$$

where

$$\kappa_f = \mathrm{var}\left[\frac{\mathbf{1}_{\{g(S)>-1\}}}{k(S)} \,\middle|\, f\right] \le k^{-2}(t_f)$$

and$t_f = \max(|t|\colon g(t) > -1)$. By means of the inequality $1/(1+x) - 1 \ge -x$, we have

$$\frac{1}{\mathrm{mes}(A_\gamma)} - \frac{1}{\widehat{\mathrm{mes}}(A_\gamma)} \le \frac{\widehat{\mathrm{mes}}(A_\gamma) - \mathrm{mes}(A_\gamma)}{\mathrm{mes}^2(A_\gamma)}.$$

Therefore,

$$\mathbb{E}\left\{\left(\frac{1}{\mathrm{mes}(A_\gamma)} - \frac{1}{\widehat{\mathrm{mes}}(A_\gamma)}\right)^2; \widehat{\mathrm{mes}}(A_\gamma) > \mathrm{mes}(A_\gamma) \,\middle|\, f\right\} \le \frac{\kappa_f \zeta^{-2d}}{m\, \mathrm{mes}^4(A_\gamma)}.$$

It is the expectation on the set $\{\widehat{\mathrm{mes}}(A_\gamma) < \mathrm{mes}(A_\gamma)\}$ that induces complications in that the factor $1/\widehat{\mathrm{mes}}(A_\gamma)$ can be very large when there are not many t_i in the excursion set A_γ. We now proceed to this case. Conditional on a particular realization of f (and, equivalently, the process $g(t)$), the analysis consists of three steps.

Step 1. Define the f-dependent probability

$$p_f \triangleq Q_b\{t_i \in A_\gamma\colon f\} = \int_{A_\gamma} k_{\tau,\zeta}(t)\,\mathrm{d}t = \int_{A^g_{-1}} k(t)\,\mathrm{d}t.$$

Using standard exponential change of measure techniques for large deviations [15], we obtain

$$Q_b\left\{\sum_{i=1}^m \mathbf{1}_{\{t_i\in A_\gamma\}} \le p_f(1-\delta_3)m \,\middle|\, f\right\} \le \mathrm{e}^{-mI_{\delta_3,p_f}} \quad \text{for all } \delta_3 \in (0,1),$$

where the rate function $I_{\delta_3,p_f} = \theta_* p_f(1-\delta_3) - \varphi(\theta_*)$, $\varphi(\theta) = \log(1 - p_f + p_f \mathrm{e}^{\theta})$, and $\theta_* = \log(1 - \delta_3/(1 - p_f(1-\delta_3)))$. By elementary calculus, if we choose $\delta_3 = \frac{1}{2}$ then we have, for some $\varepsilon_0 > 0$,

$$I_{\delta_3,p_f} \ge \varepsilon_0 p_f \quad \text{for all } p_f > 0.$$

Furthermore, we have

$$\begin{aligned}
&\mathbb{E}\left\{\left(\frac{1}{\mathrm{mes}(A_\gamma)} - \frac{1}{\widehat{\mathrm{mes}}(A_\gamma)}\right)^2; \widehat{\mathrm{mes}}(A_\gamma) \le \mathrm{mes}(A_\gamma), \max_{i=1}^m f(t_i) > b,\right.\\
&\quad \left.\sum_{i=1}^m \mathbf{1}_{\{t_i\in A_\gamma\}} \le \frac{p_f m}{2} \,\middle|\, f\right\}\\
&\qquad \le \mathbb{E}\left\{\frac{4}{\widehat{\mathrm{mes}}^2(A_\gamma)}; \widehat{\mathrm{mes}}(A_\gamma) \le \mathrm{mes}(A_\gamma), \max_{i=1}^m f(t_i) > b, \sum_{i=1}^m \mathbf{1}_{\{t_i\in A_\gamma\}} \le \frac{p_f m}{2} \,\middle|\, f\right\}.
\end{aligned}$$

https://doi.org/10.1239/aap/1444308882 Published online by Cambridge University Press

There is at least one t_i in the excursion set A_γ. Therefore, the estimator $\widehat{\mathrm{mes}}(A_\gamma) \geq m^{-1}\zeta^{-d}k^{-1}(t_f)$. Thus, the above expectation is upper bounded by

$$\leq \kappa k^{-2}(t_f)m^2\zeta^{2d}\mathrm{e}^{-\varepsilon_0 m p_f}.$$

Step 2. We consider the situation when $\sum \mathbf{1}_{\{t_i\in A_\gamma\}} > p_f m/2$. The unbiasedness of $\widehat{\mathrm{mes}}(A_\gamma)$ suggests that

$$\mathrm{mes}(A_\gamma) = \mathbb{E}\bigg\{\frac{1}{\zeta^d k(S)} \;\bigg|\; S \in A_{-1}^g\bigg\} p_f,$$

where S is a random index following density $k(s)$. Note that on the set A_{-1}^g, $k(t_f) \leq k(S) \leq \kappa_1$. Thus, if we let $\lambda_f = \kappa_1^{-1}k(t_f)$ then on the set $\{\sum \mathbf{1}_{\{t_i\in A_\gamma\}} > p_f m/2\}$, we have

$$\widehat{\mathrm{mes}}(A_\gamma) \geq \frac{\lambda_f \mathrm{mes}(A_\gamma)}{2}.$$

Thus, using a Taylor expansion, we have

$$\begin{aligned}
&\mathbb{E}_b\bigg\{\bigg(\frac{1}{\mathrm{mes}(A_\gamma)} - \frac{1}{\widehat{\mathrm{mes}}(A_\gamma)}\bigg)^2; \widehat{\mathrm{mes}}(A_\gamma) < \mathrm{mes}(A_\gamma); \sum \mathbf{1}_{\{t_i\in A_\gamma\}} > \frac{p_f m}{2} \;\bigg|\; f\bigg\}\\
&\quad\leq \mathbb{E}_b\bigg\{\frac{2^4(\mathrm{mes}(A_\gamma) - \widehat{\mathrm{mes}}(A_\gamma))^2}{\lambda_f^4 \mathrm{mes}^4(A_\gamma)}; \widehat{\mathrm{mes}}(A_\gamma) < \mathrm{mes}(A_\gamma); \sum \mathbf{1}_{\{t_i\in A_\gamma\}} > \frac{p_f m}{2} \;\bigg|\; f\bigg\}\\
&\quad\leq \frac{2^4\kappa_f\zeta^{-2d}}{m\lambda_f^4 \mathrm{mes}^4(A_\gamma)}.
\end{aligned}$$

Step 3. Combining the previous analysis, we obtain

$$\mathbb{E}_b\{J_2^2 \mid f\} \leq \frac{2^4\zeta^{-2d}}{\mathrm{mes}^4(A_\gamma)}\frac{\kappa_1^2}{k^2(t_f)m} + \frac{\kappa_f\zeta^{-2d}}{m\mathrm{mes}^4(A_\gamma)} + k(t_f)^{-2}m^2\zeta^{2d}\mathrm{e}^{-\varepsilon_0 m p_f}. \tag{39}$$

The density $k(t)$ has a heavy tail that is $k(t) \sim 1/|t|^{d+\varepsilon_1}$ and $k(t) \leq \kappa_1$ for all t. In Step 3, we provide a bound on the distributions of t_f and p_f.

We start with t_f. For each $s > 0$, $t_f > s$ if and only if $\sup_{|t-\tau|>s} g(t) > -1$. According to the results in Lemmas 2 and 3, for sufficiently large s, there exists some $\varepsilon_0 > 0$ such that

$$Q_b\{t_f > s\} = Q_b\Big\{\sup_{|t-\tau|>s} g(t) > -1\Big\} \leq \exp\{-s^{\varepsilon_0}\} \quad \text{for } s < \delta'\zeta$$

and

$$Q_b\{t_f > s\} \leq \exp(-\varepsilon_0 b^2) \quad \text{for } s > \delta'\zeta.$$

Therefore, all moments of $k^{-1}(t_f)$ are bounded. We have

$$\mathbb{E}_b\{k^{-l}(t_f)\} \leq \mathbb{E}_b\{t_f^{(d+\varepsilon_1)l}\} \leq \kappa_l$$

for some constant κ_l possibly depending on l. Thus, by the Cauchy–Schwarz inequality, the expectation of the first two terms in (39) can be bounded as follows:

$$\begin{aligned}
\mathbb{E}\bigg\{\frac{2^4\zeta^{-2d}}{\mathrm{mes}^4(A_\gamma)}\frac{\kappa_1^2}{k^2(t_f)m}; M > b\bigg\} &\leq \frac{O(1)}{m}\sqrt{\mathbb{E}\bigg\{\frac{\zeta^{-4d}}{\mathrm{mes}^8(A_\gamma)}\bigg\}\mathbb{E}\{k^{-4}(t_f)\}} \leq \frac{\kappa\zeta^{2d}}{m},\\
\mathbb{E}\bigg\{\frac{\kappa_f\zeta^{-2d}}{m \times \mathrm{mes}^4(A_\gamma)}\bigg\} &\leq \frac{O(1)}{m}\sqrt{\mathbb{E}\bigg\{\frac{\zeta^{-4d}}{\mathrm{mes}^8(A_\gamma)}\bigg\}\mathbb{E}\{k^{-4}(t_f)\}} \leq \frac{\kappa\zeta^{2d}}{m}.
\end{aligned}$$

https://doi.org/10.1239/aap/1444308882 Published online by Cambridge University Press

We now proceed to the third term in (39) concerning p_f. The expectation of this term is bounded by

$$\mathbb{E}_b\{m^2 k(t_f)^{-2}\mathrm{e}^{-m\varepsilon_0 p_f}; M > b\} \le \sqrt{\mathbb{E}_b\{m^4\mathrm{e}^{-2m\varepsilon_0 p_f}; M > b\}}\sqrt{\mathbb{E}_b\{k^{-4}(t_f)\}}.$$

The second term $\sqrt{\mathbb{E}_b\{k^{-4}(t_f)\}}$ is $O(1)$. We proceed to the first term,

$$\begin{aligned}\mathbb{E}_b\{m^4\mathrm{e}^{-2m\varepsilon_0 p_f}; M > b\} &= \mathbb{E}_b\{m^4\mathrm{e}^{-2m\varepsilon_0 p_f}; p_f \ge m^{-1/2}\}\\ &\quad + \mathbb{E}_b\{m^4\mathrm{e}^{-2m\varepsilon_0 p_f}; p_f \le m^{-1/2}, M > b\}\\ &\le m^4\mathrm{e}^{-2\varepsilon_0\sqrt{m}} + m^4 Q_b\{p_f \le m^{-1/2}, M > b\}.\end{aligned}$$

We now proceed to control $Q_b\{p_f \le m^{-1/2}, M > b\}$. Note that $p_f \ge k(t_f)\mathrm{mes}(A^g_{-1})$. For each $x > 0$,

$$\begin{aligned}Q_b\{p_f < x, M > b\} &\le Q_b\{k(t_f) < \sqrt{x} \quad\text{or}\quad \mathrm{mes}(A^g_{-1}) < \sqrt{x}, M > b\}\\ &\le Q_b\{t_f > x^{-1/2(d+\varepsilon_1)}\} + Q_b\{\mathrm{mes}(A^g_{-1}) < \sqrt{x}, M > b\}.\end{aligned}$$

According to the bounds in (27), for some $\delta_0 > 0$ and $\varepsilon_0 > 0$, we have

$$Q_b\{\mathrm{mes}(A^g_{-1}) < \sqrt{x}, M > b\} = Q_b\{\mathrm{mes}(A_\gamma) < \zeta^{-d}\sqrt{x}, M > b\} \le \exp(-x^{-\varepsilon_0/d})$$

for sufficiently small x. According to the previous result, we have

$$Q_b\{t_f > x^{-1/2(d+\varepsilon_1)}\} \le \mathrm{e}^{-x^{-\varepsilon_0}} \quad \text{for } x^{-1/2(d+\varepsilon_1)} < \delta'\zeta$$

and

$$Q_b\{t_f > x^{-1/2(d+\varepsilon_1)}\} \le \mathrm{e}^{-\varepsilon_0 b^2} \quad \text{for } x^{-1/2(d+\varepsilon_1)} \ge \delta'\zeta.$$

Thus, for large enough λ and small enough ε_0, we have

$$Q_b\{p_f \le m^{-1/2}, M > b\} \le \mathrm{e}^{-m^{\varepsilon_0}} \quad \text{for } m < b^\lambda;$$

for $m > b^\lambda$ (with sufficiently large λ), $t_f > m^{1/4(d+\varepsilon_1)}$ implies that $\tau + t_f/\zeta \notin T$, i.e. $m^{1/4(d+\varepsilon_1)}$ is too large and, thus,

$$Q_b\{p_f < m^{-1/2}\} = 0 \quad \text{for } m > b^\lambda.$$

Therefore, we have $m^4 Q_b\{p_f \le m^{-1/2}, M > b\} \le \kappa m^4 \mathrm{e}^{-m^{\varepsilon_0}}$ for sufficiently large m and, furthermore,

$$\mathbb{E}_b\{m^4 k(t_f)^{-2}\mathrm{e}^{-m\varepsilon_0 p_f}; M > b\} \le \kappa m^4\mathrm{e}^{-m^{\varepsilon_0}/2}.$$

Summarizing the results in all the three steps, we have $\mathbb{E}_b\{J_2^2\} \le \kappa\zeta^{-2d}/m$. If we choose $m = \kappa \max\{\varepsilon^{-2}, \varepsilon^{-d(2/\alpha+1/\beta_1+3\varepsilon_0)}\} = O(\varepsilon^{-d(2/\alpha+2/\beta_1)})$ then

$$\mathbb{E}_b|\hat{Z}_b - Z_b| = \mathbb{E}_b|J_1 + J_2| \int_T \mathbb{P}\{f(t) > \gamma\}\,\mathrm{d}t \le \varepsilon\zeta^d \int_T \mathbb{P}\{f(t) > \gamma\}\,\mathrm{d}t$$

and

$$\mathbb{E}_b\{\hat{Z}_b - Z_b\}^2 \le \kappa\zeta^{2d}\left(\int_T \mathbb{P}\{f(t) > \gamma\}\,\mathrm{d}t\right)^2.$$

https://doi.org/10.1239/aap/1444308882 Published online by Cambridge University Press

Acknowledgements

This research is supported in part by the National Science Foundation (grant no. NSF SES-1323977) and the Army Research Laboratory (grant no. W911NF-14-1-0020 and W911NF-12-R-0012)

References

[1] Adler, R. J. (1981). *The Geometry of Random Fields*. Wiley, Chichester.

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

[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., Samorodnitsky, G. and Taylor, J. E. (2013). High level excursion set geometry for non-Gaussian infinitely divisible random fields. *Ann. Prob.* **41,** 134–169.

[5] Aldous, D. J. (1989). *Probability Approximations via the Poisson Clumping Heuristic*. Springer, New York.

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

[9] Berman, S. M. (1972). Maximum and high level excursion of a Gaussian process with stationary increments. *Ann. Math. Statist.* **43,** 1247–1266.

[10] 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.

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

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

[13] Bucklew, J. A. (2004). *Introduction to Rare Event Simulation*. Springer, New York.

[14] Cirel'son, 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* (Tashkent, 1975), Springer, Berlin, pp. 20–41.

[15] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*, 2nd edn. Springer, New York.

[16] Dudley, R. M. (2010). Sample functions of the Gaussian process. In *Selected Works of R. M. Dudley*, Springer, New York, pp. 187–224.

[17] Juneja, S. and Shahabuddin, P. (2006). Rare event simulation techniques. In *Handbooks in Operations Research and Management Science: Simulation*, North-Holland, Amsterdam, pp. 291–350.

[18] Landau, H. J. and Shepp, L. A. (1970). Supremum of a Gaussian process. *Sankhyā Ser. 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. (2013). Rare-event simulation and efficient discretization for the Supremum of Gaussian random fields. Available at http://arxiv.org/abs/1309.7365.

[21] Liu, J. (2012). Tail approximations of integrals of Gaussian random fields. *Ann. Prob.* **40,** 1069–1104.

[22] 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.

[23] Liu, J. and Xu, G. (2013). On the density functions of integrals of Gaussian random fields. *Adv. Appl. Prob.* **45,** 398–424.

[24] 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.

[25] Marcus, M. B. and Shepp, L. A. (1970). Continuity of Gaussian processes. *Trans. Amer. Math. Soc.* **151,** 377–391.

[26] Mitzenmacher, M. and Upfal, E. (2005). *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press.

[27] Piterbarg, V. I. (1996). *Asymptotic Methods in the Theory of Gaussian Processes and Fields*. American Mathematical Society, Providence, RI.

[28] Sudakov, V. N. and Cirel'son, B. S. (1974). Extremal properties of half spaces for spherically invariant measures. *Zap. Naučn. Sem. LOMI* **41,** 14–24.

[29] Sun, J. (1993). Tail probabilities of the maxima of Gaussian random fields. *Ann. Prob.* **21,** 34–71.

[30] Talagrand, M. (1996). Majorizing measures: the generic chaining. *Ann. Prob.* **24,** 1049–1103.

[31] Taylor, J., Takemura, A. and Adler, R. J. (2005). Validity of the expected Euler characteristic heuristic. *Ann. Prob.* **33,** 1362–1396.

[32] TAYLOR, J. E. AND ADLER, R. J. (2003). Euler characteristics for Gaussian fields on manifolds. *Ann. Prob.* **31,** 533–563.
[33] TRAUB, J. F., WASILOKOWSKI, G. W. AND WOŹNIAKOWSKI, H. (1988). *Information-Based Complexity*. Academic Press, Boston, MA.
[34] TSIREL'SON, V. S. (1976). The density of the distribution of the maximum of a Gaussian process. *Theory Prob. Appl.* **20,** 847–856.
[35] WOŹNIAKOWSKI, H. (1997). Computational complexity of continuous problems. In *Nonlinear Dynamics, Chaotic and Complex Systems*, Cambridge University Press, pp. 283–295.