

Efficient Simulations for the Exponential Integrals of Hölder Continuous Gaussian Random Fields

Jingchen Liu¹ and Gongjun Xu, Columbia University

In this paper, we consider a Gaussian random field $f(t)$ living on a compact set $T \subset R^d$ and the computation of the tail probabilities $P(\int_T e^{f(t)} dt > e^b)$ as $b \rightarrow \infty$. We design asymptotically efficient importance sampling estimators for a general class of Hölder continuous Gaussian random fields. In addition to the variance control, we also analyze the bias (relative to the interesting tail probabilities) caused by the discretization.

1. INTRODUCTION

In this paper, we focus on the design and the analysis of efficient Monte Carlo methods for computing the tail probabilities of integrals of exponential functions of Gaussian random fields living on a compact domain. Suppose that $\{f(t) : t \in T\}$ is a continuous Gaussian random field with zero mean and unit variance. That is, for each finite subset $\{t_1, \dots, t_n\} \subset T$, $(f(t_1), \dots, f(t_n))$ is a multivariate Gaussian random vector with $E(f(t_i)) = 0$ and $\text{Var}(f(t_i)) = 1$, $i = 1, \dots, n$. The domain T is a d -dimensional compact subset of R^d . Further conditions will be imposed on f in the later discussions. Define

$$\mathcal{I}(T) = \int_T e^{\sigma(t)f(t)+\mu(t)} dt, \quad (1)$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ are two deterministic functions and $\sigma(\cdot)$ is strictly positive. Our focus is on the tail probabilities

$$w(b) = P(\mathcal{I}(T) > e^b) = P\left(\int_T e^{\sigma(t)f(t)+\mu(t)} dt > e^b\right) \quad \text{as } b \rightarrow \infty. \quad (2)$$

1.1. Motivations

The exponential integral of a Gaussian random field is a central object of several probability models. The integral (1) is the limiting object of the (weighted) sum of correlated lognormal random variables. The tail event of the latter sum is an important topic. An incomplete list of works is given by [Ahsan 1978; Duffie and Pan 1997; Glasserman et al. 2000; Basak and Shapiro 2001; Deutsch 2004; Foss and Richards 2010]. In the portfolio risk analysis, consider a portfolio consisting of n assets (S_1, \dots, S_n) each of which is associated with a weight (e.g.

¹This research is supported in part by NSF CMMI-1069064.

number of shares) (w_1, \dots, w_n) . One popular model assumes that $(\log S_1, \dots, \log S_n)$ is a multivariate Gaussian random vector. The value of the portfolio, $S = \sum_{i=1}^n w_i S_i$, is then the sum of correlated log-normal random variables. Without loss of generality, we let $\sum w_i = n$.

One interesting situation is that the portfolio size is large and the asset prices are usually highly correlated. One may employ a latent space approach used in the literature of social network. More specifically, we construct a Gaussian process $\{f(t) : t \in T\}$ and associate each asset i with a latent variable $t_i \in T$ so that $\log S_i = f(t_i)$. Then, the log asset prices fall into a subset of the continuous Gaussian process. Furthermore, there exists a (deterministic) process $w(t)$ so that $w(t_i) = w_i$. Then, the unit share value of the portfolio is $\frac{1}{n} \sum w_i S_i = \frac{1}{n} \sum w(t_i) e^{f(t_i)}$. For detailed discussion of latent state modeling, see [Hoff et al. 2002; Snijders 2002; Handcock et al. 2007; Xing et al. 2010].

In the asymptotic regime that $n \rightarrow \infty$ and the correlations among the asset prices become close to one, the subset $\{t_i : i = 1, \dots, n\}$ becomes dense in T . Ultimately, we obtain the limit

$$\frac{1}{n} \sum_{i=1}^n w_i S_i \rightarrow \int w(t) e^{f(t)} h(t) dt$$

where $h(t)$ indicates the limiting spatial distribution of $\{t_i : i = 1, \dots, n\}$ in T . Let $\mu(t) = \log w(t) + \log h(t)$. Then the tail probability of the (limiting) unit share price is $P(\int e^{f(t) + \mu(t)} dt > e^b)$.

Another application of the current study lies in option pricing. A stylized model for asset price $S(t)$ (as a function of time t) is geometric Brownian motion, that is, $S(t) = e^{W(t)}$, where $W(t)$ is a Brownian motion; see [Black and Scholes 1973; Merton 1973]. Then the payoff of an Asian option is at expiration time T is a function of the averaged price $\int_0^T e^{W(t)} dt$. For instance, the payoff of an Asian call option with strike price K is $\max(\int_0^T e^{W(t)} dt - K, 0)$; the expected payoff of a digital Asian call option is precisely $P(\int_0^T e^{W(t)} dt > K)$.

1.2. Literature and contribution

There is a rich rare-event simulation literature for the sum of independent random variables. The problems are mostly classified to light-tailed and heavy-tailed problems. An incomplete list of recent works includes [Asmussen and Kroese 2006; Dupuis et al. 2007; Blanchet and Glynn 2008; Blanchet and Liu 2008; Blanchet et al. 2008]. For the dependent case, [Snijders 2011] proposes several efficient Monte Carlo estimators for the sum of finitely many correlated lognormal random variables. The current problem is a substantial generalization of this work. We employ a different change of measure that is a mixture of infinitely many exponential

change of measures suggesting that the rare-event is mostly caused by the abnormal behavior of the random field at one location.

Another related literature is the extreme behaviors of Gaussian random fields. The results range from general bounds to sharp asymptotic approximations. An incomplete list of works includes [Landau and Shepp 1970; Marcus and Shepp 1970; Sudakov and Tsirelson 1974; Borell 1975; Tsirelson et al. 1976; Berman 1985; Hüsler 1990; Ledoux and Talagrand 1991; Talagrand 1996]. A few lines of investigations on the supremum norm are given as follows. Assuming locally stationary structure, the double-sum method ([Piterbarg 1996]) provides the exact asymptotic approximation of $\sup_T f(t)$ over a compact set T , which is allowed to grow as the threshold tends to infinity. For almost surely at least twice differentiable fields, [Adler 1981; Taylor et al. 2005; Adler and Taylor 2007] derive the analytic form of the expected Euler-Poincaré Characteristics of the excursion set which serves as a good approximation of the tail probability of the supremum. The tube method ([Sun 1993]) takes advantage of the Karhune-Loève expansion and Weyl's formula. The Rice method ([Azais and Wschebor 2005; 2008; 2009]) provides an implicit description of $\sup_T f(t)$. The discussions also go beyond the Gaussian fields such as Gaussian random fields with random variances ([Hüsler et al. 2011]). See also [Adler et al. 2009] for non-Gaussian and heavy-tailed processes. The corresponding rare-event simulations have been studied in [Adler et al. 2012].

The asymptotic tail behaviors of $\mathcal{I}(T)$ have been studied recently in the literature with focus mostly on the analytic approximations of $w(b)$ under smoothness conditions. [Liu 2012] derives the asymptotic approximations of $P(\int_T e^{\sigma f(t)} dt > e^b)$ as $b \rightarrow \infty$ when $f(t)$ is a three-time differentiable and homogeneous Gaussian random field and $\sigma(t)$ takes constant value σ . [Liu and Xu 2012c] further extends the results to the case when the process has a smooth varying mean function, i.e., $P(\int_T e^{\sigma f(t) + \mu(t)} dt > e^b)$.

The tail asymptotics of $\mathcal{I}(T)$ are difficult to develop when $f(t)$ is non-differentiable. Under such a setting, accurate approximations of $w(b)$ have not yet been developed except for some special cases such as $f(t)$ is a Brownian motion (c.f. [Yor 1992; Dufresne 2001]). Therefore, rare-event simulation serves as an appealing alternative from the computational point of view in that the design and the analysis do not require very sharp approximations of $w(b)$. This paper, to the authors' best knowledge, is the first analysis of $\mathcal{I}(T)$ for a general class of non-differentiable and differentiable fields. The main contribution of this paper is to develop a provably efficient rare-event simulation algorithm to compute $w(b)$. The efficiency of the algorithm only requires that $f(t)$ is uniformly Hölder continuous on the compact domain T .

Therefore, the results are applicable to essentially all the Gaussian processes practically in use.

Central to the analysis is a change of measure Q in both a continuous form (for the theoretical analysis) and a discrete form (for the simulation). The measure Q mimics the conditional distribution of f given the occurrence of the rare event $\{\mathcal{I}(T) > e^b\}$. Importance sampling estimators are then constructed based on the measure Q . An appealing feature of the proposed method is that the specific simulation schemes do not vary substantially under different situations. That is, the choice of the change of measure does not rely much on the specific mean, variance, or covariance structure of the process (such as, constant mean/variance, varying mean/variance, multi- and uni-modal mean/variance, etc.). Thus, this change of measure captures the common characteristics of the conditional distributions among all the processes in the class. With a fine tuning of the change of measure, it is conceivable that the efficiency can be further improved by taking into account more refined structures of the process such as homogeneity, smoothness, maxima of the mean and the variance function, etc. Nonetheless, a unified efficient simulation scheme is useful especially when the fine structures of the processes are not easy to obtain. We do not pursue this further improvement in this paper.

The complexity analysis of the proposed estimators consists of two elements. Firstly, since f considered in this paper is continuous, exact simulation of the entire field is usually impossible. Thus, we need to use a discrete object to approximate the continuous field and the bias caused by the discretization needs to be well controlled relative to $w(b)$. The second part of our analysis is the variance control, that is, to provide a bound of the second moments of the (discrete) importance sampling estimators.

The rest of the paper is organized as follows. Section 2 provides the construction of our importance sampling algorithm and presents the main results of this paper. Section 3 includes simulation studies. Proofs of our main theorems are given in Section 4. The proofs of supporting lemmas are included in Section 5.

2. MAIN RESULTS

2.1. Preliminaries of rare-event simulations and importance sampling

Throughout this paper, we are interested in computing $w(b) \rightarrow 0$ as $b \rightarrow \infty$. In the context of rare-event simulations [Asmussen and Glynn 2007, Chapter VI], it is more meaningful to consider the computational error relative to $w(b)$. A well accepted efficiency concept is the so-called *weak efficiency*, also known as asymptotic efficiency and logarithmic efficiency [Asmussen and Glynn 2007].

Definition 2.1. A Monte Carlo estimator L_b is said to be weakly / asymptotically / logarithmically efficient in estimating $w(b)$ if $EL_b = w(b)$ and

$$\lim_{b \rightarrow \infty} \frac{\log EL_b^2}{2 \log w(b)} = 1. \quad (3)$$

Suppose that a weakly efficient estimator L_b has been obtained. Let $\{L_b^{(j)} : j = 1, \dots, n\}$ be i.i.d. copies of L_b and $Z_b = \frac{1}{n} \sum_{j=1}^n L_b^{(j)}$ be the averaged estimator that has a *relative* mean squared error $\text{Var}^{1/2}(L_b)/n^{1/2}w(b)$. A simple consequence of Chebyshev's inequality yields

$$P(|Z_b/w(b) - 1| \geq \eta) \leq \frac{\text{Var}(L_b)}{n\eta^2 w^2(b)}. \quad (4)$$

The limit (3) suggests that for any $\lambda > 0$, $\text{Var}(L_b) \leq EL_b^2 = o(w^{2-\lambda}(b))$. For any positive η and δ , in order to achieve the following relative accuracy

$$P(|Z_b/w(b) - 1| > \eta) < \delta, \quad (5)$$

it is sufficient to generate $n = O(\eta^{-2}\delta^{-1}w^{-\lambda}(b))$ i.i.d. copies of L_b for any $\lambda > 0$.

To construct efficient estimators, in this paper, we use importance sampling for the variance reduction (see [Asmussen and Glynn 2007, Chapter V]). It is based on the following identity

$$w(b) = E\left(1_{(\mathcal{I}(T) > e^b)}\right) = E^Q\left(1_{(\mathcal{I}(T) > e^b)} \frac{dP}{dQ}\right),$$

where Q is a probability measure on \mathcal{F} such that dP/dQ is well defined on the set $\{\mathcal{I}(t) > e^b\}$. We use E and E^Q to denote the expectations under the measures P and Q respectively. Then, the random variable

$$L_b = 1_{(\mathcal{I}(T) > e^b)} \frac{dP}{dQ} \quad (6)$$

is an unbiased estimator of $w(b)$ under the measure Q .

If we choose $Q(\cdot) = P(\cdot | \mathcal{I}(T) > e^b)$ to be the conditional distribution, then the corresponding likelihood ratio $dP/dQ \equiv w(b)$ on the set $\{\mathcal{I}(T) > e^b\}$ has zero variance under Q . Thus, Q is also called the zero-variance change of measure. Note that this change of measure is of no practical value in that its implementation requires the computation of the probabilities $w(b)$. Nonetheless, the measure Q provides a guideline for the construction of an efficient change of measure to compute the probability $w(b)$. Therefore, the task lies in constructing a change

of measure Q that is a good approximation of \mathcal{Q} . In addition, from the computational point of view, we should also be able to numerically compute L_b and to simulate f from Q .

Besides the variance control, another important issue is the bias control. The random fields considered in this paper are continuous processes. Direct simulation is usually not feasible. Therefore, we need to set up an appropriate discretization scheme to approximate the continuous objects. The bias caused by the discretization also needs to be controlled relative to $w(b)$. Suppose that a biased estimator \tilde{L}_b has been constructed for $w(b)$ such that $E(\tilde{L}_b) = \tilde{w}(b)$. Thus, the computation error can be decomposed as follows

$$\left| \frac{\tilde{L}_b}{w(b)} - 1 \right| \leq \frac{|\tilde{L}_b - \tilde{w}(b)|}{w(b)} + \frac{|\tilde{w}(b) - w(b)|}{w(b)}.$$

The first term on the right-hand-side is controlled by the relative variance of \tilde{L}_b and the second term is the bias relative to $w(b)$. Both the bias and the variance control will be carefully analyzed in the following sections relative to $w(b)$. The overall computational complexity is then the necessary number of i.i.d. replicates for \tilde{L}_b multiplied by the computational cost to generate one \tilde{L}_b .

2.2. The change of measure and rare-event simulation

We propose a change of measure Q on the continuous sample path space and further propose a discrete version, denoted by Q_M , for the simulation purpose, where the subscript M indicates the size of the discretization. The description of Q requires the following quantities. For each b , we define u satisfying the following identity

$$\left(\frac{2\pi}{\sigma_T} \right)^{\frac{d}{2}} u^{-\frac{d}{2}} e^{u\sigma_T} = e^b, \quad (7)$$

where

$$\sigma_T = \sup_{t \in T} \sigma(t).$$

The left-hand-side of (7) is in a similar form as the Lambert W -function. Note that when b is large, equation (7) generally has two solutions. One is on the order of b/σ_T ; the other one is close to zero. See Figure 2.2 for an example of the left-hand-side of (7) with $d = 1$ and $\sigma_T = 1$. We choose u to be the larger solution. In the asymptotic regime that b tends to infinity, we need to have $\sup_T \sigma(t)f(t) + \mu(t)$ approximately exceeding level u so that the integral is above the level b .

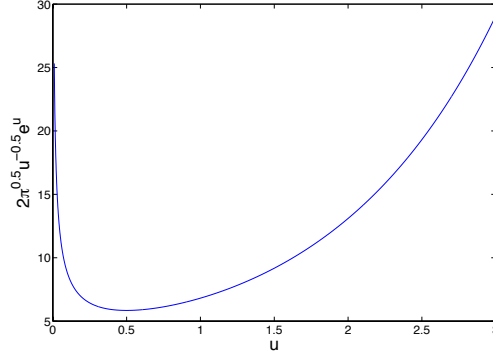


Fig. 1. Plot of $(2\pi)^{\frac{1}{2}} u^{-\frac{1}{2}} e^u$

Furthermore, we define $\mu_\sigma(t) = \mu(t)/\sigma_T$ and

$$u_t = u - \mu_\sigma(t). \quad (8)$$

We characterize the measure Q in two ways. First, we describe the simulation of the process f from Q following a three-step procedure.

- (1) Simulate a random index τ uniformly over T with respect to the Lebesgue measure.
- (2) Given the realized τ , simulate $f(\tau) \sim N(u_\tau, 1)$, where u_τ is defined as in (8).
- (3) Simulate the rest of the field $\{f(t) : t \neq \tau\}$ from the original conditional distribution under P given $(\tau, f(\tau))$.

The above simulation description induces a measure Q . To derive the Radon-Nikodym derivative between Q and P , we need to realize that sampling f from the measure P may also follow a similar three-step procedure except that in Step 2 we sample $f(\tau)$ from the standard normal distribution. Thus, the random variable τ and f are independent under P . We let \mathcal{F} be the σ -field generated by f and \mathcal{F}' be that generated by f and τ . It is straightforward to verify that the Radon-Nikodym derivative dQ/dP on \mathcal{F}' is

$$\tilde{Z}(f, \tau) = \frac{\exp\left\{-\frac{1}{2}(f(\tau) - u_\tau)^2\right\}}{\exp\left\{-\frac{1}{2}f(\tau)^2\right\}}.$$

That is, for each $A' \in \mathcal{F}'$, $Q(A') = E(\tilde{Z}(f, \tau); A')$. Then, for each $A \subset C(T)$

$$Q(f \in A) = E\left[\frac{\exp\left\{-\frac{1}{2}(f(\tau) - u_\tau)^2\right\}}{\exp\left\{-\frac{1}{2}f(\tau)^2\right\}}; f \in A\right] = E\left\{E\left[\frac{\exp\left\{-\frac{1}{2}(f(\tau) - u_\tau)^2\right\}}{\exp\left\{-\frac{1}{2}f(\tau)^2\right\}} \middle| f\right]; f \in A\right\}. \quad (9)$$

Note that τ and f are independent under P and τ is uniform on T , thus the conditional expectation given f is written as

$$Z(f) = E \left[\frac{\exp \left\{ -\frac{1}{2} (f(\tau) - u_\tau)^2 \right\}}{\exp \left\{ -\frac{1}{2} f(\tau)^2 \right\}} \middle| f \right] = \int_T \frac{1}{\text{mes}(T)} \cdot \frac{\exp \left\{ -\frac{1}{2} (f(t) - u_t)^2 \right\}}{\exp \left\{ -\frac{1}{2} f(t)^2 \right\}} dt,$$

where $\text{mes}(\cdot)$ denotes the Lebesgue measure. We insert the above form into (9) and obtain that $Q(f \in A) = E(Z(f); f \in A)$. Therefore, $Z(f)$ is the Radon-Nikodym derivative between Q and P restricted on \mathcal{F} . Given that we are only interested in the tail event regarding f (not τ), we work with $Z(f)$ most of the time. By slightly abusing notation, we write

$$\frac{dQ}{dP} = Z(f) = \int_T \frac{1}{\text{mes}(T)} \cdot \frac{\exp \left\{ -\frac{1}{2} (f(t) - u_t)^2 \right\}}{\exp \left\{ -\frac{1}{2} f(t)^2 \right\}} dt = \int_T \frac{1}{\text{mes}(T)} \cdot \exp \left\{ -\frac{1}{2} u_t^2 + f(t) u_t \right\} dt. \quad (10)$$

Remark 2.2. To better understand the connection between the above simulation procedure and the likelihood ratio (10), we present a discrete analogue for a finite dimensional multivariate Gaussian random vector $X = (X_1, \dots, X_n)$. For the finite dimensional case, τ is uniformly distributed over $\{1, \dots, n\}$. The density function of X under the change of measure is

$$q(x_1, \dots, x_n) = \frac{1}{n} \sum_{\tau=1}^n q_\tau(x_\tau) f(x_{-\tau} | x_\tau)$$

where $x_{-\tau} = (x_1, \dots, x_{\tau-1}, x_{\tau+1}, \dots, x_n)$, $q_\tau(x)$ is the sampling distribution of x_τ given τ , and f is the density under the original measure. Thus, the likelihood ratio is

$$\frac{q(x_1, \dots, x_n)}{f(x_1, \dots, x_n)} = \frac{1}{n} \sum_{\tau=1}^n \frac{q_\tau(x_\tau)}{f(x_\tau)}$$

that is the discrete analogue of (10).

Based on the above discussion, we have the (continuous version) importance sampling estimator taking the form

$$L_b = 1_{(\mathcal{I}(T) > e^b)} \left(\int_T \frac{1}{\text{mes}(T)} \cdot \exp \left\{ -\frac{1}{2} u_t^2 + f(t) u_t \right\} dt \right)^{-1}$$

and its second moment equals

$$E^Q [L_b^2] = E^Q \left[\left(\int_T \frac{1}{\text{mes}(T)} \cdot \exp \left\{ -\frac{1}{2} u_t^2 + f(t) u_t \right\} dt \right)^{-2}; \mathcal{I}(T) > e^b \right].$$

The measure Q is constructed such that the behavior of f under Q mimics the tail behavior of f given the rare event $\{\mathcal{I}(T) > e^b\}$ under P . According to the above simulation procedure, a random variable τ is first sampled uniformly over T , then $f(\tau)$ is simulated with a large mean at level u_τ . Under the zero-variance change of measure, the large value of the integral $\mathcal{I}(T)$ is mostly caused by the high excursion of $f(t)$ at one location. The random index τ searches the maximum of $f(t)$ over the index set T . It worth emphasizing that τ is not necessarily the exact maximum but should be very close to it.

In the case when $f(t)$ is differentiable and strictly stationary and $\mu(t) \equiv 0$ and $\sigma(t) \equiv 1$, the zero-variance change of measure can be more precisely quantified. For the stationary case, u_t is a constant and we write it as u . Then, the zero-variance change of measure is approximated as follows

$$\sup_A \left| \mathcal{Q}(A) - P(f \in A | \sup_T \gamma_u(t) > u) \right| \rightarrow 0 \quad \text{as } b \rightarrow \infty,$$

where $\mathcal{Q}(\cdot) = P(f \in \cdot | \mathcal{I}(T) > e^b)$ and $\gamma_u(t) = f(t) + \frac{\text{Tr}(\nabla^2 f(t))}{2\sigma u} + \kappa_0$. $\nabla^2 f(t)$ is the Hessian matrix, $\text{Tr}(\cdot)$ is the trace operator, and κ_0 is a constant only depending on the covariance function. See [Liu and Xu 2012a] for the detailed description of the above results. According to the total variation approximation, the high excursion of $\mathcal{I}(T)$ is almost the same as the high excursion of $f(t)$ with a small correction depending on the Hessian matrix.

The propose measure Q is different from \mathcal{Q} mostly in two ways. First, under \mathcal{Q} , the overshoot of $\gamma_u(t) \approx f(t)$ over the level u is of order $O(u^{-1})$; under Q , the overshoot is of order $O(1)$. Second, the measure Q does not tilt the distribution of $\nabla^2 f(t)$. These are the main sources of inefficiency. On the other hand, the Radon-Nikodym derivative dQ/dP takes a very friendly form that one can take advantage of the Jensen's inequality to prove weak efficiency for a general class of Gaussian processes, especially for non-differentiable processes, for which there is no quantitative result for the zero-variance change of measure.

2.3. The algorithm and efficiency results

For the implementation, we introduce a suitable discretization scheme on T . For any positive integer N , let $G_{N,d}$ be a subset of R^d

$$G_{N,d} = \left\{ \left(\frac{i_1}{N}, \frac{i_2}{N}, \dots, \frac{i_d}{N} \right) : i_1, \dots, i_d \in \mathbb{Z} \right\},$$

where \mathbb{Z} is the set of integers. That is, $G_{N,d}$ is a regular lattice on R^d . For each $t = (t^1, \dots, t^d) \in G_{N,d}$, define

$$T_N(t) = \{(s^1, \dots, s^d) \in T : s^j \in (t^j - 1/N, t^j] \text{ for } j = 1, \dots, d\}$$

that is the $\frac{1}{N}$ -cube intersected with T and upper-cornered at t . Furthermore, let

$$T_N = \{t \in G_{N,d} : \text{mes}(T_N(t)) > 0\}. \quad (11)$$

Since T is compact, T_N is a finite set. We enumerate the elements in $T_N = \{t_1, \dots, t_M\}$, where $M \sim \text{mes}(T)N^d$.

We use

$$w_M(b) = P(\mathcal{I}_M(T) > e^b)$$

as an approximation of $w(b)$ where

$$\mathcal{I}_M(T) = \sum_{i=1}^M \text{mes}(T_N(t_i)) \times e^{\sigma(t_i)f(t_i) + \mu(t_i)}. \quad (12)$$

We now state the main results that require the following technical conditions.

A1 $f(t)$ is almost surely continuous with respect to t and furthermore admits $E[f(t)] = 0$ and $E[f^2(t)] = 1$ for all $t \in T$;

A2 There exist $\delta, \kappa_H > 0$, and $\beta \in (0, 1]$ such that, for all $|s - t| < \delta$, the mean and variance functions satisfy

$$|\mu(t) - \mu(s)| + |\sigma(t) - \sigma(s)| \leq \kappa_H |s - t|^\beta;$$

A3 For each $s, t \in T$, define the covariance function

$$C(s, t) = \text{Cov}(f(s), f(t)).$$

For all $|s - s'| < \delta$ and $|t - t'| < \delta$, the covariance function satisfies

$$|C(t, s) - C(t', s')| \leq \kappa_H (|t - t'|^{2\beta} + |s - s'|^{2\beta}).$$

Condition A1 assumes that f has zero mean and unit variance. For a general Gaussian random field with mean $\mu(t)$ and variance $\sigma^2(t)$, we treat the mean and the standard deviation as

additional parameters. Conditions A2 and A3 essentially ensure that the process $\mu(t) + \sigma(t)f(t)$ is uniformly Hölder continuous, that is, for $|s - t| < \delta$, $\text{Var}(f(s) - f(t)) \leq 2\kappa_H |s - t|^{2\beta}$. These assumptions are weak enough such that they accommodate essentially all Gaussian processes practically in use, such as fractional Brownian motion, smooth Gaussian processes, etc. Therefore, the algorithm developed in this paper is suitable for a wide range of applications. The first result controls the relative bias of $w_M(b)$.

THEOREM 2.3. *Consider a Gaussian random field f satisfying conditions A1-A3. For any $\epsilon > 0$, there exists a constant κ_0 such that for any $\eta > 0$, $b > 1$, and the lattice size $N = \kappa_0^{1/\beta} |\log \eta|^{1/\beta} \eta^{-1/\beta} b^{2(1+\epsilon)/\beta}$,*

$$\frac{|w_M(b) - w(b)|}{w(b)} < \eta,$$

where β is given as in conditions A2 and A3.

Thanks to the continuity of $f(t)$, as N tends to infinity, the lattice T_N becomes dense in T . The finite sum $\mathcal{I}_M(T)$ converges in probability to $\mathcal{I}(T)$ and therefore $w_M(b)/w(b)$ converges to 1. On the other hand, the convergence of $w_M(b)/w(b)$ is not uniform in b . The above theorem provides a lower bound of N such that $w_M(b)/w(b)$ is close enough to unity and the relative bias can be controlled.

With N chosen as in Theorem 2.3, we proceed to estimating $w_M(b)$ by importance sampling, which is based on the change of measure proposed in (10). We make the corresponding adaptation under the above discretization T_N . In particular we define a measure Q_M as the discrete version of Q such that dQ_M/dP takes the form:

$$\frac{dQ_M}{dP} = \sum_{i=1}^M \frac{1}{M} \frac{e^{-\frac{1}{2}(f(t_i) - u_{t_i})^2}}{e^{-\frac{1}{2}f(t_i)^2}} = \sum_{i=1}^M \frac{1}{M} e^{u_{t_i} f(t_i) - \frac{1}{2}u_{t_i}^2}. \quad (13)$$

The computation of the above likelihood ratio and the event $\mathcal{I}_M(T) > b$ only consists of $\{f(t_i) : i = 1, \dots, M\}$. For the simulation under the measure Q_M , we propose the following algorithm.

It is not hard to verify that the above simulation procedure is consistent with the likelihood ratio (13) and thus $\tilde{L}_b = 1_{\{\mathcal{I}_M(T) > e^b\}} \frac{dP}{dQ_M}$ is an unbiased estimator of $w_M(b)$. The next theorem controls the variance of the estimator \tilde{L}_b .

The algorithm has two steps:

Step 1: Simulate ι uniformly from $\{1, \dots, M\}$ and generate $f(t_\iota)$ from $N(u_{t_\iota}, 1)$. Given $(\iota, f(t_\iota))$, simulate $(f(t_1), \dots, f(t_{\iota-1}), f(t_{\iota+1}), \dots, f(t_M))$ from the original conditional distribution under the measure P .

Step 2: Compute and output

$$\tilde{L}_b = \frac{1_{\{\mathcal{I}_M(T) > \epsilon^b\}}}{\sum_{i=1}^M \frac{1}{M} e^{u_{t_i} f(t_i) - \frac{1}{2} u_{t_i}^2}}. \quad (14)$$

THEOREM 2.4. *Suppose that f is a Gaussian random field satisfying conditions A1-A3. If N is chosen as in Theorem 2.3, then*

$$\lim_{b \rightarrow \infty} \frac{\log E^{Q_M} \tilde{L}_b^2}{2 \log w_M(b)} = 1.$$

The above results show that the estimator \tilde{L}_b in Algorithm 1 is asymptotically efficient in estimating $w_M(b)$. To estimate $w(b)$, we simulate n i.i.d. copies of \tilde{L}_b , $\{\tilde{L}_b^{(j)} : j = 1, \dots, n\}$ and the final estimator is $Z_b = \frac{1}{n} \sum_{j=1}^n \tilde{L}_b^{(j)}$. The estimation error is

$$|Z_b - w(b)| \leq |w_M(b) - w(b)| + |Z_b - w_M(b)|. \quad (15)$$

The first term is controlled by Theorem 2.3, i.e., $|w_M(b) - w(b)| \leq \eta w(b)$ if we choose the discretization size

$$N = O(|\log \eta|^{1/\beta} \eta^{-1/\beta} b^{2(1+\epsilon)/\beta}).$$

The second term of (15) is controlled by the discussion as in (4) if we choose the number of replicates

$$n = O(\eta^{-2} \delta^{-1} w^{-\lambda}(b)).$$

The simulation of \tilde{L}_b consists of generating a random vector of dimension $M = O(N^d)$. Note that the complexity of computing the eigenvalues and eigenvectors or the Cholesky decomposition of an M -dimension matrix is $O(M^3) = O(N^{3d})$. Thus, the total computational complexity to achieve the prescribed accuracy in (5) is $O(N^{3d} \eta^{-2} \delta^{-1} w^{-\eta'}(b))$.

The current choice of measure Q as in (10) does not depend on the particular form of the covariance structure. When there is more knowledge available, we can further tune and adapt the measure Q to more refined structures and to improve the efficiency. Notice that the measure Q twists the process f at the random location τ . The main tuning parameters are the

distribution of τ (which is currently chosen to be uniform) and the distribution of $f(\tau)$ (which is currently chosen to be $N(u_\tau, 1)$). Therefore, the general form of the change of measure Q is

$$\frac{dQ}{dP} = \int_T h(t) \frac{g_t(f(t))}{\varphi_t(f(t))} dt$$

where $h(t)$ is the density of τ , $g_t(\cdot)$ is the sampling density of $f(\tau)$ given that $\tau = t$, and $\varphi_t(\cdot)$ is the density of $f(t)$ under P . We may refine the choice of g_t and h to take into account the specific structures of f . For instance, if $\mu(t)$ has one unique maximum attained at t_* , then it would be more efficient to choose $h(t)$ concentrating around t_* . The theoretical analysis also needs to be adapted to the specific choices of h and g_t . The most difficult part of the analysis lies in obtaining more accurate asymptotic approximations of $w(b)$ so as to justify stronger type of efficiency. This is particularly challenging when $f(t)$ is not differentiable and it is beyond the topic of the current paper. In this paper, we stick to the unified simulation scheme that is applicable to a large class of processes and admits an acceptable efficiency property.

Remark 2.5. One limitation of the current analysis is that the total computational complexity grows exponentially fast with dimension d . This is mainly due to the regular discretization method. One may alternatively use other numerical methods such as quasi-Monte Carlo, randomized quasi-Monte Carlo, or Monte Carlo, whose complexities do not depend on dimension, to approximate the integral $\int e^{f(t)} dt$. In the literature of quasi-Monte Carlo and randomized quasi-Monte Carlo low-discrepancy sequences have been developed. These more refined choices may further reduce bias compared to the regular lattice $G_{N,d}$. For more detailed analysis, see [L'Ecuyer and Munger 2010; L'Ecuyer et al. 2010; L'Ecuyer 2009]. We do not perform rigorous analysis along this line that is beyond the scope of this paper.

3. SIMULATION

To illustrate the proposed algorithm, we first apply it to homogeneous Gaussian random fields $\{f(t), t \in T = [0, 1]^d\}$ with dimension $d = 1$ and 2. For the not-so-small tail probabilities, we also use crude Monte Carlo to compute them as a validation of the importance sampling algorithm.

For each case, we assume that f has zero mean and covariance function

$$C(s, t) = e^{-|t-s|^\alpha}, \tag{16}$$

where $\|\cdot\|$ is the L_2 -norm and α is taken to be 1 and 2 corresponding to different covariance structures of f . When $\alpha = 2$, f is infinitely differentiable; when $\alpha = 1$, f is non-differentiable. We discretize T following the procedure in Section 2.2.

For $d = 1$, we take $T_N = \{i/100 : i = 1, \dots, 100\}$ with discretization size $N = 100$. The detailed simulation is described in the next steps.

1. Generate a random variable $\iota \sim \text{Uniform}\{1, 2, \dots, 100\}$.
2. Simulate $f\left(\frac{\iota}{100}\right) \sim N(u, 1)$, where u is calculated from equation (7).
3. Given ι and $f\left(\frac{\iota}{100}\right)$, simulate $\{f\left(\frac{i}{100}\right), i = 1, \dots, \iota - 1, \iota + 1, \dots, 100\}$ under the covariance structure specified in (16).

First, we consider the one-dimensional case with constant $\sigma(t)$ and $\mu(t)$. Let $\mu(t) = 0$ and $\sigma(t) = 1$. Then the tail probability of interest takes the form

$$w(b) = P\left(\int_0^1 e^{f(t)} dt > e^b\right).$$

The estimated tail probabilities $w(b)$ along with the estimated standard deviations $Std^Q(\tilde{L}_b) = \sqrt{\text{Var}^Q(\tilde{L}_b)}$ are shown in Table I. All the results are based on 10^4 independent simulations. The standard deviation of the final estimate (in the column ‘‘Est.’’) is the reported standard deviation (in the column of ‘‘Std.’’) divided by 100. Comparing the simulation results of $\alpha = 1$ and 2, we can see that the algorithm has a smaller relative error when $\alpha = 2$. The CPU time to generate 10^4 samples is less than one second. To validate the simulation results, we use crude Monte Carlo for $b = 3$ and 5. Based on 10^6 independent simulations, for $b = 3$, the estimated tail probabilities are $4.7\text{e-}4$ (Std. $2\text{e-}5$) and $8.2\text{e-}4$ (Std. $3\text{e-}5$) when $\alpha = 1$ and 2, respectively; based on 10^9 independent simulations, for $b = 5$, the estimated tail probabilities are $1.2\text{e-}8$ (Std. $3\text{e-}9$) and $7.7\text{e-}8$ (Std. $9\text{e-}9$) when $\alpha = 1$ and 2, respectively. These results are consistent with those computed by the importance sampling estimators.

We also consider the non-constant mean and variances. In particular, we choose $\sigma(t) = 1 - |t - 0.5|^2$ and $\mu(t) = |2t - 1|$. The corresponding simulation results for $w(b) = P\left(\int_0^1 e^{\sigma(t)f(t)+\mu(t)} dt > e^b\right)$ are shown in Table II. The CPU time to generate 10^4 samples is less than one second. For $b = 3$, crude Monte Carlo estimator based on 10^6 independent simulations gives the estimated tail probabilities $1.4\text{e-}3$ (Std. $4\text{e-}5$) and $2.3\text{e-}3$ (Std. $5\text{e-}5$) when $\alpha = 1$ and $\alpha = 2$, respectively; for $b = 5$, crude Monte Carlo estimator based on 10^9 independent simulations gives the estimated tail probabilities $1.8\text{e-}8$ (Std. $4\text{e-}9$) and $1.4\text{e-}7$ (Std. $1\text{e-}8$) when $\alpha = 1$ and $\alpha = 2$, respectively.

Table I. Estimates of $w(b)$ on $T = [0, 1]$, $\sigma(t) = 1$, and $\mu(t) = 0$

	b	$Est.$	$Std.$	$Std./Est.$
$\alpha = 1$	$b = 3$	4.47e-04	1.41e-03	3.14
	$b = 5$	1.13e-08	6.33e-08	5.60
	$b = 7$	1.80e-15	1.66e-14	9.23
$\alpha = 2$	$b = 3$	8.49e-04	2.09e-03	2.46
	$b = 5$	7.03e-08	2.30e-07	3.27
	$b = 7$	8.27e-14	3.46e-13	4.19

Estimates of $w(b)$ for $T = [0, 1]$ with $\sigma(t) = 1$ and $\mu(t) = 0$. All results are based on 10^4 independent simulations. The standard deviation of the estimate is $Std./100$.

For the case that $d = 2$, we start with constant mean and variance, $\mu(t) = 0$ and $\sigma(t) = 1$. Then the tail probability of interest takes the form $w(b) = P(\int_{[0,1]^2} e^{f(t)} dt > e^b)$. Similarly, we discretize T and take $T_N = \{(i, j)/100 : i, j = 1, \dots, 100\}$ with the discretization size $N = 100 \times 100$. Table III shows the estimated tail probabilities $w(b)$ along with $Std^Q(\tilde{L}_b)$. In addition, we perform crude Monte Carlo for $b = 3$. The estimated tail probabilities, based on 10^5 independent simulations, are 1.8e-4 (Std. 4e-5) and 5.3e-4 (Std. 7e-5) when $\alpha = 1$ and $\alpha = 2$, respectively. Furthermore, Table IV gives estimated tail probabilities for non-constant functions $\sigma(t) = 1 - \|t - (0.5, 0.5)\|_2^2$ and $\mu(t) = \|2t - (1, 1)\|_1$, where $\|\cdot\|_1$ and $\|\cdot\|_2$ are the L_1 - and L_2 -norms respectively. The CPU time to generate 10^4 samples varies from from 1 to 15 minutes. For $b = 3$, the crude Monte Carlo based on 10^5 independent simulations gives 3.3e-3 (Std. 2e-4) and 5.6e-3 (Std. 2e-4) for $\alpha = 1$ and $\alpha = 2$, respectively.

The simulation results show that the coefficients of variation ($Std^Q(L_b)/w(b)$) increase as the tail probabilities become smaller. Nonetheless, the coefficients of variation stay reasonably small when the probability is as small as 10^{-7} . The continuity of the process and the dimension do affect the empirical performance of the algorithm. More precisely, the algorithm admits smaller coefficients of variation when the process is more continuous (corresponding to a larger value of α) and the domain T is of a lower dimension. In addition, for stationary processes, the algorithm has a slightly better performance than the nonstationary cases. This is because, for the stationary cases, the uniform distribution of τ is closer to the distribution of the maximum of f under \mathcal{Q} .

4. PROOF OF THE THEOREMS

The proofs of the theorems need several supporting lemmas. To smooth the discussion, we provide their statements at places where they are used and delay their proofs to Section 5.

Table II. Estimates of $w(b)$ on $T = [0, 1]$, $\sigma(t) = 1 - |t - 0.5|^2$, and $\mu(t) = |2t - 1|$

	b	<i>Est.</i>	<i>Std.</i>	<i>Std./Est.</i>
$\alpha = 1$	$b = 3$	1.39e-03	3.57e-03	2.56
	$b = 5$	2.22e-08	1.31e-07	5.87
	$b = 7$	2.97e-15	4.45e-14	14.93
$\alpha = 2$	$b = 3$	2.30e-03	4.93e-03	2.14
	$b = 5$	1.31e-07	4.66e-07	3.57
	$b = 7$	1.19e-13	6.86e-13	5.77

Estimates of $w(b)$ on $T = [0, 1]$ with $\sigma(t) = 1 - |t - 0.5|^2$ and $\mu(t) = |2t - 1|$. All results are based on 10^4 independent simulations. The standard deviation of the estimate is *Std./100*.

Table III. Estimates of $w(b)$ on $T = [0, 1]^2$, $\sigma(t) = 1$, and $\mu(t) = 0$

	b	<i>Est.</i>	<i>Std.</i>	<i>Std./Est.</i>
$\alpha = 1$	$b = 3$	2.01e-04	1.77e-03	8.81
	$b = 5$	1.16e-09	1.58e-08	13.64
	$b = 7$	1.51e-17	5.79e-16	38.48
$\alpha = 2$	$b = 3$	5.04e-04	2.81e-03	5.57
	$b = 5$	1.46e-08	7.54e-08	5.18
	$b = 7$	4.04e-15	3.11e-14	7.70

Estimates of $w(b)$ on $T = [0, 1]^2$ with $\sigma(t) = 1$ and $\mu(t) = 0$. All results are based on 10^4 independent simulations. The standard deviation of the estimate is *Std./100*.

Table IV. Estimates of $w(b)$ on $T = [0, 1]^2$, $\sigma(t) = 1 - |t - (0.5, 0.5)|^2$, and $\mu(t) = |2t - (1, 1)|_1$

	b	<i>Est.</i>	<i>Std.</i>	<i>Std./Est.</i>
$\alpha = 1$	$b = 3$	2.89e-03	1.18e-02	4.07
	$b = 5$	5.02e-09	5.60e-08	11.16
	$b = 7$	1.30e-17	4.04e-16	31.13
$\alpha = 2$	$b = 3$	5.15e-03	1.58e-02	3.06
	$b = 5$	6.33e-08	3.09e-07	4.88
	$b = 7$	4.25e-15	3.61e-14	8.50

Estimates of $w(b)$ on $T = [0, 1]^2$ with $\sigma(t) = 1 - |t - (0.5, 0.5)|^2$ and $\mu(t) = |2t - (1, 1)|_1$. All results are based on 10^4 independent simulations. The standard deviation of the estimate is *Std./100*.

PROOF OF THEOREM 2.3. For any $\eta, \epsilon > 0$ and a large constant $\kappa_0 > 0$ (to be determined later), let

$$\mathcal{L} = \left\{ f(\cdot) \in \mathcal{C}(T) : \sup_{s,t:|t-s|\leq dN^{-1}} |\sigma(s)f(s) + \mu(s) - (\sigma(t)f(t) + \mu(t))| \leq \eta \kappa_0^{-1/3} b^{-(1+\epsilon)} \right\}, \quad (17)$$

where $\mathcal{C}(T)$ is the set of all continuous functions on T and N is chosen as in the statement of the theorem. The following lemma suggests that we only need to focus on the set \mathcal{L} .

LEMMA 4.1. *Under the conditions in Theorem 2.3, for any $\epsilon > 0$, there exists a constant κ_0 such that, for any $\eta > 0$, $b > 1$, and $N = \kappa_0^{1/\beta} |\log \eta|^{1/\beta} \eta^{-1/\beta} b^{(2+2\epsilon)/\beta}$, we have*

$$P(\mathcal{I}(T) > e^b, \mathcal{L}^c) \leq \eta w(b), \quad P(\mathcal{I}_M(T) > e^b, \mathcal{L}^c) \leq \eta w(b). \quad (18)$$

With N chosen as in Lemma 4.1, we have that

$$\begin{aligned} & |w_M(b) - w(b)| \\ & \leq P(\mathcal{I}_M(T) > e^b, \mathcal{I}(T) < e^b, \mathcal{L}) + P(\mathcal{I}_M(T) < e^b, \mathcal{I}(T) > e^b, \mathcal{L}) + 2\eta w(b). \end{aligned} \quad (19)$$

On the set \mathcal{L} , we have that

$$\begin{aligned} & |\mathcal{I}_M(T) - \mathcal{I}(T)| \\ & = \left| \sum_{i=1}^M \int_{T_N(t_i)} \left[e^{\sigma(t_i)f(t_i)+\mu(t_i)} - e^{\sigma(t)f(t)+\mu(t)} \right] dt \right| \\ & \leq 2 \left| \sum_{i=1}^M \text{mes}(T_N(t_i)) \times e^{\sigma(t_i)f(t_i)+\mu(t_i)} \times \sup_{|t-t_i| \leq dN^{-1}} |\sigma(t_i)f(t_i) + \mu(t_i) - (\sigma(t)f(t) + \mu(t))| \right| \\ & \leq 2\eta\kappa_0^{-1/3} b^{-(1+\epsilon)} \mathcal{I}_M(T). \end{aligned}$$

Recall that t_i is the corner point of $T_N(t_i)$. The first inequality in the above display is an application of Taylor's expansion. A similar argument yields that

$$|\mathcal{I}_M(T) - \mathcal{I}(T)| \leq 2\eta\kappa_0^{-1/3} b^{-(1+\epsilon)} \mathcal{I}(T).$$

Thus, the first term in (19) is bounded by

$$P(\mathcal{I}_M(T) > e^b, \mathcal{I}(T) < e^b, \mathcal{L}) \leq P\left(e^b(1 - 2\eta\kappa_0^{-1/3} b^{-(1+\epsilon)}) < \mathcal{I}(T) < e^b, \mathcal{L}\right)$$

and the second term is bounded by

$$P(\mathcal{I}_M(T) < e^b, \mathcal{I}(T) > e^b, \mathcal{L}) \leq P\left(e^b < \mathcal{I}(T) < e^b(1 + 2\eta\kappa_0^{-1/3} b^{-(1+\epsilon)}), \mathcal{L}\right).$$

We insert the above bounds back to (19). There exists some constant $c > 0$ such that

$$\begin{aligned}
& |w_M(b) - w(b)| \\
& \leq P\left(e^b(1 - 2\eta\kappa_0^{-1/3}b^{-(1+\epsilon)}) < \mathcal{I}(T) < e^b(1 + 2\eta\kappa_0^{-1/3}b^{-(1+\epsilon)}), \mathcal{L}\right) + 2\eta w(b) \\
& \leq P\left(b + \log(1 - 2\eta\kappa_0^{-1/3}b^{-(1+\epsilon)}) < \log \mathcal{I}(T) < b + \log(1 + 2\eta\kappa_0^{-1/3}b^{-(1+\epsilon)}), \mathcal{L}\right) + 2\eta w(b) \\
& \leq c\eta\kappa_0^{-1/3}b^{-(1+\epsilon)}F'(b - c\eta\kappa_0^{-1/3}b^{-(1+\epsilon)}) + 2\eta w(b), \tag{20}
\end{aligned}$$

where $F'(x)$ is the probability density function of $\log \mathcal{I}(T)$. The following lemma provides an upper bound of the density function $F'(x)$.

LEMMA 4.2. *Under the conditions of Theorem 2.3, let $F'(x)$ be the probability density function of $\log \mathcal{I}(T)$. Then $F'(x)$ exists almost everywhere. Moreover, for all $\epsilon > 0$ and $\lambda > 0$,*

$$F'(x) = o(1)x^{1+\epsilon/2} \cdot w(x + x^{-1-\lambda}) \quad \text{as } x \rightarrow \infty. \tag{21}$$

We apply Lemma 4.2 to (20) by setting $x = b - c\eta\kappa_0^{-1/3}b^{-(1+\epsilon)}$. Furthermore, we choose λ small such that $x + x^{-1-\lambda} = b$ and thus

$$F'(b - c\eta\kappa_0^{-1/3}b^{-(1+\epsilon)}) \leq o(1)b^{1+\epsilon/2}w(b).$$

We insert the above bound to the right-hand-side of (20) and obtain that for κ_0 large enough,

$$|w_M(b) - w(b)| \leq c\kappa_0^{-1/3}\eta b^{-(1+\epsilon)}o(1)b^{1+\epsilon/2}w(b) + 2\eta w(b) \leq 3\eta w(b).$$

Then, we can redefine η and κ_0 and conclude the conclusion. ■

PROOF OF THEOREM 2.4.

The main idea. The key element of this proof is an application of the Jensen's inequality that for all u and $f(t)$

$$\frac{1}{M} \sum_{i=1}^M e^{uf(t_i)} \geq \left[\frac{1}{M} \sum_{i=1}^M e^{f(t_i)} \right]^u.$$

Thus, if $\frac{1}{M} \sum_{i=1}^M e^{f(t_i)} > e^b$, then $\frac{1}{M} \sum_{i=1}^M e^{uf(t_i)} > e^{ub}$. The following technical proof is to write the likelihood ratio and $I_M(T)$ in a form that the above inequality is applicable. Thus, we are able to develop an upper bound of the second moment of the likelihood ratio on the set $\{\mathcal{I}_M(T) > e^b\}$.

The technical proof. We first present a lemma that provides not-so-accurate but useful bounds of $w(b)$.

LEMMA 4.3. *Under Conditions A1-A3, there exist constants \tilde{c}_0 , \tilde{c}_1 and \tilde{c}_2 such that*

$$\exp\left(-\frac{b^2 + \tilde{c}_1 b \log b + \tilde{c}_0}{2\sigma_T^2}\right) \leq w(b) \leq \exp\left(-\frac{b^2}{2\sigma_T^2} + \tilde{c}_2 b\right).$$

Consequently,

$$-\lim_{b \rightarrow \infty} \frac{\log w(b)}{b^2} \rightarrow \frac{1}{2\sigma_T^2}.$$

Under the change of measure, the discrete likelihood ratio is

$$\frac{dQ_M}{dP} = \sum_{i=1}^M \frac{1}{M} \frac{e^{-\frac{1}{2}(f(t_i) - u_{t_i})^2}}{e^{-\frac{1}{2}f(t_i)^2}} = \sum_{i=1}^M \frac{1}{M} e^{-\frac{1}{2}u_{t_i}^2 + f(t_i)u_{t_i}}.$$

Note that most $T_N(t_i)$ are rectangles and there are just a few on the boundary of T that are not rectangles. If we let $\kappa_1 = 2\text{mes}(T)$, then $\mathcal{I}_M(T)$ is bounded from the above by

$$\frac{\kappa_1}{N^d} \sum_{i=1}^M e^{\sigma(t_i)f(t_i) + \mu(t_i)} \geq \sum_{i=1}^M e^{\sigma(t_i)f(t_i) + \mu(t_i)} \cdot \text{mes}(T_N(t_i)) = \mathcal{I}_M(T).$$

That is, those $T_N(t_i)$'s on the boundary of T are replaced by rectangles. Thus, the second moment of the estimator is bounded by

$$\begin{aligned} & E^{Q_M} \left[\left(\frac{dP}{dQ_M} \right)^2 ; \mathcal{I}_M(T) > e^b \right] \\ & \leq E^{Q_M} \left[\left(\sum_{i=1}^M \frac{1}{M} e^{-\frac{1}{2}u_{t_i}^2 + f(t_i)u_{t_i}} \right)^{-2} ; \frac{\kappa_1}{N^d} \sum_{i=1}^M e^{\sigma(t_i)f(t_i) + \mu(t_i)} > e^b \right] \end{aligned} \quad (22)$$

For the next step, we wish to take the term “ $-\frac{1}{2}u_{t_i}^2$ ” in the exponential out of the expectation. Note that $-\frac{1}{2}u_{t_i}^2 \geq -\frac{1}{2}(u - \min_{t \in T} \mu_\sigma(t))^2$ and we continue the above calculation

$$\leq e^{(u - \min_{t \in T} \mu_\sigma(t))^2} E^{Q_M} \left[\left(\sum_{i=1}^M \frac{1}{M} e^{f(t_i)u_{t_i}} \right)^{-2} ; \frac{\kappa_1}{N^d} \sum_{i=1}^M e^{\sigma(t_i)f(t_i) + \mu(t_i)} > e^b \right]. \quad (23)$$

We now split the set $\{t_1, \dots, t_M\}$ into $\{t_i : f(t_i) > 0\}$ and $\{t_i : f(t_i) \leq 0\}$. Furthermore, there exists κ such that

$$\frac{1}{N^d} \sum_{\{t_i: f(t_i) \leq 0\}} e^{\sigma(t_i)f(t_i)+\mu(t_i)} \leq \kappa.$$

Thus, there exists $\delta_0 > 0$ such that, for b sufficiently large and on the set $\{\frac{\kappa_1}{N^d} \sum_{i=1}^M e^{\sigma(t_i)f(t_i)+\mu(t_i)} > e^b\}$, we have $\kappa < (1 - \delta_0)e^b$ and

$$\sum_{\{i:f(t_i)>0\}} e^{\sigma(t_i)f(t_i)+\mu(t_i)} \geq \delta_0 \sum_{i=1}^M e^{\sigma(t_i)f(t_i)+\mu(t_i)}. \quad (24)$$

Therefore, $\frac{\kappa_1}{N^d} \sum_{i=1}^M e^{\sigma(t_i)f(t_i)+\mu(t_i)} > e^b$ implies that

$$\frac{1}{M} \sum_{i=1}^M e^{\sigma_T f(t_i) + \max_{t \in T} \mu(t)} \geq \frac{1}{M} \sum_{\{i:f(t_i) \geq 0\}} e^{\sigma(t_i)f(t_i)+\mu(t_i)} \geq \frac{\delta_0 N^d}{M \kappa_1} e^b. \quad (25)$$

We now consider the behavior of the likelihood ratio on the set $\{\frac{\kappa_1}{N^d} \sum_{i=1}^M e^{\sigma(t_i)f(t_i)+\mu(t_i)} > e^b\}$.

For $f(t) > 0$, we have that

$$u_t f(t) \geq (u - \max \mu_\sigma(t)) f(t) = \frac{u - \max \mu_\sigma(t)}{\sigma_T} [\sigma_T f(t) + \max \mu(t)] - (u - \max \mu_\sigma(t)) \times \max \mu_\sigma(t).$$

For u large enough, we have that

$$\begin{aligned} \frac{1}{M} \sum_{i=1}^M e^{u_{t_i} f(t_i)} &\geq \frac{1}{M} \sum_{i:f(t_i)>0} e^{(u - \max \mu_\sigma(t)) f(t_i)} \\ &= \frac{1}{M} \sum_{i:f(t_i)>0} \exp \left\{ \frac{u - \max_t \mu_\sigma(t)}{\sigma_T} [\sigma_T f(t_i) + \max_{t \in T} \mu(t)] - (u - \max_t \mu_\sigma(t)) \max_{t \in T} \mu_\sigma(t) \right\} \end{aligned} \quad (26)$$

Notice that $\frac{1}{M} \sum_{i:f(t_i) \leq 0} e^{(u - \max \mu_\sigma(t)) f(t_i)} \leq 1$. We continue the above calculations and obtain that

$$\geq \frac{e^{-(u - \max_t \mu_\sigma(t)) \max_{t \in T} \mu_\sigma(t)}}{M} \sum_{i=1}^M \exp \left\{ \frac{u - \max_t \mu_\sigma(t)}{\sigma_T} [\sigma_T f(t_i) + \max_{t \in T} \mu(t)] \right\} - 1.$$

We apply Jensen's inequality to the above summation and obtain that

$$\begin{aligned} & \frac{1}{M} \sum_{i=1}^M e^{u_{t_i} f(t_i)} \\ & \geq e^{-\frac{u - \max_{t \in T} \mu_{\sigma}(t)}{\sigma_T} \max_{t \in T} \mu(t)} \left[\frac{1}{M} \sum_{i=1}^M e^{\sigma_T f(t_i) + \max_{t \in T} \mu(t)} \right]^{(u - \max_{t \in T} \mu_{\sigma}(t))/\sigma_T} - 1 \end{aligned}$$

We apply the result in (25) and continue the above calculations. There exist $\kappa_2, \kappa_3 > 0$ such that on the set $\{\frac{\kappa_1}{Nd} \sum_{i=1}^M e^{\sigma(t_i) f(t_i) + \mu(t_i)} > e^b\}$

$$\begin{aligned} \frac{1}{M} \sum_{i=1}^M e^{u_{t_i} f(t_i)} & \geq e^{-\kappa_2 u} \left(\frac{\delta_0 N^d}{M \kappa_1} e^b \right)^{(u - \max_{t \in T} \mu_{\sigma}(t))/\sigma_T} - 1 \\ & \geq e^{-\kappa_3 u \log b + u^2}. \end{aligned} \quad (27)$$

For the last step, we use the definition of u as in (7) and obtain that $\sigma_T u - \kappa_4 \log b \leq b \leq \sigma_T u$ where $\kappa_4 > 0$ and further

$$\lim_{b \rightarrow \infty} \frac{b}{\sigma_T u} = 1.$$

Combining (22), (23), and (27), there exists a constant $\kappa_5 > 0$ such that

$$e^{(u - \min_{t \in T} \mu_{\sigma}(t))^2} E^{Q_M} \left[\left(\sum_{i=1}^M \frac{1}{M} e^{f(t_i) u_{t_i}} \right)^{-2} ; \frac{\kappa_1}{Nd} \sum_{i=1}^M e^{\sigma(t_i) f(t_i) + \mu(t_i)} > e^b \right] \leq e^{-\sigma_T^{-2} b^2 + \kappa_5 b \log b}.$$

Combining the above results with Theorem 2.3 and Lemma 4.3, we have

$$\liminf_{b \rightarrow \infty} \frac{\log E^{Q_M} \tilde{L}_b^2}{2 \log w_M(b)} \geq \liminf_{b \rightarrow \infty} \frac{-b^2/\sigma_T^2 + \kappa_5 b \log b}{-b^2/\sigma_T^2} = 1.$$

On the other hand, Jensen's inequality suggests that $\log E^{Q_M} \tilde{L}_b^2 \geq 2 \log w_M(b)$ and

$$\limsup_{b \rightarrow \infty} \frac{\log E^{Q_M} \tilde{L}_b^2}{2 \log w_M(b)} \leq 1.$$

■

5. PROOFS OF LEMMAS

In this section, we present the proofs of the supporting lemmas in the main proof. The first lemma is known as the Borel-TIS lemma, which is proved independently by [Borell 1975; Tsirelson et al. 1976].

LEMMA 5.1 (BORELL-TIS). *Let $f(t)$, $t \in \mathcal{U}$ (\mathcal{U} is a parameter set) be a mean zero Gaussian random field. f is almost surely bounded on \mathcal{U} . Then, $E[\sup_{t \in \mathcal{U}} f(t)] < \infty$, and*

$$P\left(\sup_{t \in \mathcal{U}} f(t) - E[\sup_{t \in \mathcal{U}} f(t)] \geq b\right) \leq \exp\left(-\frac{b^2}{2\sigma_{\mathcal{U}}^2}\right),$$

where $\sigma_{\mathcal{U}}^2 = \sup_{t \in \mathcal{U}} \text{Var}[f(t)]$.

The Borel-TIS lemma provides a very general bound of the tail probability

$$P(\sup_{t \in \mathcal{U}} f(t) > b) \leq \exp\left(-\frac{(b - E[\sup_{t \in \mathcal{U}} f(t)])^2}{2\sigma_{\mathcal{U}}^2}\right).$$

In most cases, $E[\sup_{t \in \mathcal{U}} f(t)]$ is much smaller than b . Thus, for b sufficiently large, the tail probability can be further bounded by

$$P(\sup_{t \in \mathcal{U}} f(t) > b) \leq \exp\left(-\frac{b^2}{4\sigma_{\mathcal{U}}^2}\right).$$

The following result by [Dudley 1973] (c.f. Theorem 6.7 in [Adler et al. 2012]) is often used to control $E[\sup_{t \in \mathcal{U}} f(t)]$.

LEMMA 5.2. *Let \mathcal{U} be a compact subset of R^n , and let $\{f(t) : t \in \mathcal{U}\}$ be a mean zero, continuous Gaussian random field. Define the canonical metric d on \mathcal{U} as*

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

and put $\text{diam}(\mathcal{U}) = \sup_{s, t \in \mathcal{U}} d(s, t)$, which is assumed to be finite. Then there exists a finite universal constant $\kappa > 0$ such that

$$E[\max_{t \in \mathcal{U}} f(t)] \leq \kappa \int_0^{\text{diam}(\mathcal{U})/2} [\log(\mathcal{N}(\varepsilon))]^{1/2} d\varepsilon,$$

where the entropy $\mathcal{N}(\varepsilon)$ is the smallest number of d -balls of radius ε whose union covers \mathcal{U} .

By means of the above lemma, one can often establish that $E[\max_{t \in \mathcal{U}} f(t)] = O(\delta |\log \delta|)$ where $\delta^2 = \sup_{t \in \mathcal{U}} \text{Var}(f(t))$. We now proceed to the proof of the supporting lemmas.

PROOF OF LEMMA 4.1. We present the proof of the first bound in the lemma. The proof of the second bound is completely analogous. Consider the change of measure

$$\frac{dQ}{dP} = \int_T \frac{1}{\text{mes}(T)} \exp\left(u_t f(t) - \frac{u_t^2}{2}\right) dt. \quad (28)$$

With a similar argument as in (23), we have

$$\begin{aligned} P(\mathcal{I}(T) > e^b, \mathcal{L}^c) &= E^Q \left[\frac{dP}{dQ}; \mathcal{I}(T) > e^b, \mathcal{L}^c \right] \\ &\leq e^{(u - \min_t \mu_\sigma(t))^2 / 2} \cdot E^Q \left[\frac{\text{mes}(T)}{\int_T e^{u_t f(t)} dt}; \mathcal{I}(T) > e^b, \mathcal{L}^c \right], \end{aligned}$$

where E^Q is the expectation under measure Q . Note that $\mathcal{I}(T) > e^b$ implies that for all large b ,

$$\begin{aligned} \frac{1}{\text{mes}(T)} \int_T e^{\sigma_T f(t) + \max_{t \in T} \mu(t)} dt &\geq \frac{1}{\text{mes}(T)} \int_{T \cap \{f(t) \geq 0\}} e^{\sigma(t) f(t) + \mu(t)} dt \\ &\geq \frac{1}{\text{mes}(T)} e^b - \frac{1}{\text{mes}(T)} \int_{T \cap \{f(t) < 0\}} e^{\sigma(t) f(t) + \mu(t)} dt \\ &\geq \frac{1}{\text{mes}(T)} e^b - e^{\max_{t \in T} \mu(t)}. \end{aligned} \quad (29)$$

Furthermore, on the set $\{\mathcal{I}(T) > e^b\}$, we have that

$$\frac{1}{\text{mes}(T)} \int_T e^{u_t f(t)} dt \geq \frac{1}{\text{mes}(T)} \int_{T \cap \{f(t) > 0\}} \exp\left\{(u - \max_t \mu_\sigma(t)) f(t)\right\} dt.$$

We add and subtract the term $\frac{u - \max_t \mu_\sigma(t)}{\sigma_T} \max_{t \in T} \mu(t)$ in the exponent and continue the above calculation

$$= \frac{1}{\text{mes}(T)} \int_{T \cap \{f(t) > 0\}} \exp\left\{\frac{u - \max_t \mu_\sigma(t)}{\sigma_T} \left(\sigma_T f(t) + \max_{t \in T} \mu(t)\right) - \frac{u - \max_t \mu_\sigma(t)}{\sigma_T} \max_{t \in T} \mu(t)\right\} dt$$

Since $\frac{1}{\text{mes}(T)} \int_{T \cap \{f(t) \leq 0\}} e^{(u - \max_t \mu_\sigma(t)) f(t)} dt \leq 1$, the above display is bounded from below by

$$\geq \frac{1}{\text{mes}(T)} \int_T \exp\left\{\frac{u - \max_t \mu_\sigma(t)}{\sigma_T} \left(\sigma_T f(t) + \max_{t \in T} \mu(t)\right) - \frac{u - \max_t \mu_\sigma(t)}{\sigma_T} \max_{t \in T} \mu(t)\right\} dt - 1$$

By means of Jensen's inequality and the lower bound in (29), the above display is further lower bounded by

$$\begin{aligned} &\geq \exp \left\{ -\frac{u - \max_t \mu_\sigma(t)}{\sigma_T} \max_{t \in T} \mu(t) \right\} \left[\frac{1}{\text{mes}(T)} \int_T \exp\{\sigma_T f(t) + \max_{t \in T} \mu(t)\} dt \right]^{(u - \max_t \mu_\sigma(t))/\sigma_T} - 1 \\ &\geq \exp \left\{ -\frac{u - \max_t \mu_\sigma(t)}{\sigma_T} \max_{t \in T} \mu(t) \right\} \left[\frac{e^b}{\text{mes}(T)} - e^{\max_{t \in T} \mu(t)} \right]^{(u - \max_t \mu_\sigma(t))/\sigma_T} - 1. \end{aligned} \quad (30)$$

Therefore, we have the following bound

$$\begin{aligned} &P(\mathcal{I}(T) > e^b, \mathcal{L}^c) \\ &\leq e^{(u - \min_t \mu_\sigma(t))^2/2} \cdot E^Q \left[\frac{1}{\text{mes}(T)^{-1} \int_T e^{u_t f(t)} dt}; \mathcal{I}(T) > e^b, \mathcal{L}^c \right] \\ &\leq e^{(u - \min_t \mu_\sigma(t))^2/2} \left[e^{-\frac{u - \max_{t \in T} \mu_\sigma(t)}{\sigma_T} \max_{t \in T} \mu(t)} \left(\frac{e^b}{\text{mes}(T)} - e^{\max_{t \in T} \mu(t)} \right)^{\frac{u - \max_{t \in T} \mu_\sigma(t)}{\sigma_T}} - 1 \right]^{-1} \\ &\quad \times Q(\mathcal{I}(T) > e^b, \mathcal{L}^c) \end{aligned}$$

Notice the facts that $(u - \min_t \mu_\sigma(t))^2/2 = u^2/2 + O(u)$ and

$$\left[\frac{e^b}{\text{mes}(T)} - e^{\max_{t \in T} \mu(t)} \right]^{\frac{u - \max_{t \in T} \mu_\sigma(t)}{\sigma_T}} = e^{(b/\sigma_T + O(1))(u + O(1))}.$$

By the facts that $u = b/\sigma_T + O(\log b)$ and $Q(\mathcal{I}(T) > e^b, \mathcal{L}^c) \leq Q(\mathcal{L}^c)$, there exists a constant c_1 such that for b sufficiently large,

$$P(\mathcal{I}(T) > e^b, \mathcal{L}^c) \leq \exp \left\{ -\frac{b^2}{2\sigma_T^2} + c_1 b \log b \right\} Q(\mathcal{L}^c).$$

Then Lemma 5.3 (presented momentarily) together with Lemma 4.3 implies that there exist constants κ_1 and λ such that

$$P(\mathcal{I}(T) > e^b, \mathcal{L}^c) \leq \eta \exp \left(-\frac{b^2}{2\sigma_T^2} + c_1 b \log b - \lambda b^{1+\epsilon} \right) \leq \kappa_1 \eta w(b).$$

Note that κ_1 can be chosen such that the above inequality holds for all b and η . Then, we can redefine η and the constant κ_0 and obtain the conclusion.

The proof of the second inequality in (18) is similar. The only difference is that the integral in the above derivation is replaced by a summation over discretization T_N . Hence, we omit the details. ■

LEMMA 5.3. *Under the conditions in Theorem 2.3, there exists a constant λ such that*

$$Q(\mathcal{L}^c) \leq \eta \exp(-\lambda b^{1+\epsilon}),$$

where \mathcal{L} is the set defined as in (17).

PROOF OF LEMMA 5.3. We focus on a given τ and define another process

$$f_*(t) = f(t) - u_\tau C(t, \tau).$$

Then under the measure of Q , $f_*(t)$ has the same distribution as that of $f(t)$ under the original measure P . Under the above notation and $N = \kappa_0^{1/\beta} |\log \eta|^{1/\beta} \eta^{-1/\beta} b^{(2+2\epsilon)/\beta}$, we have that

$$\mathcal{L}^c = \left\{ \sup_{s,t:|t-s| \leq dN^{-1}} \left| \sigma(s)f_*(s) + \sigma(s)u_\tau C(s, \tau) + \mu(s) - (\sigma(t)f_*(t) + \sigma(t)u_\tau C(t, \tau) + \mu(t)) \right| > \eta \kappa_0^{-1/3} b^{-(1+\epsilon)} \right\}.$$

According to the Borel-TIS lemma,

$$Q\left(\sup_{t \in T} |f_*(t)| > \sqrt{|\log \eta| \kappa_0^{1/2} b^{1/2+\epsilon}}\right) \leq 2 \exp\left(-\frac{\left(\sqrt{|\log \eta| \kappa_0^{1/2} b^{1/2+\epsilon}} - E^Q[\sup_{t \in T} f_*(t)]\right)^2}{2\sigma_T^2}\right).$$

Since $E^Q[\sup_{t \in T} f_*(t)] = O(1)$, there exists $\lambda_1 > 0$ such that for large κ_0

$$Q\left(\sup_{t \in T} |f_*(t)| > \sqrt{|\log \eta| \kappa_0^{1/2} b^{1/2+\epsilon}}\right) \leq \exp(-\lambda_1 |\log \eta| \kappa_0 b^{1+2\epsilon}) = \eta^{\lambda_1 \kappa_0 b^{1+2\epsilon}} \leq \eta \exp(-\lambda_1 \kappa_0 b^{1+2\epsilon}). \quad (31)$$

In what follows, we bound the tail of $\mathcal{L}^c \cap \{\sup_{t \in T} |f_*(t)| \leq \sqrt{|\log \eta| \kappa_0^{1/2} b^{1/2+\epsilon}}\}$. For all s, t satisfying $|t - s|^\beta \leq d^\beta N^{-\beta} = d^\beta \kappa_0^{-1} |\log \eta|^{-1} \eta b^{-(2+2\epsilon)}$ and $b > 1$, there exist positive constants c_1, c_2 , and c such that

$$\begin{aligned} & \left| \sigma(s)f_*(s) + \sigma(s)u_\tau C(s, \tau) + \mu(s) - (\sigma(t)f_*(t) + \sigma(t)u_\tau C(t, \tau) + \mu(t)) \right| \\ & \leq |\sigma(s)| \cdot |f_*(s) - f_*(t)| + |f_*(t)| \cdot |\sigma(s) - \sigma(t)| \\ & \quad + u_\tau \cdot |\sigma(s)| \cdot |C(s, \tau) - C(t, \tau)| + u_\tau \cdot |C(t, \tau)| \cdot |\sigma(s) - \sigma(t)| + |\mu(s) - \mu(t)| \\ & \leq \sigma_T \cdot |f_*(s) - f_*(t)| + c\kappa_0^{-1/2} |\log \eta|^{-1/2} \eta b^{-1-2\epsilon}. \end{aligned}$$

Applying the above bound, we obtain that

$$\begin{aligned} & Q\left(\mathcal{L}^c, \sup_{t \in T} |f_*(t)| \leq \sqrt{|\log \eta|} \kappa_0^{1/2} b^{1/2+\epsilon}\right) \\ & \leq Q\left(\sup_{s,t:|t-s| \leq dN^{-1}} \sigma_T |f_*(s) - f_*(t)| + c\kappa_0^{-1/2} |\log \eta|^{-1/2} \eta b^{-1-2\epsilon} > \eta \kappa_0^{-1/3} b^{-(1+\epsilon)}\right) \\ & \leq Q\left(\sup_{s,t:|t-s| \leq dN^{-1}} \sigma_T |f_*(s) - f_*(t)| > \frac{1}{2} \eta \kappa_0^{-1/3} b^{-(1+\epsilon)}\right). \end{aligned}$$

For the component $f_*(s) - f_*(t)$ and $|t - s| \leq dN^{-1}$, the variance function has an upper bound:

$$\text{Var}(f_*(s) - f_*(t)) = 2(1 - C(s, t)) \leq 2\kappa_H |s - t|^{2\beta} \leq 2\kappa_H d^{2\beta} N^{-2\beta} = 2\kappa_H d^{2\beta} \kappa_0^{-2} |\log \eta|^{-2} \eta^2 b^{-4(1+\epsilon)}.$$

An application of Lemma 5.2 to the double-indexed process $\xi(s, t) = f_*(s) - f_*(t)$ (that is the supremum of the difference $\sup_{|t-s| \leq dN^{-1}} (f_*(s) - f_*(t))$ is the supremum of $\xi(s, t)$ on a set where its variance is small) yields that

$$E^Q \left[\sup_{|t-s| \leq dN^{-1}} (f_*(s) - f_*(t)) \right] = O(\eta b^{-2(1+\epsilon)} \log b).$$

We apply the Borel-TIS inequality

$$\begin{aligned} & Q\left(\sup_{s,t:|t-s| \leq dN^{-1}} \sigma_T |f_*(s) - f_*(t)| > \frac{1}{2} \eta \kappa_0^{-1/3} b^{-(1+\epsilon)}\right) \\ & \leq 2 \exp \left\{ - \frac{\left(\frac{1}{2} \eta \kappa_0^{-1/3} b^{-(1+\epsilon)} - E_Q[\sup_{|t-s| \leq dN^{-1}} (f_*(s) - f_*(t))] \right)^2}{4\sigma_T^2 \kappa_H d^{2\beta} \kappa_0^{-2} |\log \eta|^{-2} \eta^2 b^{-4(1+\epsilon)}} \right\} \\ & \leq \exp\left(-\lambda_2 (\log \eta)^2 \kappa_0^{4/3} b^{2(1+\epsilon)}\right) \end{aligned} \quad (32)$$

for some $\lambda_2 > 0$. Combining the results in (31) and (32), there exists λ such that for large κ_0 ,

$$Q(\mathcal{L}^c) \leq \exp(-\lambda_1 |\log \eta| \kappa_0 b^{1+2\epsilon}) + \exp\left(-\lambda_2 (\log \eta)^2 \kappa_0^{4/3} b^{2(1+\epsilon)}\right) \leq \eta \exp(-\lambda b^{1+\epsilon}).$$

■

PROOF OF LEMMA 4.2. Let $\Phi(\cdot)$ be the cumulative distribution function of the standard Gaussian distribution. Define cumulative distribution function of $\mathcal{I}(T)$ and the associated quantiles with respect to $\Phi(\cdot)$ as

$$F(x) = P(\log \mathcal{I}(T) \leq x) \text{ and } t_x = \Phi^{-1}(F(x)).$$

We cite one result (Proposition 1 in [Liu and Xu 2012b]) that gives an upper bound of $F'(x)$:

LEMMA 5.4. *Under the conditions of Lemma 4.2, $F'(x)$ exists almost everywhere. Choose $y < x$ (depending on x) in a way that $x - y \rightarrow 0$ and $x(x - y) \rightarrow \infty$ when we send x to infinity. Then,*

$$\limsup_{x \rightarrow \infty} \sqrt{2\pi}\sigma_T \exp\left(\frac{\sigma_T^2 t_y^2 + 2(x-y)y}{2\sigma_T^2}\right) F'(x) \leq 1, \quad (33)$$

where $\sigma_T = \sup_{t \in T} \sigma(t)$.

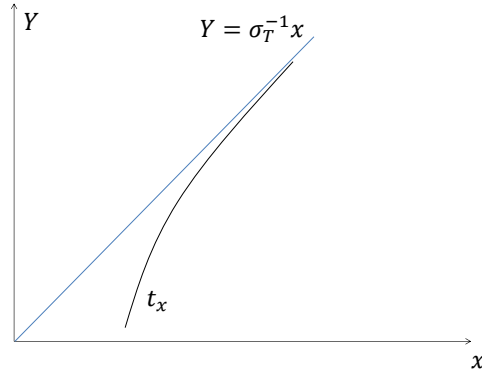


Fig. 2. Plot of t_x

Lemma 5.4 is equivalent to

$$F'(x) \leq \frac{1 + o(1)}{\sqrt{2\pi}\sigma_T} \exp\left(-\frac{\sigma_T^2 t_y^2 + 2(x-y)y}{2\sigma_T^2}\right). \quad (34)$$

To prove Lemma 4.2, it is sufficient to show that the right-hand-side of (34) is $o(1)x^{1+\epsilon/2}w(x+x^{-1-\lambda})$. Lemma 4.3 implies that

$$\lim_{x \rightarrow \infty} \frac{t_x}{x} = \sigma_T^{-1}. \quad (35)$$

Then, for $\lambda > 0$ we let $\tilde{x} = x + x^{-1-\lambda}$ and $y = x - \log \log x/x$. With such choices of \tilde{x} and y , the following holds

$$\begin{aligned} t_{\tilde{x}}^2 - (t_y^2 + 2\sigma_T^{-2}(x-y)y) &= (2 + o(1))\sigma_T^{-1}x(t_{\tilde{x}} - t_y) - 2\sigma_T^{-2}(x-y)y \\ &= (2 + o(1))\sigma_T^{-1}x(x-y) \left(\frac{t_{\tilde{x}} - t_y}{\tilde{x} - y} \cdot \frac{\tilde{x} - y}{x - y} - \sigma_T^{-1}(1 + o(1)) \right). \end{aligned} \quad (36)$$

According to Theorem 4.4.1 in [Bogachev 1998] (see also [Ehrhard 1983]), t_x is a concave function of x , and thus $(t_z - t_x)/(z - x)$ is a monotone non-increasing function of z when $z > x$ and $\lim_{z \rightarrow x^+} (t_z - t_x)/(z - x)$ exists. Figure 5 illustrates the function t_x . Then for $y = x - \log \log x/x$ and $\tilde{x} = x + x^{-1-\lambda}$, we have that

$$(t_{\tilde{x}} - t_y)/(\tilde{x} - y) \leq \lim_{z \rightarrow y^+} (t_z - t_y)/(z - y). \quad (37)$$

The concavity of t_x and (35) imply that $\lim_{z \rightarrow y^+} (t_z - t_y)/(z - y)$ is a non-increasing function of y and converges to σ_T^{-1} . Sending x to infinity on both sides of (37) gives $\limsup_{x \rightarrow \infty} (t_{\tilde{x}} - t_y)/(\tilde{x} - y) \leq \sigma_T^{-1}$. Thus, the first term in the parenthesis of (36) is bounded by $\sigma_T^{-1} + o(1)$. Therefore, there exists a constant $c_0 > 0$ such that

$$t_{\tilde{x}}^2 - (t_y^2 + 2\sigma_T^{-2}(x-y)y) \leq c_0x(x-y) = c_0 \log \log x.$$

Then, we obtain an upper bound for (34):

$$F'(x) \leq \frac{1}{\sqrt{2\pi}\sigma_T} \exp\left(-\frac{\sigma_T^2 t_y^2 + 2(x-y)y}{2\sigma_T^2}\right) \leq \frac{1}{\sqrt{2\pi}\sigma_T} \exp\left(-\frac{t_{\tilde{x}}^2 - c_0 \log \log x}{2}\right).$$

Note that

$$w(\tilde{x}) = 1 - F(\tilde{x}) = \frac{1 + o(1)}{\sqrt{2\pi}t_{\tilde{x}}} \exp\left(-\frac{t_{\tilde{x}}^2}{2}\right) = (1 + o(1)) \frac{\sigma_T}{\sqrt{2\pi}x} \exp\left(-\frac{t_{\tilde{x}}^2}{2}\right).$$

Therefore, for any $\epsilon > 0$, we have that

$$F'(x) \leq (1 + o(1))\sigma_T^{-2} \cdot x(\log x)^{c_0/2} \cdot w(\tilde{x}) = o(1)x^{1+\epsilon/2}w(x + x^{-1-\lambda}).$$

■

PROOF OF LEMMA 4.3. We start to derive a lower bound for $w(b)$. Let $\sigma(t_*) = \sup_{t \in T} \sigma(t) = \sigma_T$. On the set \mathcal{L} as defined in (17) (with $\eta = \eta_0$ fixed), there exists a $\delta_0 > 0$ such that

$$\int_{|t-t_*| < N^{-1}} e^{\sigma(t)f(t)+\mu(t)} dt \geq \delta_0 N^{-d} e^{\sigma_T f(t_*)}.$$

Then, on the set \mathcal{L} if $\sigma_T f(t_*) > b + d \log N - \log \delta_0$ then

$$\int_{|t-t_*| < N^{-1}} e^{\sigma(t)f(t)+\mu(t)} dt \geq e^b.$$

Therefore, we have the lower bound

$$\begin{aligned} w(b) &\geq P(\mathcal{I}(T) > e^b, \mathcal{L}) \\ &\geq P\left(\int_{|t-t_*| < N^{-1}} e^{\sigma(t)f(t)+\mu(t)} dt > e^b, \mathcal{L}\right) \\ &\geq P(\sigma_T f(t_*) > b + d \log N - \log \delta_0, \mathcal{L}) \\ &\geq P(\sigma_T f(t_*) > b + d \log N - \log \delta_0) \times P(\mathcal{L} | \sigma_T f(t_*) > b + d \log N - \log \delta_0) \end{aligned} \quad (38)$$

There exist $\delta_0, \delta_1 > 0$ such that the conditional probability $P(\mathcal{L} | \sigma_T f(t_*) > b + d \log N - \log \delta_0) > \delta_1$. Therefore,

$$w(b) \geq \delta_1 P(\sigma_T f(t_*) > b + d \log N - \log \delta_0). \quad (39)$$

Note that $N = O(1)b^{(1+\epsilon)/\beta}$ and $f(t_*)$ follows a standard normal distribution. Thus, there exist positive constants \tilde{c}_0 and \tilde{c}_1 such that $w(b) \geq \exp\left(-\frac{b^2 + \tilde{c}_1 b \log b + \tilde{c}_0}{2\sigma_T^2}\right)$.

We continue to construct an upper bound. Since $\mathcal{I}(T) > e^b$ implies that $\sup_{t \in T} \{\sigma(t)f(t)\} > b - \max_{t \in T} \mu(t) - \log \text{mes}(T)$. Therefore, there exists a constant \tilde{c}_2 such that

$$w(b) \leq P\left(\sup_T \{\sigma(t)f(t)\} > b - \max_{t \in T} \mu(t) - \log \text{mes}(T)\right) \leq \exp\left(-\frac{b^2}{2\sigma_T^2} + \tilde{c}_2 b\right). \quad (40)$$

Combining (39) and (40), we conclude the proof. ■

REFERENCES

- ADLER, R. 1981. *The Geometry of Random Fields*. Wiley, Chichester, U.K.; New York, U.S.A.
- ADLER, R., BLANCHET, J., AND LIU, J. 2012. Efficient Monte Carlo for large excursions of Gaussian random fields. *Annals of Applied Probability* 22, 3, 1167–1214.

- ADLER, R., SAMORODNITSKY, G., AND TAYLOR, J. 2009. High level excursion set geometry for non-Gaussian infinitely divisible random fields. *preprint*.
- ADLER, R. AND TAYLOR, J. 2007. *Random fields and geometry*. Springer.
- AHSAN, S. 1978. Portfolio selection in a lognormal securities market. *Zeitschrift Fur Nationalokonomie-Journal of Economics* 38, 1-2, 105–118.
- ASMUSSEN, S. AND GLYNN, P. 2007. *Stochastic Simulation: Algorithms and Analysis*. Springer, New York, NY, USA.
- ASMUSSEN, S. AND KROESE, D. 2006. Improved algorithms for rare event simulation with heavy tails. *Advances in Applied Probability* 38, 545–558.
- AZAIS, J. M. AND WSCHEBOR, M. 2005. On the distribution of the maximum of a Gaussian field with d parameters. *Annals of Applied Probability* 15, 1A, 254–278.
- AZAIS, 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. *Stochastic Processes and Their Applications* 118, 7, 1190–1218.
- AZAIS, J. M. AND WSCHEBOR, M. 2009. *Level sets and extrema of random processes and fields*. Wiley, Hoboken, N.J.
- BASAK, S. AND SHAPIRO, A. 2001. Value-at-risk-based risk management: Optimal policies and asset prices. *Review of Financial Studies* 14, 2, 371–405.
- BERMAN, S. M. 1985. An asymptotic formula for the distribution of the maximum of a Gaussian process with stationary increments. *Journal of Applied Probability* 22, 2, 454–460.
- BLACK, F. AND SCHOLES, M. 1973. Pricing of options and corporate liabilities. *Journal of Political Economy* 81, 3, 637–654.
- BLANCHET, J. AND GLYNN, P. 2008. Efficient rare event simulation for the maximum of heavy-tailed random walks. *to appear in Annals of Applied Probability* 18, 4, 1351–1378.
- BLANCHET, J., GLYNN, P., AND LEDER, K. 2008. Strongly efficient algorithms for light-tailed random walks: An old folk song sung to a faster new tune. In *MCQMC 2008*, P. LECuyer and A. Owen, Eds. Springer, 227–248.
- BLANCHET, J. H. AND LIU, J. 2008. State-dependent importance sampling for regularly varying random walks. *Advances in Applied Probability* 40, 1104–1128.
- BOGACHEV, V. 1998. *Gaussian Measures*. Mathematical Surveys and Monographs Volume 62. American Mathematical Society.
- BORELL, C. 1975. The Brunn-Minkowski inequality in Gauss space. *Inventiones Mathematicae* 30, 2, 207–216.
- DEUTSCH, H. P. 2004. *Derivatives and internal models* 3rd Ed. Finance and capital markets. Palgrave Macmillan, Houndmills, Basingstoke, Hampshire ; New York, N.Y.
- DUDLEY, R. 1973. Sample functions of the Gaussian process. *Ann. Probab.* 1, 66–103.
- DUFFIE, D. AND PAN, J. 1997. An overview of value at risk. *The Journal of Derivatives* 4, 3, 7–49.
- DUFRESNE, D. 2001. The integral of geometric Brownian motion. *Advances in Applied Probability* 33, 1, 223–241.
- DUPUIS, P., LEDER, K., AND WANG, H. 2007. Importance sampling for sums of random variables with regularly varying tails. *ACM Trans. Model. Comput. Simul.* 17, 3, 14.
- EHRHARD, A. 1983. Symétrisation dans l'espace de Gauss. *Math. Scand.* 53, 281–301.
- FOSS, S. AND RICHARDS, A. 2010. On sums of conditionally independent subexponential random variables. *Mathematics of Operations Research* 35, 102–119.

- GLASSERMAN, P., HEIDELBERGER, P., AND SHAHABUDDIN, P. 2000. Variance reduction techniques for estimating value-at-risk. *Management Science* 46, 10, 1349–1364.
- HANDCOCK, M., RAFTERY, A., AND TANTRUM, J. 2007. Model-based clustering for social network. *Journal of the Royal Statistical Society: Series A* 170, 1–22.
- HOFF, P. D., RAFTERY, A. E., AND HANDCOCK, M. S. 2002. Latent space approaches to social network analysis. *Journal of the American Statistical Association* 97, 460, 1090–1098.
- HÜSLER, J. 1990. Extreme values and high boundary crossings of locally stationary Gaussian processes. *Annals of Probability* 18, 1141–1158.
- HÜSLER, J., PITERBARG, V., AND ZHANG, Y. 2011. Extremes of Gaussian processes with random variance. *Electronic Journal of Probability* 16, 45, 1254–1280.
- LANDAU, H. J. AND SHEPP, L. A. 1970. Supremum of a Gaussian process. *Sankhya-the Indian Journal of Statistics Series A* 32, Dec, 369–378.
- L'ECUYER, P. 2009. Quasi-monte carlo methods with applications in finance. *Finance and Stochastics* 12, 3, 307–349.
- L'ECUYER, P. AND MUNGER, D. 2010. On figures of merit for randomly-shifted lattice rules. In *Monte Carlo and Quasi Monte Carlo Methods*, L. Plaskota and H. Wozniakowski, Eds. Springer-Verlag, Berlin, 133–159.
- L'ECUYER, P., MUNGER, D., AND TUFFIN, B. 2010. On the distribution of integration error by randomly-shifted lattice rules. *Electronic Journal of Statistics* 4, 950–993.
- LEDOUX, M. AND TALAGRAND, M. 1991. Probability in Banach spaces : isoperimetry and processes.
- LIU, J. 2012. Tail approximations of integrals of Gaussian random fields. *Annals of Probability* 40, 3, 1069–1104.
- LIU, J. AND XU, G. 2012a. On the conditional distributions and the efficient simulations of exponential integrals of Gaussian random fields. *preprint <http://arxiv.org/abs/1204.5546>*.
- LIU, J. AND XU, G. 2012b. On the density functions of integrals of Gaussian random fields. *Advances in Applied Probability to appear. Available at <http://stat.columbia.edu/~jliu/paper/grf10.pdf>*.
- LIU, J. AND XU, G. 2012c. Some asymptotic results of Gaussian random fields with varying mean functions and the associated processes. *Annals of Statistics* 40, 262–293.
- MARCUS, M. B. AND SHEPP, L. A. 1970. Continuity of Gaussian processes. *Transactions of the American Mathematical Society* 151, 2.
- MERTON, R. 1973. The theory of rational option pricing. *Bell Journal of Economics and Management Science* 4, 141–183.
- PITERBARG, V. I. 1996. *Asymptotic methods in the theory of Gaussian processes and fields*. American Mathematical Society, Providence, R.I.
- SNIJDERS, T. 2002. Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structures* 3.
- SNIJDERS, T. 2011. Efficient simulation of tails probabilities of sums of correlated lognormals. *Annals of Operations Research* 189, 5–23.
- SUDAKOV, V. AND TSIRELSON, B. 1974. Extremal properties of half spaces for spherically invariant measures. *Zap. Nauchn. Sem. LOMI* 45, 75–82.
- SUN, J. Y. 1993. Tail probabilities of the maxima of Gaussian random-fields. *Annals of Probability* 21, 1, 34–71.

- TALAGRAND, M. 1996. Majorizing measures: The generic chaining. *Annals of Probability* 24, 3, 1049–1103.
- TAYLOR, J., TAKEMURA, A., AND ADLER, R. J. 2005. Validity of the expected Euler characteristic heuristic. *Annals of Probability* 33, 4, 1362–1396.
- TSIRELSON, B., IBRAGIMOV, I., AND SUDAKOV, V. 1976. Norms of Gaussian sample functions. *Proceedings of the Third Japan-USSR Symposium on Probability Theory (Tashkent, 1975)* 550, 20–41.
- XING, E. P., FU, W. J., AND SONG, L. 2010. A state-space mixed membership blockmodel for dynamic network tomography. *Annals of Applied Statistics* 4, 2, 535–566.
- YOR, M. 1992. On some exponential functionals of Brownian-motion. *Advances in Applied Probability* 24, 3, 509–531.