A REVIEW OF SOME METHODS TO ESTIMATE THE TAIL OF THE DISTRIBUTION OF THE MAXIMUM OF A GAUSSIAN FIELD

MARIO WSCHEBOR

Abstract. We review two methods aiming to compute the distribution of the maximum of a Gaussian process: an updated version of the Rice series and the so-called "direct method". The Rice series applies to one-parameter processes while the direct method can be used also for multi-parameter random fields.

1. Introduction

Let $\mathcal{X} = \{X(t) : t \in T\}$ be a real-valued Gaussian random function and $M = \sup_{t\in T} X(t)$ its supremum. Our aim in this paper is to review without proofs some of the results on the computation of the probability distribution of the random variable M, that is, the function

$$F(u) = P(M \leq u),\ u \in R.$$

In all cases, we will assume that the parameter domain T is a subset of Euclidean space possessing some geometric regularity and the paths $t \rightsquigarrow X(t)$ are continuously differentiable of some order. If T is one-dimensional we will call $\mathcal{X}$ a "random process", if it has dimension greater than 1, we will call it a "random field".

Exact useful formulas for the function F exist only for a short list of Gaussian processes. They are based upon ad-hoc methods and the general question of computing or getting good estimations of this distribution remains, generally speaking, an open problem, interesting both from the point of view of the mathematical theory and the applications.

In the first half of the 1970's, a number of fundamental inequalities on the tail of the distribution of M, that is $1 - F(u)$, as well as asymptotic results as $u \to +\infty$ were proved, starting with a paper by Landau and Shepp [19] followed by a long list of papers and books. A partial account of this work are: Marcus and Shepp [23], Sudakov and Tsirelson [29], Borell [14] [15], Fernique [18], Ledoux and Talagrand [21], Berman [12] [13], Adler [2], Talagrand [31], Ledoux [20] and Lifshits [22].

A typical inequality, proved independently by Borell and Sudakov-Tsirelson, states that if T is countable, the Gaussian process $\mathcal{X}$ is centered and almost surely

Received 10-5-27.

2000 *Mathematics Subject Classification.* Primary 60G70 Secondary 60G15.

Key words and phrases. Gaussian processes, Gaussian fields, Rice formula, Rice series, distribution of the maximum, density of the maximum, direct method.

its paths are bounded, then

$$\sigma^2 = \sup_{t\in T} Var(X(t)) < +\infty$$

and for every $u > 0$,

$$P\big(|M - \mu(M)| > u\big) \le \exp\big(-\frac{u^2}{2\sigma^2}\big),$$

where $\mu(M)$ is a median of the distribution of M. Of course, this extends immediately to separable bounded Gaussian processes.

These classical inequalities are essential for the development of most of the mathematical theory. However, in a wide number of applications they are not good enough since there appear certain constants (such as the median above) for which the estimations one is able to obtain differ substantially from the true values and as a consequence, the relative error that one gets for the tail probability of M becomes exponentially large (for large values of u).

The bounds given by these results can be applied to a wide class of Gaussian processes, but become inaccurate for given ones or for special classes of Gaussian processes. This situation implies that for statistical applications the usual practice is to estimate the distribution of M by simulation of the paths, which is expensive for continuous parameter processes and poor from theoretical viewpoint if one is willing to understand a number of natural problems, such as changing the values of certain parameters in the probability law of the random process or field.

Since the 1990's several methods have been introduced to obtain more precise results under certain restrictions on the process $\mathcal{X}$. These are interesting from the point of view of the mathematical theory as well as in applications. Some examples of these contributions are the double sum method (see Piterbarg [27]); the Euler-Poincaré Characteristic (EPC) approximation, Taylor, Takemura and Adler [33], Adler and Taylor [3]; the tube method, Sun [30], the Rice method, revisited by Azaïs and Delmas [5], the Rice series, Azaïs and Wschebor [6] and the "direct method" [9],[10].

In this paper, we describe the results that one can obtain using the Rice series and the direct method and the main theorems in which they are based.

The Rice series concerns only one-parameter processes. The main point is to write the function F as the sum of a series whose terms are computed from the moments of up-crossings of the process. Its application to real cases requires to consider numerical problems, some of which appear to be very difficult and remain unsolved for the time being. For certain classes of processes, it is numerically efficient to compute the function F using the Rice series, we mention some examples.

With respect to the direct method, one would like to write, whenever it is possible

$$P(M > u) = \ A(u)\ \exp\big(-\frac{1}{2}\frac{u^2}{\sigma^2}\big)\ +\ B(u) \tag{1.1}$$

where $A(u)$ is a known function having polynomially bounded growth as $u \to +\infty$ and $B(u)$ is an error bounded by a centered Gaussian density with variance σ_1^2, $\sigma_1^2 < \sigma^2$. We will call the first (respectively the second) term in the right-hand side of (1.1) the "first (resp second) order approximation of $P(M > u)$."

Generally speaking, both $A(u)$ and the exponential order of $B(u)$ remain unknown, excepting very special cases. The only situation in which this author knows a precise result is the following:

Let $\mathcal{X}$ be a centered stationary Gaussian processes defined on the interval $[0,T]$, $\lambda_0 = 1,\ \lambda_8 < +\infty$, the joint distribution of $X(s), X(t), X'(s), X'(t), X''(s), X''(t)$ is non-degenerate, plus some additional condition on the covariance, such as being decreasing for $0 < t < T$. Under these conditions, one has as $u \to +\infty$:

$$P(M > u) = 1 - \Phi(u) + \sqrt{\frac{\lambda_2}{2\pi}} T\varphi(u) \\ - \frac{3\sqrt{3}(\lambda_4 - \lambda_2^2)^{\frac{9}{2}}}{2\pi\lambda_2^{\frac{9}{2}}(\lambda_2\lambda_6 - \lambda_4^2)} \frac{T}{u^5} \varphi\left(\sqrt{\frac{\lambda_4}{\lambda_4 - \lambda_2^2}} u\right) [1 + o(1)] .$$

Here and in what follows, Φ (resp. φ) denotes the standard normal distribution (resp. density), and for a stationary process, λ_k its k-th spectral moment. This equivalence was proved in [26] for sufficiently small length T and in [4] for any T.

In [3], [33] first order is computed by means of the expectation of the EPC of the excursion set $E_u := \{t \in T : X(t) > u\}$. This works for large values of u, and in certain cases, these authors have also given bounds on the exponential order of $B(u)$ as $u \to +\infty$. We will describe some related results using the direct method, which also allows to estimate the tail probabilities for any fixed $u-$value.

2. The Rice Series

We use the following notations:

- In this section, $\mathcal{X} = \{X(t) : t \in [0,T]\}$ denotes a real-valued random process defined on the interval [0,T] of the real line having $\mathcal{C}^1$ paths.
- If ξ is a random variable taking values in R^d, we denote $p_\xi(z)$ the value at the point $z \in R^d$ of the density of its probability distribution, whenever it exists.
- $U_u = U_u(\mathcal{X}, [0,T])$ is the number of up-crossings of the level u by the function $X(.)$ on the interval $[0,T]$, that is:
  $U_u = \sharp\{t \in [0,T] : X(t) = u, X'(t) > 0\}$.
- For integers $k, m,\ k \geq 0, m \geq 1$ we put $k^{[m]} = k(k-1)...(k-m+1)$.
- $\widetilde{\nu}_m := E(U_u^{[m]} \chi_{\{X(0) \leq u\}})$ $(m = 1, 2, ...)$ denotes the factorial moment of order m of the number of up-crossings of the path over interval $[0,T]$ restricted to the paths starting below the level u at $t = 0$.
- $\nu_m := E(U_u^{[m]})$ $(m = 1, 2, ...)$ denotes the factorial moment of order m of the total number of up-crossings of the path over interval $[0,T]$. Note that $\widetilde{\nu}_m \leq \nu_m$.

The Rice formulas to compute ν_m and $\widetilde{\nu}_m$, whenever they hold true are the following:

$$\nu_m = \int_{[0,T]^m} E\left(X'^+(t_1)...X'^+(t_m) \middle| X(t_1) = ... = X(t_m) = u\right) \\ \times p_{X(t_1),...,X(t_m)}(u, ..., u) dt_1...dt_m. \tag{2.1}$$

$$\begin{aligned}\widetilde{\nu}_m = &\int_{[0,T]^m} dt_1 ... dt_m \\ &\int_{-\infty}^{u} E\left(X'^{+}(t_1)...X'^{+}(t_m)\big|X(0)=x, X(t_1)=...=X(t_m)=u\right) \\ &\times p_{X(0),X(t_1),...,X(t_m)}(x,u,...,u)dx\end{aligned} \tag{2.2}$$

For conditions which assure the validity of these formulas, either for Gaussian processes or non-Gaussian ones, as well as for proofs, one can see [11], Ch. 3. Notice that these formulas are valid for each value of u, so that one has to make an appropriate choice of the versions, both for the conditional expectation and the density. For Gaussian processes, the density is the usual normal density function and the conditional expectation is the one obtained using the regression formulas.

The basic result that relates the distribution of the maximum with the factorial moments of the number of up-crossings, via the Rice series, is the following:

Theorem 2.1. *Assume that a.s. the paths of the stochastic process $\mathcal{X}$ are of class $\mathcal{C}^\infty$ and that the density $p_{X_{T/2}}(.)$ is bounded by some constant D.*
(i) If there exists a sequence of positive numbers $\{c_k\}_{k=1,2,...}$ such that:

$$\gamma_k := P\left(\|X^{(2k-1)}\|_\infty \geq c_k.T^{-(2k-1)}\right) + \frac{Dc_k}{2^{2k-1}\,(2k-1)!} = o\left(2^{-k}\right)\,(k\to\infty) \tag{2.3}$$

then

$$1 - F(u) = P\left(X(0) > u\right) + \sum_{m=1}^{\infty} (-1)^{m+1}\,\frac{\widetilde{\nu}_m}{m!} \tag{2.4}$$

(ii) In formula (2.4) the error when one replaces the infinite sum by its m_0-th partial sum is bounded by $\gamma^*_{m_0+1}$ where:

$$\gamma^*_m := \sup_{k\geq m}\left(2^{k+1}\gamma_k\right).$$

The series in the right-hand side of (2.4) is called the "Rice Series". We make several remarks on Theorem 2.1 below:

(1) The proof of this theorem can be found in [6].
(2) At first glance, the statement of Theorem 2.1 does not appear to be very useful in practice, for various reasons, the first one being that one has to check the existence of a sequence $\{c_k\}_{k=1,2,...}$ such that the hypothesis (2.3) holds true.

In fact, this can be done for some general classes of random processes, that is, one can reduce the existence of the sequence $\{c_k\}$ to some natural conditions on the law of the process. This is the case of Theorem 2.2 below, in which this is done for stationary Gaussian processes having a sufficiently regular covariance function.

Other interesting cases can also be considered with respect to formula (2.4), even if some related numerical problems remain still unsolved. Two relevant examples are the following:

- Gaussian processes possessing continuous but not $\mathcal{C}^\infty$ paths. Then we can proceed as follows: 1) regularize the paths by means of some linear deterministic device (such as convolution with an appropriate

kernel); 2) apply the Rice series to compute the distribution of the new process obtained in this way and 3) estimate the error that the regularization has produced. This appears to be a promising method to compute the distribution of M in many interesting Gaussian cases. (see [11], chap. 4 for an example of this kind of procedure).

- Consider in dimension 1 a strong solution of a stochastic differential equation with regular coefficients excited with Brownian motion. Under fairly general conditions one can regularize the solution and prove the existence of the sequence $\{c_k\}$ for the regularized process, hence use the Rice series to compute the distribution of its maximum (see [11], chap. 5). The drawback is that even when the Rice formulas hold true, in this case one is unable to compute numerically the factorial moments of up-crossings. This remains an open problem: one has to find numerical methods adapted to this kind of computation in non-Gaussian cases, which do not exist for the time being.

(3) Using formula (2.4) to compute the distribution of M in a concrete case implies first that we have to replace the infinite sum by a finite one, that is, to choose the number m_0 of terms to sum up and second, to compute the factorial moments $\widetilde{\nu}_m, 1 \leq m \leq m_0$.

With respect to the choice of m_0 the Rice series has the nice enveloping property that if one replaces the infinite sum by its m_0-partial sum, then the error is bounded by the absolute value of the m_0-th term i.e. $\widetilde{\nu}_{m_0}/m_0!$. Moreover, if m_0 is odd (respectively even), replacing the infinite sum by the m_0- partial sum in (2.4) gives an upper (resp. lower) bound for the tail $1-F(u)$ (see again [6] for a proof; notice that the Rice series has alternate signs but is not a Leibnitz series).

Since one is going to compute the terms up to the m_0-th one, this gives a step by step bound for this source of error, allowing an adaptive procedure not to surpass a given bound.

It turns out that in interesting cases, this adaptive procedure ends up after computing a few terms of the Rice series, implying that the cost of the computation becomes much smaller than simulation of the paths or using m_0 from a priori estimation of the error. We will return to this point below.

The second point refers to the computation of the factorial moments. As mentioned above, in the non-Gaussian case, even when Rice formulas hold true, the computation of the integrands becomes intractable.

For Gaussian processes, this can be handled in a variety of situations, even though there remain a number of unsolved numerical analysis problems. In any case, for $m \geq 2$, one must use some numerical approximation procedure, whose complexity increases rapidly with m, so that an essential practical question is that the number of terms needed not to surpass a given error bound, remains small.

By the way, for Gaussian stationary processes possessing some path regularity, we have reasonably simple analytic formulas for the moments of up-crossings for $m = 1$ and $m = 2$ (see for example [16]), but notice that

the requirement $X(0) \leq u$ in the formula for $\widetilde{\nu}_m$ implies that the relevant process to compute the integrals is non-stationary even if the starting $\mathcal{X}$ is stationary.

(4) The Rice series as a tool to compute the distribution of the maximum of Gaussian processes has been used since a long time for certain Gaussian stationary processes (see [24], [25]). In fact, to verify the convergence of the series and the validity of the equality in these special cases requires lengthy calculations, and is not easier than the proofs of the general theorems 2.1 and 2.2.

(5) For Gaussian processes, we can relax the $\mathcal{C}^\infty$ requirement for the paths to $\mathcal{C}^p$ differentiability. In this case, we can not write an equality such as (2.4), but one can give upper and lower bounds for the approximation of the distribution of M ([6]).

Let us now turn to the application of Theorem 2.1 to Gaussian stationary processes.

Theorem 2.2. *Let $\mathcal{X} = \{X(t) : t \in R\}$ be Gaussian, centered and stationary, with covariance Γ normalized by $\Gamma(0) = 1$. Assume that Γ has a Taylor expansion at the origin which is absolutely convergent at $t = 2T$. Then the conclusion of Theorem 2.1 holds true, so that the Rice series converges and $F(u)$ can be computed by means of (2.4).*

The proof can be found in [6]. It consists in showing that the sequence $\{c_k\}$ defined by:

$$c_k := (Bk\lambda_{4k-2})^{1/2} \quad \text{if} \quad \frac{\lambda_{4k}}{\lambda_{4k-2}} \leq Bk$$

$$c_k := (\lambda_{4k})^{1/2} \quad \text{if} \quad \frac{\lambda_{4k}}{\lambda_{4k-2}} > Bk.$$

where $B = 4 \log 2$, verifies the condition required in Theorem 2.1.

Notice that any centered Gaussian stationary process *with bounded spectrum* verifies the condition in Theorem 2.2 for any value of the length interval T.

3. Computation of Moments, Numerical Problems, and Examples

We assume now that the process $\mathcal{X}$ is Gaussian. The use of the Rice series implies the computation of the factorial moments $\widetilde{\nu}_m$. This leads to use numerical approximation, and for $m \geq 2$, Montecarlo methods are in principle the most efficient for the calculation of the integrals. An efficient procedure requires understanding the behavior of the integrands

$$\tilde{A}(u; t_1, ..., t_m) =$$
$$\int_{-\infty}^{u} E\left(X'^{+}(t_1)...X'^{+}(t_m) \middle| X(0) = x, X(t_1) = ... = X(t_m) = u\right)$$
$$\times\, p_{X(0),X(t_1),...,X(t_m)}(x, u, ..., u)dx$$

near the diagonal

$$D_m = \{(t_1, ..., t_m) \in [0, T]^m, t_i = t_j \text{ for some pair } i, j,\ i \neq j\},$$

where the density collapses and the conditional expectation tends to zero.

It turns out that the integrand $\tilde{A}(u; t_1, ..., t_m)$ approaches zero near the diagonal, with a speed that grows rapidly with m and it is convenient to use importance sampling when applying Montecarlo to compute the integral, taking into account the vanishing order of the integrand. We list some of the results that can be useful for this purpose (For proofs, see [6] or [11]):

- We start with $m = 2$, case in which we have a satisfactory general result. Suppose that $\mathcal{X}$ has C^5 paths and that for each $t \in [0, T]$ the joint distribution of $X(t), X'(t), X^{(2)}(t), X^{(3)}(t)$ does not degenerate. Then

$$\tilde{A}(u; t_1, t_2) \approx J(t^*)(t_2 - t_1)^4 \text{ as } t_1, t_2 \to t^* \tag{3.1}$$

  where $J(.)$ is a continuous non-zero function depending on u, which can be expressed (in a complicated form) in terms of the mean and covariance functions of the process and its derivatives.
- For $m \geq 3$, we have the following result for centered processes and $u = 0$. A general result describing the behavior of the integrand near the diagonal for non-centered Gaussian processes or $u \neq 0$ does not exist at present.

  Assume that $\mathcal{X}$ is centered, has C^{2m-1} paths and that for each pairwise distinct values of the parameter $t_1, t_2, ..., t_m \in [0, T]$ the joint distribution of $(X(t_h), X'(t_h), ..., X^{(2m-1)}(t_h),\ h = 1, 2, ..., m)$ is non-degenerate. Then as $t_1, t_2, ..., t_m \to t^*$, one has:

$$\tilde{A}(0; t_1, ..., t_m) \approx\ J_m(t^*) \prod_{1 \leq i < j \leq m} (t_j - t_i)^4$$

  where J_m is a continuous non-zero function of t.

One should notice that the last result describes the behavior of $\tilde{A}(0; t_1, ..., t_m)$ when $t_1, ..., t_m$ *all* approach *the same* value t^*. From this and the non-degeneracy hypothesis, one can obtain the order of the integrand when $(t_1, ..., t_m)$ approaches the diagonal D_m in some other form.

Let us now turn to numerical questions and some few examples. Suppose that we want to compute $P(M > u)$ with an error bounded by δ, where $\delta > 0$ is a given positive number.

To proceed by simulation of the paths, we approximate the paths by means of the polygonal corresponding to a uniform partition of the domain in n intervals of length T/n and estimate the error in the distribution when replacing the actual maximum M by the maximum $M^{(n)}$ on the partition points. Then we estimate $P(M^{(n)} > u)$ using standard simulation of stationary Gaussian random variables, choosing the number of replications so that the mean square error remains controlled. The total mean number of elementary operations required to get a mean square error bounded by δ in the estimation of $P(M > u)$ takes the form $(const)\, \delta^{-5/2} \log(1/\delta)$. On the other hand, the enveloping property of the Rice series implies that the actual number of terms required by the application of Theorem 2.1 can be *much* smaller than the one resulting from this or from the a priori bound for γ_m^*.

More precisely, suppose that we have obtained each numerical approximation $\widetilde{\nu}_m^*$ of $\widetilde{\nu}_m$ with a precision η

$$|\widetilde{\nu}_m^* - \widetilde{\nu}_m| \leq \eta,$$

and that we stop when

$$\frac{\widetilde{\nu}_{m_0+1}^*}{(m_0+1)!} \leq \eta. \tag{3.2}$$

Then it follows that

$$\left| \sum_{m=1}^{\infty} (-1)^{m+1} \frac{\widetilde{\nu}_m}{m!} - \sum_{m=1}^{m_0} (-1)^{m+1} \frac{\widetilde{\nu}_m^*}{m!} \right| \leq (e+1)\eta.$$

Putting $\eta = \delta/(e+1)$, we get the desired bound.

Next, we give the results of the evaluation of $P(M_T > u)$ using up to three terms in the Rice series in a certain number of cases. In fact, in the statistical literature, the "Davies bound" [17] has been widely used. It is based upon the obvious inequality

$$P(M > u) \leq P(X(0) > u) + \nu_1$$

Notice that the right-hand side of this inequality is obtained by taking only one term in the Rice series - this already gives an upper bound for $P(M > u)$ - and replacing $\tilde{\nu}_1$ by its upper bound ν_1. For fixed T and high level u the Davies bound is sharp. But when both T and u are fixed, the situation becomes essentially different and using more than one term of the Rice series supplies a remarkable improvement in the computation. Generally speaking, for fixed u, increasing length T requires more terms in the Rice series to keep the error below a given bound.

We consider three stationary Gaussian processes for which Rice formulas can be applied for every $m = 1, 2, ..$ in order to compute $\tilde{\nu}_m$. One can show that Rice formula is valid for Gaussian processes whenever the joint distribution of $X(t_1), ..., X(t_m)$ does not degenerate for $(t_1, ..., t_m) \notin D_m$ (see [11], Chap. 3 for a proof). This non-degeneracy condition holds true for any stationary process such that the spectrum is not purely atomic (see [16] for a proof). This condition is verified in the examples since their spectrum have a density.

The examples are the following, given by their covariance functions.

(1) $\Gamma_1(t) = \exp(-t^2/2)$
(2) $\Gamma_2(t) = (\cosh(t))^{-1}$
(3) $\Gamma_3(t) = \left(3^{1/2}t\right)^{-1} \sin(3^{1/2}t)$

Notice that in the three cases one has $\lambda_0 = \lambda_2 = 1$.

Γ_1 and Γ_3 have analytic extensions to the whole plane, so that Theorem 2.2 applies for any length T of the parameter interval. Theorem 2.2 applies only for length smaller than $\pi/4$ since the meromorphic extension of $\Gamma_2(.)$ has poles at the points $i\pi/2 + k\pi i$, k an integer.

The numerical results are taken from [6] where one can find details and some extensions. They correspond to $T = 1, 4, 6, 8, 10$, $u = -2, -1, 0, 1, 2, 3$ using three terms of the Rices series. If $T \leq 8$, the error in the computation of $P(M > u)$ is smaller than $0,01$ in all cases, excepting for $u = 0$.

In all cases, a lower-bound (resp. an upper-bound) is given using two (resp. three) terms of the Rices series. (In fact, in these computations, for simpler and

faster calculation instead of $\widetilde{\nu}_3$, ν_3 has been used, providing a slightly weaker upper-bound).

4. The Direct Method

We start with some notation and hypotheses that concern the Gaussian field $\mathcal{X}$ and the parameter set T, a subset of Euclidean space R^d.

We assume that T satisfies the following conditions [**A1**]:

(1) T is compact.
(2) T is the disjoint union of $S_d, S_{d-1}..., S_0$, where S_j is an orientable $\mathcal{C}^3$ manifold of dimension j without boundary.

The S_j's will be called faces. S_{d_0}, $d_0 \leq d$ is the non empty face having largest dimension. σ_j denotes the j-dimensional geometric measure on S_j.

Each S_j has an atlas such that the second derivatives of the inverse functions of all charts (viewed as diffeomorphisms from an open set in R^j to S_j) are bounded by a fixed constant. For $t \in S_j$, we denote L_t the maximum curvature of S_j at the point t. It follows that L_t is bounded for $t \in T$.

Notice that the decomposition

$$T = S_d \cup ... \cup S_0$$

is not unique. Quite general domains will satisfy these assumptions on T: as typical examples, smooth manifolds with boundary, convex polyhedra, unions of these kind of sets that verify the conditions, as well as smooth deformations of these sets.

As for $\mathcal{X}$ we make assumptions A2-A5 which can be verified using well-known results:

[**A2**]: $\mathcal{X}$ is defined on an open set containing T and has $\mathcal{C}^2$ paths

[**A3**]: for every $t \in T$ the distribution of $\big(X(t), X'(t)\big)$ does not degenerate; for every $s, t \in T$, $s \neq t$, the distribution of $\big(X(s), X(t)\big)$ does not degenerate

[**A4**]: Almost surely the maximum of $X(.)$ on T is attained at a single point.

For $t \in S_j$, $X'_j(t)$ $X'_{j,N}(t)$ denote respectively the derivative along S_j and the normal derivative. Both quantities are viewed as vectors in R^d, and the density of their distribution will be expressed respectively with respect to an orthonormal basis of the tangent space $T_{t,j}$ of S_j at the point t, or its orthogonal complement $N_{t,j}$. $X''_j(t)$ denotes the second derivative of X along S_j, at the point $t \in S_j$ and will be viewed as a matrix expressed in an orthonormal basis of $T_{t,j}$. Similar notations will be used for any function defined on S_j.

[**A5**]: Almost surely, for every $j = 1, ..., d$ there is no point t in S_j such that $X'_j(t) = 0$, $\det(X''_j(t)) = 0$.

4.1. The density of M. The fundamental property that we will use is a representation of the density of the maximum, given by the next theorem. See [9] for a proof.

Theorem 4.1. *Let $M = \max_{t\in T} X(t)$. Under assumptions A1 to A5, the distribution of M has the density*

$$p_M(x) = \sum_{t\in S_0} E\big(\chi_{A_x}\big|X(t)=x\big)p_{X(t)}(x) + \sum_{j=1}^{d}\int_{S_j} E\big(|\det(X_j''(t))|\chi_{A_x}\big|X(t)=x, X_j'(t)=0\big)p_{X(t),X_j'(t)}(x,0)\sigma_j(dt),$$

where $A_x = \{M \le x\}$.

We make some remarks concerning Theorem 4.1:

(1) Originally, this result was obtained for T an interval on the real line, with the aim of studying the regularity of the distribution of M. In fact, one can obtain very strong results concerning this regularity for one-parameter Gaussian processes, exploiting equalities of type (4.1) (see [7]).

The equality is stated in terms of the density, but it is obvious that integrating once, one obtains a similar equality for the distribution of the maximum.

These exact formulas are only implicit in what concerns p_M, since the maximum M appears in the right-hand side, inside the event A_x. However, they can be used to obtain estimations of the density (and the distribution). The most obvious way to do this is to replace the indicator functions by 1; this already provides some non-trivial upper-bounds.

The simplest example is the following (see [11], Ch. 7): Let $\mathcal{X}$ be a Gaussian, centered process defined on $[0,1]$ with unit variance and sufficient regular paths. Denote by $r(s,t) = E(X(s)X(t))$ its covariance and $r_{ij}(s,t) = \frac{\partial^{i+j}}{\partial^i s \partial^j t} r(s,t)$. We also assume the non-degeneracy condition that $r_{11}(t,t) = E\big[(X'(t))^2\big]$ does not vanish. By means of the deterministic time change, given by

$$Y(t) = X(\gamma^{-1}(t)) \text{ where } \gamma(t) = \int_0^t \sqrt{r_{11}(s,s)}ds,$$

one can check that the process $\{Y(t)\}$ has "unit speed", i.e., $E[(Y'(t))^2] \equiv 1$. So, to study the distribution of the maximum, we can also assume that the original process has "unit speed", which means that $Var(X'(t)) \equiv 1$, changing at the same time the domain $[0,1]$ into an interval of length $L = \gamma^{-1}(1)$. Then one can prove that:

$$p_M(u) \le p^+(u) := \varphi(u)\left[1 + (2\pi)^{-1/2}\int_0^L C(t)\varphi(u/C(t)) + u\Phi(u/C(t))dt\right], \tag{4.1}$$

with

$$C(t) := \sqrt{r_{22}(t,t) - 1}.$$

In particular, if the process $\mathcal{X}$ is stationary, the above bound can be simplified, since $C(t) = \sqrt{\lambda_4 - 1}$ and $L = 1/\sqrt{\lambda_2}$.

We will see some more refined bounds in what follows, when instead of 1 we replace the indicator function χ_{A_x} by some other upper-bound.

(2) One can replace $|\det(X_j''(t))|$ in the above conditional expectation by $(-1)^j \det(X_j''(t))$, since under the conditioning and whenever the event $\{M \leq x\}$ holds true, $X_j''(t)$ is negative semi-definite.
(3) One should be careful in using formula (4.1) since its form depends on the choice of an atlas of charts of each manifold S_j and put them together to compute the integral, for example, using a partition of unity. It is not difficult to see that in fact the result is independent of that choice (see [11] chapters 6 and 7).

4.2. A general bound for the density p_M. We introduce some further notations. For t in S_j, $j \leq d_0$ we define $C_{t,j}$ as the closed convex cone generated by the set of directions: $\{\lambda \in R^d : \ \|\lambda\| = 1 \ ; \exists \ s_n \in S, (n = 1, 2, ...)$ such that $s_n \to t$, $\frac{t-s_n}{\|t-s_n\|} \to \lambda$ as $n \to +\infty\}$, whenever this set is non-empty and $C_{t,j} = \{0\}$ if it is empty. We will denote by $\widehat{C}_{t,j}$ the dual cone of $C_{t,j}$, that is:

$$\widehat{C}_{t,j} := \{z \ \in \ R^d : \langle z, \lambda \rangle \geq 0 \text{ for all } \lambda \in C_{t,j}\}.$$

Notice that these definitions easily imply that $T_{t,j} \subset C_{t,j}$ and $\widehat{C}_{t,j} \subset N_{t,j}$. We will say that the function $X(.)$ has an "extended outward" derivative at the point t in S_j, $j \leq d_0$ if $X'_{j,N}(t) \in \widehat{C}_{t,j}$. An instant reflection shows that the name corresponds to the directions in which the function increases.

The next theorem is easy to prove.

Theorem 4.2. *Under assumptions A1 to A5, one has :*

(a) $p_M(x) \leq \overline{p}(x)$ *where*

$$\overline{p}(x) := \sum_{t \in S_0} E\big(\chi_{X'(t) \in \widehat{C}_{t,0}} \big| X(t) = x\big) p_{X(t)}(x) \ +$$

$$\sum_{j=1}^{d_0} \int_{S_j} E\big(|\det(X_j''(t))| \chi_{X'_{j,N}(t) \in \widehat{C}_{t,j}} \big| X(t) = x, X_j'(t) = 0\big) p_{X(t), X_j'(t)}(x, 0) \sigma_j(dt). \tag{4.2}$$

(b) $P\{M > u\} \leq \int_u^{+\infty} \overline{p}(x) dx.$

The actual interest of this Theorem depends on the feasibility of computing $\overline{p}(x)$. It turns out that this can be done in some relevant cases. The results can be compared with the approximation of $P\{M > u\}$ by means of $\int_u^{+\infty} p^E(x) dx$ given in [3] and [33] where $p^E(x)$ is obtained on replacing in the formula for $\overline{p}(x)$, $|\det(X_j''(t))|$ by $(-1)^j \det(X_j''(t))$. Under certain general conditions, $\int_u^{+\infty} p^E(x) dx$ is the expected value of the Euler-Poincaré characteristic (EPC) of the excursion set $E_u = \{t \in T : X(t) > u\}$. The advantage of $p^E(x)$ over $\overline{p}(x)$ is that one can have nice expressions for it for a large number of random fields (see [1]). Conversely $\overline{p}(x)$ has the advantage that it is an upper-bound of the true density $p_M(x)$. Hence, upon integrating once, it provides an upper-bound for the tail probability, *for every u value*. On the other hand, it is clear from the definitions

that

$$|p^E(x) - p_M(x)| \le \overline{p}(x) - p_M(x)$$

that is, the error when replacing $p_M(x)$ by $\overline{p}(x)$ is an upper-bound of the error when replacing by $p^E(x)$.

4.3. Computing $\overline{p}(x)$. In this subsection we will give a tractable formula for $\overline{p}(x)$, under certain assumptions on the random field $\mathcal{X}$ and the parameter set T.

We will assume that the process $\mathcal{X}$ is centered Gaussian, with a covariance function having the form

$$E\big(X(s)X(t)\big) = \rho\big(\|s-t\|^2\big), \tag{4.3}$$

where $\rho : R^+ \to R$ is of class $\mathcal{C}^4$. Without loss of generality, we assume that $\rho(0) = 1$. Assumption (4.3) is equivalent to saying that the law of $\mathcal{X}$ is invariant under orthogonal linear transformations and translations of the underlying parameter space R^d. We will also assume that the function $\rho\big(\|s-t\|^2\big)$ is a covariance for every dimension d. This is equivalent to saying that $\rho : R^+ \to R$ is the Laplace transform of a finite Borel measure on R^+ (see [28]).

As for the set T, we assume it has a polyhedral shape. More precisely, we mean that each $S_j (j = 1, ..., d)$ is a union of subsets of affine manifolds of dimension j in R^d. For short, we write ρ' for $\rho'(0)$ and ρ'' for $\rho''(0)$.

We denote by H_n (resp. $\bar{H}_n$) the Hermite (resp. probabilistic Hermite) polynomials, that is, defined for $n = 0, 1, 2, ...$ as:

$$H_n(x) := e^{x^2}\big(-\frac{\partial}{\partial x}\big)^n e^{-x^2},\ \overline{H}_n(x) := e^{x^2/2}\big(-\frac{\partial}{\partial x}\big)^n e^{-x^2/2}.$$

We will also need the integrals $I_n(v) = \int_v^{+\infty} e^{-t^2/2} H_n(t)dt$ $(n = 0, 1, 2, ...)$. An elementary computation gives:

$$\begin{aligned} I_n(v) = \ & 2e^{-v^2/2} \sum_{k=0}^{[\frac{n-1}{2}]} 2^k \ \frac{(n-1)!!}{(n-1-2k)!!} H_{n-1-2k}(v) \\ & + \chi_{\{n \text{ even}\}}\ 2^{\frac{n}{2}}\ (n-1)!!\sqrt{2\pi}(1-\Phi(x)) \end{aligned}$$

where for each positive integer m, $m!!$ is defined by

$$m!! = \prod(m-2p), \quad p : \text{integer},\ 0 < m-2p \le m,\ 0!! = 1.$$

We are now ready to state the expression for $\overline{p}(x)$ (see [9] for a proof):

Theorem 4.3. *Assume that the random field $\mathcal{X}$ is centered Gaussian, satisfies A1-A5 and has a covariance having the form (4.3) which verifies the conditions of the beginning of this subsection. Moreover, let T have polyhedral shape. Then $\overline{p}(x)$ can be expressed by means of the formula:*

$$\overline{p}(x) = \varphi(x) \left\{ \sum_{t \in S_0} \widehat{\sigma}_0(t) + \sum_{j=1}^{d_0} \big[\big(\frac{|\rho'|}{\pi}\big)^{j/2} \overline{H}_j(x) + R_j(x)\big] g_j \right\}\ , \tag{4.4}$$

where

- *g_j is a geometric parameter of the face S_j defined by*

$$g_j = \int_{S_j} \widehat{\sigma}_j(t)\sigma_j(dt), \tag{4.5}$$

where $\widehat{\sigma}_j(t)$ is the normalized solid angle of the cone $\widehat{C}_{t,j}$ in $N_{t,j}$, that is:

$$\widehat{\sigma}_j(t) = \frac{\sigma_{d-j-1}(\widehat{C}_{t,j} \cap \mathcal{S}^{d-j-1})}{\sigma_{d-j-1}(\mathcal{S}^{d-j-1})} \text{ for } j = 0, ..., d-1, \tag{4.6}$$

$$\widehat{\sigma}_d(t) = 1. \tag{4.7}$$

Notice that for convex or other usual polyhedra, each S_j can be partitioned into a finite number of pieces such that $\widehat{\sigma}_j(t)$ is constant in each one of them, so that g_j is equal to the sum of this constants multiplied by the j-dimensional geometric measure of the corresponding pieces.

- *For $j = 1, ..., d$,*

$$R_j(x) = \big(\frac{2\rho''}{\pi|\rho'|}\big)^{\frac{j}{2}} \frac{\Gamma((j+1)/2}{\pi} \int_{-\infty}^{+\infty} T_j(v) \exp\big(-\frac{y^2}{2}\big)\, dy \tag{4.8}$$

where

$$v := -\frac{1}{\sqrt{2}}\big((1-\gamma^2)^{1/2}y - \gamma x\big) \quad \textit{with} \quad \gamma := |\rho'|(\rho'')^{-1/2}, \tag{4.9}$$

$$T_j(v) := \Big[\sum_{k=0}^{j-1} \frac{H_k^2(v)}{2^k k!}\Big] e^{-v^2/2} \; - \; \frac{H_j(v)}{2^j(j-1)!} I_{j-1}(v), \tag{4.10}$$

.

We make some remarks on Theorem 4.3:

(1) The "principal term" is

$$\varphi(x)\left\{\sum_{t\in S_0} \widehat{\sigma}_0(t) + \sum_{j=1}^{d_0} \big[\big(\frac{|\rho'|}{\pi}\big)^{j/2} \overline{H}_j(x)\big] g_j\right\}, \tag{4.11}$$

which is the product of a standard normal density times a polynomial with degree d_0. Integrating once, we get -in our special case- the formula for the expectation of the EPC of the excursion set given in Adler and Taylor (2007).

(2) The "complementary term"

$$\varphi(x) \sum_{j=1}^{d_0} R_j(x) g_j, \tag{4.12}$$

can be computed by means of a formula, as it follows from the statement of the theorem. These formulas will be in general quite unpleasant due to the complicated form of $T_j(v)$. However, for low dimensions they are simple. For example:

$$T_1(v) = \sqrt{2\pi}\big[\varphi(v) - v(1-\Phi(v))\big], \tag{4.13}$$

$$T_2(v) = 2\sqrt{2\pi}\varphi(v), \tag{4.14}$$

$$T_3(v) = \sqrt{\frac{\pi}{2}}\big[3(2v^2+1)\varphi(v) - (2v^2-3)v(1-\Phi(v))\big]. \tag{4.15}$$

(3) One can prove that the complementary term (4.12) is equivalent, as $x \to +\infty$, to

$$\varphi(x)\ g_{d_0} K\ x^{2d_0-4}\ e^{-\frac{1}{2}\frac{\gamma^2}{3-\gamma^2}x^2}, \tag{4.16}$$

where the constant K is given by:

$$K = 2^{3d_0-2}\frac{\Gamma\big(\frac{d_0+1}{2}\big)}{\sqrt{\pi}(2\pi\gamma)^{d_0/2}(d_0-1)!}\rho''^{d_0/4}\Big(\frac{\gamma}{3-\gamma^2}\Big)^{2d_0-4}. \tag{4.17}$$

4.4. On the error $\overline{p}(x) - p_M(x)$. Next, we present the statement of two theorems in which "second order approximation" appears for random fields. For proofs and other more complicated results, especially in what concerns the geometry of the parameter set T, see [9].

The next theorem gives sufficient conditions to ensure that the error

$$\overline{p}(x) - p_M(x)$$

is bounded by a constant times a Gaussian density having strictly smaller variance than the maximum variance of $\mathcal{X}$. We assume that the maximum of the variance is not attained in $S \backslash S_{d_0}$. This excludes constant variance or some other stationary-like condition. For parameter dimension $d_0 > 1$, a related result is Theorem 3.3 of [33].

Theorem 4.4. *Assume that the random field $\mathcal{X}$ satisfies conditions A1-A5. With no loss of generality, we assume that* $\max_{t\in T} Var(X(t)) = 1$. *Assume that*

$$\max_{t\in T\backslash S_{d_0}} Var(X(t)) < 1.$$

Then there exist positive constants C, δ such that for every $x > 0$,

$$\overline{p}(x) - p_M(x) \le C\varphi(x(1+\delta)). \tag{4.18}$$

In our last result we show a general class of random fields for which we compute the exact asymptotic exponential order of the error. For the time being, it is the only general case with $d_0 > 1$ in which we are able to do this.

We assume that the random field $\mathcal{X}$ is centered, satisfies A1-A5, its law is isotropic and stationary so that the covariance has the form (4.3), and verifies the conditions in subsection 4.3. To simplify somewhat the computations, with no loss of generality, we add the normalization $\rho' = -1/2$.

Theorem 4.5. *Assume that $\mathcal{X}$ is centered, satisfies the conditions above and T is convex. Then*

$$\lim_{x\to+\infty} -\frac{2}{x^2}\log\big[\overline{p}(x) - p_M(x)\big] = 1 + \frac{1}{12\rho'' - 1}. \tag{4.19}$$

References

1. Adler, R. J.: *The Geometry of Random Fields*, Wiley, 1981.
2. Adler, R. J.: *An Introduction to Continuity, Extrema and Related Topics for General Gaussian Processes*, IMS, Hayward, Ca., 1990.
3. Adler, R. J. and Taylor J. E.: *Random fields and geometry*, Springer-Verlag, 2007.
4. Azaïs J-M., Bardet J-M. and Wschebor M.: On the Tails of the distribution of the maximum of a smooth stationary Gaussian Process. *ESAIM: Prob. et Stat.* **6** (2002), 177–184.
5. Azaïs, J-M. and Delmas, C.: Asymptotic expansions for the distribution of the maximum of a Gaussian random fields. *Extremes* **5** (2002), 2, 181–212.
6. Azaïs, J-M and Wschebor, M.: The Distribution of the Maximum of a Gaussian Process: Rice Method Revisited. *In and out of equilibrium: probability with a physical flavour*, Progress in Probability, 321-348, Birkhaüser, 2002.
7. Azaïs J-M. and Wschebor M.: On the regularity of the distribution of the Maximum of one parameter Gaussian processes. *Probab. Theory Relat. Fields*, **119** (2001) 70–98.
8. Azaïs J-M. and Wschebor M.: On the Distribution of the Maximum of a Gaussian Field with d Parameters. *Annals Applied Probability*, **15** (2005) 1A, 254–278.
9. Azaïs J-M. and Wschebor M.: A general expression for the distribution of the maximum of a Gaussian field and the approximation of the tail. *Stoch. Proc. Appl*, **118** (2008) 7, 1190–1218.
10. Azaïs J-M. and Wschebor M.: Erratum to A general expression for the distribution of the maximum of a Gaussian field and the approximation of the tail. *Stoch. Proc. Appl*, **120** (2010) 10, 2100–2101.
11. Azaïs J-M. and Wschebor, M.: *Level sets and extrema of random processes and fields*, Wiley, 2009.
12. Berman, S. M. An asymptotic formula for the distribution of the maximum of a Gaussian process with stationary increments. *J. Appl. Prob.*, **22** (1985) 454–460.
13. Berman, S. M.: *Sojourns and extremes of stochastic processes*, The Wadworth and Brooks, Probability Series, 1992.
14. Borell, C.: The Brunn-Minkowski inequality in Gauss space. Invent. Math., **30** (1975), 207–216.
15. Borell, C.: (2003). The Ehrhard inequality. *C.R. Acad. Sci. Paris*, Sér. I, **337** (2003), 663–666.
16. Cramér, H. and Leadbetter, M. R.: *Stationary and Related Stochastic Processes*, Wiley, 1967.
17. Davies, R. B.: Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **64** (1977), 247–254.
18. Fernique, X.: Régularité des trajectoires des fonctions aléatoires gaussiennes. *Ecole d'Eté de Probabilités de Saint Flour (1974). Lecture Notes in Mathematics*, 480, Springer-Verlag, 1975.
19. Landau, H. J. and Shepp, L. A.: On the supremum of a Gaussian process. *Sankya Ser. A* **32** (1970), 369–378.
20. Ledoux, M.: *The Concentration of Measure Phenomenon*. American Math. Soc., 2001.
21. Ledoux, M. and Talagrand, M.: *Probability in Banach Spaces*, Springer-Verlag, 1991.
22. Lifshits, M. A.: *Gaussian Random Functions*, Kluwer, 1995.
23. Marcus, M. B. and Shepp, L. A.: Sample behaviour of Gaussian processes, in: *Proc. 6th Berkeley Symp. Math. Stat. and Probab.* **2** (1972) 423–442, University of California Press, Berkeley.
24. Miroshin, R. N.: Rice series in the theory of random functions. *estn. Leningrad Univ. Math.* **1** (1974), 143-155.
25. Miroshin, R. N.: The use of Rice series. *Theor. Prob. Appl.* **28** (1983), 714-726.
26. Piterbarg, V; I.: Comparison of distribution functions of maxima of Gaussian processes. *Th. Probab. Appl.*, **26** (2001), 687-705.
27. Piterbarg, V. I.: *Asymptotic Methods in the Theory of Gaussian Processes and Fields.* Amer. Math. Society, 1996.
28. Shoenberg, I. J.: Metric spaces and completely monotone functions *Ann. Math.*, **39** (1938), 4, 811-841.

29. Sudakov, V. N. and Tsirelson, B. S.: Extremal properties of half spaces for spherically invariant measures (in Russian). *Zap. Nauchn. Sem. LOMI*, **45** (1974), 75-82.
30. Sun, J.: Tail Probabilities of the Maxima of Gaussian Random Fields, *Ann. Probab.*, **21** (1993), 34-71.
31. Talagrand, M.: Majorising measures: the general chaining. *Ann. Probab.*, **24** (1996), 1049-1103.
32. Taylor, J.E. and Adler, R. J.: Euler characteristics for Gaussian fields on manifolds. *Ann. Probab.*, **31** (2003), 533-563.
33. Taylor J. E., Takemura A. and Adler R. J.: Validity of the expected Euler Characteristic heuristic. *Ann. Probab.*, **33** (2005), 1362-1396.

Mario Wschebor: Centro de Matemática, Facultad de Ciencias, Universidad de la República, Montevideo, 11200 Iguá 4225, Uruguay

E-mail address: wschebor@cmat.edu.uy