The Annals of Applied Probability
2000, Vol. 10, No. 1, 1–74

SPECIAL INVITED PAPER

ON EXCURSION SETS, TUBE FORMULAS AND MAXIMA OF RANDOM FIELDS[1]

BY ROBERT J. ADLER

Technion–Israel Institute of Technology

This is a rambling review of what, with a few notable and significant exceptions, has been a rather dormant area for over a decade. It concentrates on the septuagenarian problem of finding good approximations for the excursion probability $P\{\sup_{t\in T} X_t \geq \lambda\}$, where λ is large, X is a Gaussian, or "Gaussian-like," process over a region $T \subset \Re^N$ and, generally, $N > 1$.

A quarter of a century ago, there was a flurry of papers out of various schools linking this problem to the geometrical properties of random field sample paths. My own papers made the link via Euler characteristics of the excursion sets $\{t \in T\colon X_t \geq \lambda\}$. A decade ago, Aldous popularized the Poisson clumping heuristic for computing excursion probabilities in a wide variety of scenarios, including the Gaussian. Over the past few years, Keith Worsley has been the driving force behind the computation of many new Euler characteristic functionals, primarily driven by applications in medical imaging. There has also been a parallel development of techniques in the astrophysical literature. Meanwhile, somewhat closer to home, Hotelling's 1939 "tube formulas" have seen a renaissance as sophisticated statistical hypothesis testing problems led to their reapplication toward computing excursion probabilities, and Sun and others have shown how to apply them in a purely Gaussian setting.

The aim of the present paper is to look again at many of these results and tie them together in new ways to obtain a few new results and, hopefully, considerable new insight. The "Punchline of this paper," which relies heavily on a recent result of Piterbarg, is given in Section 6.6: "In computing excursion probabilities for smooth enough Gaussian random fields over reasonable enough regions, the expected Euler characteristic of the corresponding excursion sets gives an approximation, for large levels, that is accurate to as many terms as there are in its expansion."

CONTENTS

July 1998; revised June 1999.

[1] Supported in part by the NSF and the Israel Science Foundation

AMS 1991 *subject classifications*. Primary 60G15, 60G60, 60G70; secondary 60G17, 62M40, 62H10.

Key words and phrases. Random fields, excursion sets, tube formula, extremal distributions, maxima, Euler characteristic.

1. Introduction.

1.1. *On Gaussian maxima.* One of the oldest, and most difficult, problems in the study of random processes has been precise determination of the probability

$$P\left\{\sup_{t\in T} X(t) \ \geq \ \lambda\right\}, \tag{1.1.1}$$

where X is a (real-valued) random process, T is its (possibly quite general) parameter space and $\lambda \in \Re$ is arbitrary.

When $T \in \Re$, and so is totally ordered, and X is Markovian, there are a number of quite successfull ways to attack this problem. We, however, shall be mainly concerned with the situation in which X is either Gaussian (and, in general, non-Markov) or, in a sense to be made precise later, somehow distributionally "related" to a Gaussian process, and in which T might be quite general. Most of the discussion will center around $T \subset \Re^N$, $N \geq 1$, in which case we shall call X a *random field*.

It is a fact that seems remarkable that, even today, there is no explicit formula for this simple probability in the general Gaussian situation, despite the fact that it appears in a large number of different problems. However, despite, or perhaps because of this, there are multitudes of approximations and techniques for deriving approximations to (1.1.1), particularly when λ is large.

My aim in this paper will be to review a number of different approaches to computing, or, to be more precise, approximating (1.1.1) and to compare their effectiveness. The most general techniques involve the notion of metric entropy (and, more recently, maximising measures) and give general bounds that are attractive to theoreticians but which, because of their dependence on unknown constants, are close to useless in practice[2] unless supplemented with additional information.

Another way of looking at (1.1.1) is via the *excursion sets*

$$A_\lambda \equiv A_\lambda(X, T) := \{t \in T : X(t) \geq \lambda\} \tag{1.1.2}$$

of the random field X on the set A over the level λ, since

$$P\left\{\sup_{t\in T} X(t) \geq \lambda\right\} \equiv P\{A_\lambda \neq \varnothing\}. \tag{1.1.3}$$

Of course, finding $P\{A_\lambda \neq \varnothing\}$ is no easier, either in principle or in practice, than finding $P\{\sup_T X(t) \geq \lambda\}$ directly, but there is an approximation based on excursion sets that turns out to be very helpful. In one dimension, the argument goes as follows:

$$\begin{aligned} P\left\{\sup_{t\in[0,T]} X(t) \geq \lambda\right\} &= P\{N_\lambda(T) \geq 1 \text{ or } X(0) \geq \lambda\} \\ &\leq P\{N_\lambda(T) \geq 1\} + P\{X(0) \geq \lambda\} \\ &\leq E\{N_\lambda(T)\} + P\{X(0) \geq \lambda\}, \end{aligned} \tag{1.1.4}$$

where $N_\lambda(T)$ is the number of upcrossings[3] of the level λ during $[0, T]$.

---

[2]By terms like "in practice" and "applied," I will be referring to situations in which one wants a numerical value for the probability in (1.1.1) correct to within a percentage point or two, and *not* applications in which (1.1.1) is used as a tool in some other theoretical endeavour.

[3]If $F(t)$ is a continuous function on $[0, T]$ such that $F(t)$ is not identically equal to λ in any interval and neither $F(0)$ nor $F(T)$ equals λ, then F is said to have an upcrossing of the level λ at the point t_0 if there exists an $\varepsilon > 0$ such that $F(t) \leq \lambda$ in $(t_0 - \varepsilon, t_0)$ and $F(t) \geq \lambda$ in

If X is a zero-mean, stationary, almost surely continuous Gaussian process on $[0, T]$, then a famous formula due initially to Rice [84] in 1945 (with extensions, generalizations and rigorizations due to a number of authors, including [16, 42, 124]) gives the mean number of level crossings as

$$E\{N_\lambda\} = \frac{T\lambda_2^{1/2}}{2\pi\sigma} \exp\left(-\frac{\lambda^2}{2\sigma^2}\right), \tag{1.1.5}$$

where $\sigma^2 = E\{|X(t)|^2\}$ and λ_2 is a spectral parameter that will be defined later.

Throughout this paper we shall use time and again the following basic Gaussian tail inequalities, in which

$$\Psi(\lambda) = (2\pi)^{-1/2} \int_\lambda^\infty \exp\left(-\frac{1}{2}x^2\right) dx, \tag{1.1.6}$$

denotes the tail of the standard Gaussian distribution function: For all $\lambda > 0$,

$$\left(1 - \frac{\sigma^2}{\lambda^2}\right) \frac{\sigma}{\lambda\sqrt{2\pi}} \exp\left(\frac{-\lambda^2}{2\sigma^2}\right) \le \Psi\left(\frac{\lambda}{\sigma}\right) \le \frac{\sigma}{\lambda\sqrt{2\pi}} \exp\left(\frac{-\lambda^2}{2\sigma^2}\right). \tag{1.1.7}$$

Putting this together with (1.1.4) and (1.1.5) yields, for $\lambda > 0$,

$$\begin{aligned} P\left\{\sup_{t\in[0,T]} X(t) \ge \lambda\right\} &\le \left[\frac{\sigma}{\lambda\sqrt{2\pi}} + \frac{T\lambda_2^{1/2}}{2\pi\sigma}\right] \exp\left(\frac{-\lambda^2}{2\sigma^2}\right) \\ &\equiv \left[C_1 + C_2\lambda^{-1}\right] \exp\left(\frac{-\lambda^2}{2\sigma^2}\right). \end{aligned} \tag{1.1.8}$$

In fact, for large λ, it turns out that the right-hand side of (1.1.8) actually serves as the first two terms of an expansion in decreasing powers of λ for the probability on the left-hand side, and this expansion is accurate up to $O(\lambda^{-2})$[4]; that is,

$$P\left\{\sup_{t\in[0,T]} X(t) \ge \lambda\right\} = \left[C_1 + C_2\lambda^{-1} + O(\lambda^{-2})\right] \exp\left(\frac{-\lambda^2}{2\sigma^2}\right). \tag{1.1.9}$$

for large λ. Note that from our derivation of (1.1.8) the first two constants here are known, and only the details surrounding the error term remain unspecified.

This simple result can be taken as the archetype of the results to be discussed in this review, in which, for more general processes over more general spaces, (1.1.9) will be replaced by

$$P\left\{\sup_{t\in T} X(t) \ge \lambda\right\} = \lambda^\alpha \exp\left(\frac{-\lambda^2}{2\sigma^2}\right) \left[\sum_{n=0}^{n'} C_n\lambda^{-n} + o(\lambda^{-n'})\right] \tag{1.1.10}$$

$(t_0, t_0 + \varepsilon)$. Note that this implies that only relatively smooth functions can have a finite number of upcrossings. Brownian motion paths, for example, do not fall into this category.

[4]Even better levels of accuracy are often available; cf. Theorem 4.5.2 and the surrounding discussion.

for some α and n' that will depend on specific properties of both the process and its parameter space.

It will turn out that the general metric entropy techniques hinted at before usually suffice to identify α in (1.1.10), with $n' = 0$ and C_0 undeterminable. A very brief discussion of these results forms the content of the following section.

Level crossing, and, in the general case, excursion set techniques can generally be made to *rigorously* give (1.1.10) with $n' = 0$, and with precise values of α and C_0. (In higher dimensions, the number of level crossings must be replaced by a generalization of this concept known as the *Euler characteristic* of the excursion set, to be defined soon.) However, we shall see that, in fact, they seem to give the expansion up to $n' = 2[N/2]$, where N is the dimension of the parameter space and throughout this paper $[x]$ denotes the integer part of x. Furthermore, the constants $C_0, \ldots, C_{2[N/2]}$ are explicitly computable. Why this is so, and how the argument involves the geometry of excursion sets, will be the topic of Section 4 and parts of Section 6.

Section 5 looks at a technique that works when the excursion set approach does not, a technique developed by Pickands [71, 72] in the late 1960's for processes on the line, extended to random fields by Bickel and Rosenblatt [13] and Qualls and Watanabe [81] in 1973 and surveyed and developed in the recent monograph of Piterbarg [76]. To explain what goes wrong with the excursion set approach, it is important to note that while (1.1.5) is *always* correct under the conditions stated, if X is not differentiable (e.g., an Ornstein–Uhlenbeck process) then the constant λ_2 is infinite. Consequently, everything that we built on the basis of this result fails to work in this scenario. The Pickands–Piterbarg approach avoids this issue with an entirely different approach, which, however, requires a lot more work. Following Piterbarg, and for reasons that will become clearer later, we call this approach the "double-sum technique."

Elements of both of the preceding approaches can be found in David Aldous' superb monograph *Probability Approximations via the Poisson Clumping Heuristic*[5] [6], and the reader who wishes to find the distribution of the maximum of a specific process should read Aldous' book in conjunction with, or preferably before, reading this paper. The relationship between our formulas and Aldous' clumping heuristic will be made clear in Section 5.

In the penultimate Section 6 an old/new, classical/innovative approach to Gaussian maxima problems involving "tube formulas" will be discussed. This is an approach that, in principle, leads to expansions of the form of (1.1.10)

[5]Actually, the title of this book is somewhat of a misnomer. The word "heuristic" gives the reader the misguided impression that "Poisson clumping" is a rough-and-ready way for illegitimately guessing answers that sometimes works and sometimes does not. The word "principle" is probably more appropriate, since with the idea of Poisson clumping Aldous has managed to distill the essence of a large number of related problems to identify a common underlying principle, which almost always works. (Of course, I do have to admit that "almost always" here depends on the measure one places on spaces of problems.)

of a comparatively high order (n') when other methods fail. However, what is possible "in principle" is not always doable in practice, and so we shall discuss both the strengths of, and the problems associated with, this new technique. This approach, however, is applicable only to Gaussian processes.

Since the techniques based on Poisson clumping and excursion sets are more widely applicable, at various points in the paper we will also discuss Euler characteristics for a variety of non-Gaussian random fields. All of these fields, however, have some "relationship" to their Gaussian counterparts, having marginal distributions that are chi-squared, F, t and so on.

In general, with a few notable exceptions that will be discussed later, the last decade has been a rather dry one for Gaussian random fields and associated excursion problems. The time, and applications, seem to be ripe for renewed activity in the area. Thus, in the closing Section 7, I have listed what seem to me to be a number of interesting open problems and general directions for future research.

1.2. *Gaussian and related fields.* Here is a crash course on the basics of stationary Gaussian processes and some basic notation that will remain constant throughout the paper. It can be skipped on first (and subsequent) reading and referred to when necessary.

The parameter spaces of random processes, or subsets of them, will be denoted by capital Latin letters. The most common will be T, and this might be quite general, at least in the following section, where examples would include families of test functions, all measurable subsets of some measure space and so forth. Throughout, we assume that T is totally bounded in the pseudo metric $d(s,t) = (E\{|X(t) - X(s)|^2\})^{1/2}$. From Section 3 onward, where we will be looking at detailed tail probability computations, we will generally have $T \subset \Re^N$ for some $N \geq 1$. Random processes will share the same alphabet, but closer to its end. If $T \subset \Re^N$ and the process is $\Re^d$-valued, we shall call it an (N,d) random field, or, generally only when $d = 1$, simply a random field.

The mean function of a real- or complex-valued process $\{X(t)\}_{t\in T}$ is denoted by $m(t) = E\{X(t)\}$, and the covariance function by

$$R(s,t) \;=\; E\left\{[X(s) - m(s)]\left[\bar{X}(t) - \bar{m}(t)\right]\right\}, \tag{1.2.1}$$

where the complex conjugation represented by the overbar is only meaningful if X is complex valued. If X is real valued, then R is clearly a symmetric function of s and t. If m is constant and R is a function only of $(s - t)$, then we call X stationary, and, with some abuse of notation, write

$$R(s,t) \;=\; R(t - s).$$

If, in addition, $T \in \Re^N$ and $R(t)$ is only a function of $\|t\| = \|(t_1, \ldots, t_N)\| := (\sum_{n=1}^N t_n^2)^{1/2}$, then X is called *isotropic*.

By Bochner's theorem, for every stationary covariance function R on $\Re^N$ there is a spectral distribution function F such that

$$R(t) \;=\; \int_{\Re^N} \exp(i\, t \cdot \lambda)\, dF(\lambda), \tag{1.2.2}$$

where we shall use either $t \cdot \lambda$ or $\langle t, \lambda \rangle$ to denote the inner product in $\Re^N$. (Bochner's theorem has, of course, natural extensions to more general spaces with a group structure, but we shall not need these.) Furthermore, there is a spectral process Z, with independent increments, such that, in this case, we can write X via its spectral representation as

$$X(t) = \int_{\Re^N} \exp(i\, t \cdot \lambda)\, dZ(\lambda). \tag{1.2.3}$$

Putting (1.2.1)–(1.2.3) together, it is easy to check that, for $\alpha, \beta, \gamma, \delta \in \{0, 1, 2, \ldots\}$,

$$\begin{aligned} E\left\{ \frac{\partial^{\alpha+\beta} X(t)}{\partial^\alpha t_i \partial^\beta t_j} \cdot \frac{\partial^{\gamma+\delta} X(t)}{\partial^\gamma t_k \partial^\delta t_l} \right\} &= (-1)^{\alpha+\beta} \frac{\partial^{\alpha+\beta+\gamma+\delta}}{\partial^\alpha t_i \partial^\beta t_j \partial^\gamma t_k \partial^\delta t_l} R(t)\Big|_{t=0} \\ &= i^{\alpha+\beta+\gamma+\delta} \int_{\Re^N} \lambda_i^\alpha \lambda_j^\beta \lambda_k^\gamma \lambda_l^\delta \, dF(\lambda). \end{aligned} \tag{1.2.4}$$

Since they will arise often, we shall denote the second spectral moments $\int_{\Re^N} \lambda_i \lambda_j \, dF(\lambda)$ by λ_{ij}, $i, j = 1, \ldots, N$, and denote the $N \times N$ matrix of these moments by Λ. Note that by (1.2.4) we could also define the components of Λ using only derivatives of R, without ever referring to the spectrum.

It follows from (1.2.4) and appropriate choices of $\alpha, \beta, \gamma, \delta$ that for a real-valued stationary field $X(t)$ [for which F is symmetric in the sense that $F(A) = F(-A)$ for all measurable $A \in \Re^N$] the following relationships hold for all $1 \le i, k, l \le N$:

$$X(t) \text{ and } X_i(t) \text{ are uncorrelated} \qquad (\beta = \gamma = \delta = 0,\ \alpha = 1), \tag{1.2.5}$$

$$X_i(t) \text{ and } X_{kl}(t) \text{ are uncorrelated} \qquad (\alpha = \gamma = \delta = 1,\ \beta = 0). \tag{1.2.6}$$

The preceding discussion is quite general and has nothing to do with Gaussian processes, to which we now turn. Recall that a (zero-mean) Gaussian process is a family $\{X_t\}_{t \in T}$ of random variables, indexed by a parameter set T, such that each linear combination $\sum \alpha_t X_t$ is (centered) Gaussian. Equivalently, we could require that all finite collections $(X(t_1), \ldots, X(t_k))$ are multivariate Gaussian with covariance matrix $(R(t_i, t_j))_{i,j=1,\ldots,k}$. It therefore follows that the mean and covariance functions determine the law of the entire process, one of the things that makes Gaussian processes so easy to study. Furthermore, since derivatives are limits of linear combinations, if $X(t)$ is a Gaussian field on $\Re^N$ then so are its derivatives, and the joint distribution of $X(t)$ with its derivatives is also multivariate Gaussian. Thus, in the Gaussian case, the variables considered in (1.2.5) and (1.2.6) are not only uncorrelated but also independent. We shall make repeated use of this fact.

Now the time has come to explain what "related" means in the title of this subsection. In essence, it means "any random field to which one can extend the kind of excursion theory analysis that works for Gaussian processes." In practice, this means any random field for which it is possible to compute a precise

formula for the expected value of the Euler characteristic of its excursion sets. One such example is given by the so-called χ^2 processes, representable as

$$\chi^2_K(t) = \sum_{n=1}^{K} X_n^2(t), \tag{1.2.7}$$

where the X_n are independent, identically distributed, zero-mean Gaussian processes. It is easy to check that if the X_n are stationary then so is χ^2_K, and to compute the spectral moments of χ^2_K in terms of those of the X_n. When it comes to looking at the joint distribution of $\chi^2_K(t)$ and its various derivatives, (1.2.5) is still relevant, but now lack of correlation no longer implies independence, a fact that is crucial for calculations in the Gaussian case. Nevertheless, judicious conditioning arguments allow one to find independence where unconditionally there is none, and expected Euler characteristics can be computed.

With χ^2_K fields defined, we leave it to the reader to guess how fields with names like "F," "t," "noncentral χ^2_K" and so on may be defined. In most of these cases, excursion characteristic calculations can be carried through, and an excursion theory parallel to the Gaussian one developed. The examples of the following subsection will give some indication of why these fields are important in practice.

1.3. *Some motivating examples.* In this subsection, I shall describe, very briefly, two problems related to Gaussian maxima that are nicely solved by applications of excursion set techniques. The solutions themselves will come in Section 3.4.

We begin at the *very* beginning: the beginning of the Universe. According to current astrophysical theory, 99.97% of the radiant energy of the Universe was released within the first year after the Big Bang, and much of the structure of that time is still measurable in terms of today's background microwave radiation. Theories that attempt to explain the origin of large-scale structure seen in the Universe today must therefore conform to the constraints imposed by these observations.

Figure 1 shows, in gray scale,[6] the anomalies in the cosmic microwave background (CMB) radiation, divided by their standard deviation. This was the first evidence of anomalies in the CMB radiation, a sort of signature left over from the creation of the Universe. It was a widely published result; the reader is referred to the article in *Scientific American* [80] or the 5-page, 28-author, original paper [102].

The initial data are actually directional, as they represent radiation coming into a point from the surrounding universe. As such, it is actually a random field on a two-dimensional sphere. The map shown here is a projection of the full sky in galactic coordinates, with the plane of the Milky Way placed

[6]For far more attractive and informative color figures, see, for example, www.gsfc.nasa.gov/astro/cobe/ and follow the links. I "borrowed" the figure from [122], where it also exists in color.

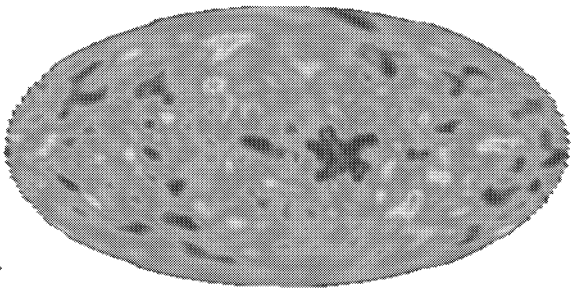

FIG. 1. *Galactic radiation.*

horizontally in the middle of the map and with the galactic center at the center.

Very briefly, astronomers are interested in the apparent randomness in the patterns in maps like Figure 1 and in determining whether or not the image can be considered as the realisation of a Gaussian (or other) random field. One very simple way to test a null hypothesis of normality is to look at the statistic

$$\sup_{t\in\text{sphere}} (\text{Standardized CMB anomaly}(t)) \tag{1.3.1}$$

and to compare it to the maximum of a parametrically matched Gaussian (or other) field. For this, of course, one needs the probability (1.1.1). However, there are better ways to test such a hypothesis and to understand galactic topography than merely looking at the maximum, and excursion sets provide an appropriate, and, over the last 10 years, heavily used, tool.

We shall return to this later.

For our second example, we take a problem in biostatistics, related to brain imaging. A small number of experimental subjects are injected with a positron emitting radio isotope, which enables a positron emission tomography (PET) machine to follow its flow through the bloodstream, and, in particular, in the brain. Images of the brain are taken while the subjects are at rest, and then when performing a task, such as the silent reading of words projected onto a screen. The underlying principle is that those parts of the brain involved in performing the task will require oxygen, and hence blood, so that there will be a heavier blood flow to, and hence positron emission from, these regions. The difference between the rest and task images, averaged over the subjects and standardised at each point by the standard deviation of the sample, is shown in Figure 2, where the dark regions depict excursion sets over two different levels. As with the galactic example, the main problem of interest lies in determining whether these pictures are consistent with a noise model or whether the excursion sets above the highest levels contain information about which part of the brain handles a specific task. Of course, in this case, the averaging over a small number of subjects, differencing and standardisation, lead to a "t," rather than Gaussian, random field.

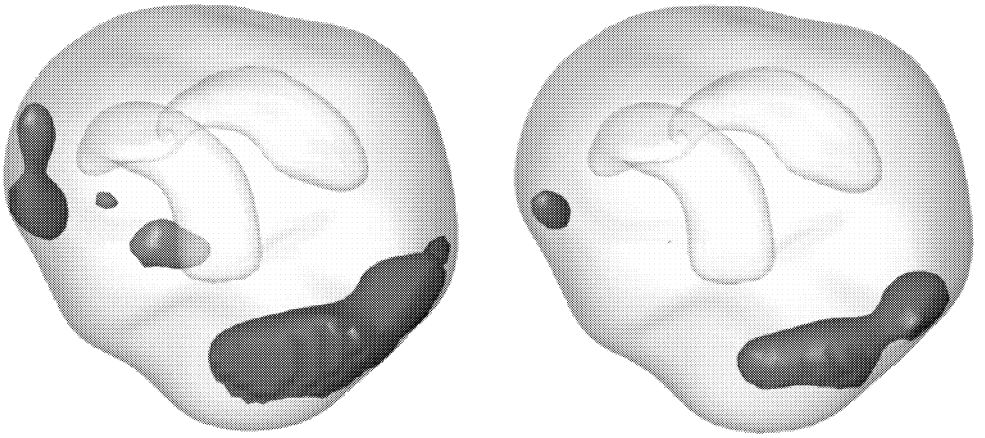

FIG. 2. *Excursion sets for PET brain images.*

1.4. *On the importance of formulas: Some philosophy.* Since it has been quite some time since I have written a paper about random fields, and since this paper is already so long that a few more words will not make much of a difference, I am going to allow myself a few words of "wisdom" and see if I can slip them by the editor.

In 1981, 5 years out of graduate school, and with a beard somewhat shorter and colored somewhat differently from my present one, I published a book entitled *The Geometry of Random Fields*. As with most young authors, I was very pleased with myself.

The book contained clever arguments involving integral and differential geometry, and had a nice long chapter on Hausdorff dimension, just before fractals became a household word. As with most young authors, I was also certain that the clever ideas expressed in the book (most of which, of course, were culled from the works of others) would leave a lasting impression on the academic world.

Close to two decades later, and with the wisdom of hindsight (there have to be some advantages to a white beard!), it is clear that what was important in my book was not so much the ideas, but one very specific formula. This was an *explicit* formula for the expectation of a topological characteristic of the excursion sets of Gaussian fields [cf. formula (3.2.5) of this paper].

Its importance lay in the fact that, as for Rice's formula (1.1.5), its one-dimensional predecessor, its simple form allowed one to substitute real parameter values and obtain actual, meaningful numbers that could, and have been, used for the solving of problems outside of mathematics.[7]

[7]There is a widespread belief among modern pure mathematicians that the major contribution they have to make to "science" is the development of "understanding," generally at the expense of explicit results. Strangely enough, most subject matter scientists do not share the mathematicians'

In many ways, this paper, the new parts of which primarily report on the work of others, is an outgrowth of this formula, its extensions and its applications, how they are related and, in part, how they are derived. Throughout, however, the emphasis will be on presenting, in a coordinated and reasonably comprehensive fashion, as many useful formulas as possible, since I now understand their relative importance. Discussion of techniques will be limited to what is needed to understand and apply the formulas, and to help the reader develop new ones appropriate to his or her own specific needs. Thus the reader should look upon this paper far more as a cookbook for the extrema of random fields, and not expect a connoisseur's description of the taste of each dish. However, as a referee noted, in a perhaps apochryphal remark he had heard attributed to Richard Hamming, "The goal of computations is understanding. The goal of understanding is to compute."

2. First-order results for maxima.

2.1. *Basic large deviations.* Let X be a centered Gaussian random variable with variance σ^2. Then from (1.1.7) we have, for all $\lambda > 0$,

$$\left(1 - \frac{\sigma^2}{\lambda^2}\right) \frac{\sigma}{\lambda\sqrt{2\pi}} \exp\left(\frac{-\lambda^2}{2\sigma^2}\right) \le P\{X > \lambda\} \le \frac{\sigma}{\lambda\sqrt{2\pi}} \exp\left(\frac{-\lambda^2}{2\sigma^2}\right). \tag{2.1.1}$$

An immediate, and trivial, consequence of (2.1.1) is

$$\lim_{\lambda\to\infty} \lambda^{-2} \log P\{X > \lambda\} = -(2\sigma^2)^{-1}. \tag{2.1.2}$$

A classical result of Landau and Shepp [52] and Marcus and Shepp [63] gives a result closely related to (2.1.2), but for the supremum of a general centered Gaussian process. If we assume that $\{X_t\}_{t\in T}$ has bounded sample paths with probability 1, then they showed that

$$\lim_{\lambda\to\infty} \lambda^{-2} \log P\left\{\sup_{t\in T} X_t > \lambda\right\} = -(2\sigma_T^2)^{-1}, \tag{2.1.3}$$

where

$$\sigma_T^2 := \sup_{t\in T} EX_t^2.$$

In view of (2.1.2), the asymptotics of (2.1.3) seem at first rather surprising, since they seem to imply that the supremum of a centered, bounded Gaussian process behaves much like a single Gaussian variable with a suitably chosen variance. However, it is important to note that the large deviation result (2.1.3) is consistent with any of the forms

$$C\lambda^\alpha \exp\left(\frac{-\lambda^2}{2\sigma_T^2}\right), \qquad C \exp(\lambda^\beta) \exp\left(\frac{-\lambda^2}{2\sigma_T^2}\right) \tag{2.1.4}$$

---

enthusiasm for insight. It seems that they generally know their subject well enough to develop their own insight. However, useful formulas are quite a different issue.

for any $\alpha \in \Re$ and any $\beta < 2$, as well as combinations of these and similar forms. What *Gaussian* large deviation theory is about, as opposed to general large deviation theory, is going beyond (2.1.3) and identifying the correct form, along with constants like α and β, in (2.1.4). What is perhaps surprising is that this is possible in quite wide generality.

The basic result, from which the modern proofs of all others follow, is the following inequality, a version of what is generally known as "Borell's inequality" but is, in fact, due independently to Borell [14] (who gave a proof based on isoperimetric inequalities for the n-sphere) and Cirelśon, Ibragimov and Sudakov [21] (who gave a stochastic analytic proof later rediscovered by Maurey and Pisier [73]). Details[8] can be found in [2]. The inequality states that, for all $\lambda > 0$,

$$P\left\{\left|\sup_{t\in T} X_t - E\left\{\sup_{t\in T} X_t\right\}\right| > \lambda\right\} \le 2\exp\left(\frac{-\lambda^2}{2\sigma_T^2}\right). \tag{2.1.5}$$

An immediate consequence of (2.1.5) is that, for all $\lambda > E \sup X$,

$$P\left\{\sup_{t\in T} X_t > \lambda\right\} \le 2\exp\left(-\tfrac{1}{2}\left(\lambda - E\left\{\sup_{t\in T} X_t\right\}\right)^2 \Big/ \sigma_T^2\right). \tag{2.1.6}$$

Noting the trivial lower bound $P\{\sup_{t\in T} X_t > \lambda\} \ge P\{X_{t^*} > \lambda\}$, where $t^* \in T$ is the point at which the supremum variance σ_T^2 is achieved, both (2.1.2) and (2.1.3) follow from the previous equation and (2.1.1). More delicate inequalities involving terms as in (2.1.4) would also follow if one knew how to calculate $E\{\sup X\}$. However, it is almost as difficult to compute this expectation as it is to compute exceedence probabilities.

Nevertheless, there are techniques to obtain bounds for $E\{\sup X\}$ in quite general situations, and these lead us to the approach of the following subsection.

2.2. *Entropy.* Perhaps surprisingly, we now lose no precision by going from random fields on $\Re^N$ to very general parameter spaces, as all we really need is (T, d) to be a totally bounded metric space, where d is the canonical pseudo metric given by $d^2(s, t) = E(X_t - X_s)^2$.

Let $N(T, d, \varepsilon) \equiv N(\varepsilon)$ be the smallest number of (open) d-balls of radius ε needed to cover T. The natural logarithm of N is known as the entropy function for (T, d).

---

[8]Actually, this is a good place to point out that the third line from the bottom of page 46 in [2], where I gave a proof of (2.1.5), is somewhat misleading. The problem, as usual, lies in the sixth word in the sentence starting, "To complete the proof note simply that ... ." Missing details can be obtained from iew3.technion.ac.il:8080/Adler.phtml.

Then results going back to [30, 31, 104, 105] (cf. [2] and [54] for details[9]) yield the existence of universal constants C_1 and C_2 for which

$$C_1 \sup_{\varepsilon>0} \varepsilon\big(\log N(\varepsilon)\big)^{1/2} \leq E \sup_{t\in T} X_t \leq C_2 \int_0^{\operatorname{diam}(T)} (\log N(\varepsilon))^{1/2}\, d\varepsilon. \tag{2.2.1}$$

Since entropies are generally quite simple to compute (cf. [2] for 20 pages of worked examples) and to rephrase in terms of the covariance function, (2.2.1) yields workable bounds that can be substituted into (2.1.6) to obtain bounds on Gaussian maxima distributions which will improve on the large deviation bound of (2.1.3). However, since the new bounds involve universal constants that hold for a very wide variety of situations, one cannot expect them to be sharp.

If one is prepared to assume more about the form of the entropy function, and work considerably harder, then the upper bound of (2.1.6) can be considerably improved. For example, in a series of papers [86, 87, 110, 88, 112] by Samorodnitsky and Talagrand,with a leapfrogging of ideas and techniques, the results of the following three theorems, among others, were obtained.

THEOREM 2.2.1. *If for some $A > \sigma_T$, some $\alpha > 0$ and some $\varepsilon_0 \in [0, \sigma_T]$ we have*

$$N(T, d, \varepsilon) \leq \left(\frac{A}{\varepsilon}\right)^{\alpha} \tag{2.2.2}$$

for all $\varepsilon < \varepsilon_0$, then for $\lambda \geq \sigma_T^2[(1+\sqrt{\alpha})/\varepsilon_0]$ we also have

$$P\left\{\sup_{t\in T} X_t \geq \lambda\right\} \leq \left(\frac{KA\lambda}{\sqrt{\alpha}\sigma_T^2}\right)^{\alpha} \Psi\left(\frac{\lambda}{\sigma_T}\right), \tag{2.2.3}$$

where K is a computable numerical constant independent of X and T.

THEOREM 2.2.2. *Set*

$$T_\delta = \left\{t \in T : EX_t^2 \geq \sigma_T^2 - \delta^2\right\}. \tag{2.2.4}$$

Suppose there exist $\alpha > \beta > 1$ such that, for all $\delta > 0$, $\varepsilon \in (0, \delta(1+\sqrt{\alpha})/\sqrt{\beta})$, we have

$$N(T_\delta, d, \varepsilon) \leq A\delta^{\beta}\varepsilon^{-\alpha}. \tag{2.2.5}$$

[9]Better bounds than (2.2.1) exist in terms of so-called *majorising measures*, for details of which both [2] and [54], and especially the simpler approach of [111], can be consulted. However, these more delicate techniques have little to offer in terms of the bounds, with *good* constants, that interest us here.

Then, for $\lambda \geq 2\sigma_T\sqrt{\beta}$, *we also have*

$$P\left\{\sup_{t\in T} X_t \geq \lambda\right\} \leq \frac{A\beta^{\beta/2}}{\alpha^{\alpha/2}} K^{\alpha+\beta} \left(\frac{\lambda}{\sigma_T^2}\right)^{\alpha-\beta} \Psi\left(\frac{\lambda}{\sigma_T}\right). \tag{2.2.6}$$

If $1 < \alpha < \beta$, *then there exist universal* K_1, K_2 *such that*

$$K_1\Psi\left(\frac{\lambda}{\sigma_T}\right) \leq P\left\{\sup_{t\in T} X_t > \lambda\right\} \leq K_2\Psi\left(\frac{\lambda}{\sigma_T}\right). \tag{2.2.7}$$

It is also known that, when $\alpha > \beta > 1$, the right-hand side of (2.2.6), albeit with different constants, also serves as a lower bound to the probability there if multiplied by a factor of $\lambda^{\alpha-\beta}$.

These results treat only entropy functions in which the growth of N in ε is of a power form, which, at least for processes indexed by points in Euclidean spaces, are the most common. In the case of exponential entropy structures, the following theorem holds.

THEOREM 2.2.3. *Suppose there exist* $A, B > 0$ *and* $\alpha \in (0, 2)$ *such that*

$$N(T, \varepsilon) < A\exp(B\varepsilon^{-\alpha}).$$

Then, for all $\lambda > 0$,

$$P\left\{\sup_{t\in T} X_t > \lambda\right\} \leq K_1 \exp(K_2\lambda^{2\alpha/(2+\alpha)})\,\Psi\left(\frac{\lambda}{\sigma_T}\right),$$

where K_1 *and* K_2 *are universal.*

Note that one cannot set $\alpha = 0$ in this result to recover either Theorem 2.2.1 or Theorem 2.2.2. The upper bound given here is, under mild side conditions, also a lower bound.

[There are also nice asymptotic bounds due to Lifshits [56], who, in a quite general setting, has shown that

$$P\left\{\sup_{t\in T} X_t > \lambda\right\} \asymp (2-p)^{1/2}\Psi_p\left(\frac{\lambda^{2-p} - d\lambda^{1-p}}{p\sigma_T^2}\right)$$
$$\times \exp\left\{-\frac{(2-p)\lambda^2}{2p\sigma_T^2} - \frac{d(p-1)\lambda}{p\sigma_T^2} + \frac{d^2}{2\sigma_T^2}\right\}\Psi\left(\frac{\lambda - d}{\sigma_T}\right) \tag{2.2.8}$$

for all $1 \leq p < 2$, where $\Psi_p(x) \equiv E\{\exp(x\sup_{t\in T} {X_t}^p)\}$ and d is a constant (not easily) determined by X. Of course, this result is somewhat circular, since one needs the Laplace transform of $\sup_{t\in T} {X_t}^p$ before being able to compute anything. Nevertheless, (2.2.9) leads to some nice theoretical results about ℓ^p-valued Gaussian processes.]

I am not going to say anything here about the way the proofs of the preceding theorems work, other than to note that the basic element of the most general approach will also lie at the core of the more detailed results to follow.

One looks for a subset $T_{\max} \subset T$ (very often a unique point in T) where the maximal variance is achieved, and one studies two things: the "size" of $T_{\max}$ (e.g., as measured in terms of metric entropy) and how rapidly EX_t^2 decays as we move out of $T_{\max}$. The underlying idea is that the supremum of the process will occur in the region of maximal variance, and the rate of decay of EX_t^2 outside that region determines how important, if at all, the behavior is elsewhere. All this should become clearer when we come to the Poisson clumping heuristic.

Now, all these results are very nice, and very elegant, and certainly show how much more one can say about large deviations in the Gaussian case than in the general situation. However, when it comes to applications, there are two problems with them. The first is that, in the notation of (1.1.10), a result like (2.2.6) gives only the first term in the expansion of the exceedence probability, and we have already decided that we need more. However, beyond this, there is the issue that in order to get good bounds universal constants really need to be replaced by constants tailored to specific examples.

As for the possibility that this could be done using the approach developed in the general setting, we can do no better than to quote the master in [112]: "In particular, it must be pointed out that while our approach is unlikely ever to yield optimal constants, it does not use chaining (that makes the search of sharp numerical constants hopeless). We have, however, felt that the search of sharp numerical constants is better left to others with the talent and taste for it."

And, of course, to other tools. We start developing these now, with what, at first, will seem like (and, in part, is) quite a digression from the main theme of this review.

3. Euler characteristic techniques.

3.1. *The Euler (and related) characteristics.* Since Rice's formula (1.1.5) leads so simply and naturally to a rigorous bound and an approximation to excursion probabilities for stationary Gaussian processes on $\Re$, it is natural to try to find an analogue to it in higher dimensions as well.

There are a number of routes to go about this, involving either integral geometry or differential topology, the essential details of which are described in [1]. (Would-be experts should look also at [37, 38, 90].) We shall mix and match the two approaches in what follows.

For the moment, let F be a smooth function on $\Re^N$, $N \geq 1$, and let $T \subset \Re^N$ be a set with smooth boundaries. I shall not bother with defining "smooth" for the moment, but, later on, when giving results about Gaussian processes, will give conditions sufficient for everything to be correct.

Our interest centres on the excursion set $A_\lambda(T) = \{t \in T : F(t) \geq \lambda\}$. However, for the moment, A could be any sufficiently nice set in $\Re^N$, so that we can drop the explicit dependence on λ and T, along with the implicit dependence on F. So, let us assume that A is just such a set and that we are looking for an integer-valued functional, say φ, defined on the collection of all nice A,

that will do something like count the number of connected components of A. Thus, at the very least, we must require that

$$\varphi(A) = \begin{cases} 0, & \text{if } A = \varnothing, \\ 1, & \text{if } A \text{ is sphere-like,} \end{cases} \tag{3.1.1}$$

where by "sphere-like" we mean topologically equivalent to an N-sphere, and

$$\varphi(A \cup B) = \varphi(A) + \varphi(B) - \varphi(A \cap B). \tag{3.1.2}$$

An important result of integral geometry states that not only does a functional possessing these two properties exist but it is *uniquely* determined by them. It is known as the *Euler characteristic* (or Euler–Poincaré or Hadwiger characteristic, depending on how one approaches the definition and precisely what the meaning of "nice" is).

Perhaps the most basic definition of the Euler characteristic of a set A is to think of A as a manifold with a simplectic triangulation. Suppose that there are, in total, α_N N-dimensional simplices in the triangulation, which, in total, have α_n, $n = 0, \ldots, N-1$, n-dimensional faces. Then the Euler characteristic of A is given by

$$\phi(A) = \alpha_0 - \alpha_1 + \cdots + (-1)^N \alpha_n. \tag{3.1.3}$$

As always, Spivak [103] is a good place to read more details about this.

More global definitions follow from this. For example, in two dimensions, the Euler characteristic of a set A is simply the number of its connected components minus the number of holes. Thus the two figures in Figure 3 clearly have Euler characteristics of 1 and 0. In three dimensions, the Euler characteristic is given by the number of connected components, minus the number of "handles," plus the number of holes. Thus, for example, the Euler characteristics of a solid ball, an empty sphere and a coffee cup are, respectively, 1, 2 and 0.

In principle, and in practice, it is thus very easy in dimensions 2 and 3 to look at a specific set and compute its Euler characteristic. However, this is one of those unfortunate cases in which what is easy for the human visual system to do quickly and effectively requires a lot more care when mathematicised.

The first formal approach we shall adopt to compute the Euler characteristic is via the following iterative fashion, due essentially to Hadwiger [37, 38].

THEOREM 3.1.1. *The functional φ characterized by* (3.1.1) *and* (3.1.2) *has the following iterative definition for nice sets $A \subset \Re^N$:*

$$\varphi(A) = \begin{cases} \textit{Number of disjoint closed intervals in } A, & \textit{if } N = 1, \\ \sum\{\varphi(A \cap \mathscr{C}_x) - \varphi(A \cap \mathscr{C}_{x^-})\}, & \textit{if } N > 1, \end{cases} \tag{3.1.4}$$

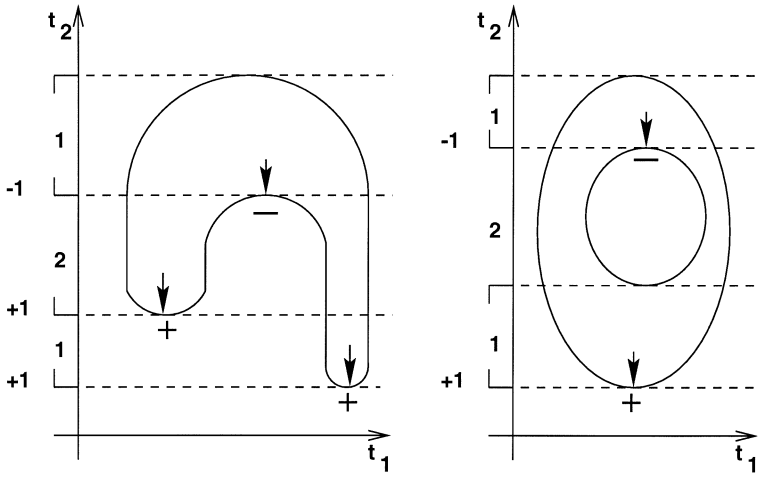

FIG. 3. *Computing the Euler characteristic for a horseshoe and an annulus.*

where

$$\varphi(A \cap \mathscr{C}_{x^-}) = \lim_{y \downarrow 0} \varphi(A \cap \mathscr{C}_{x-y}), \tag{3.1.5}$$

and where $\mathscr{C}_x$ denotes the $(N-1)$ plane of points in $\Re^N$, all of which have their jth coordinate equal to x, $j \in [1, N]$ is arbitrary and the (finite) summation is over all real x for which the summand is nonzero.

Two examples are given in Figure 3. Ignore the arrows for the moment. Here x runs on the vertical, t_2, axis, the regions of constancy of $\varphi(A \cap \mathscr{C}_x)$ along with its value are marked alongside the vertical axis and the contributions to the sum in (3.1.4) are shown to the left of those. In the first case, we have the rather expected result that $\varphi(A) = 1$, and in the second the somewhat less fortunate one that $\varphi(A) = 0$. This means that the Euler characteristic, alone, does not distinguish, in $\Re^2$, among a disk with a hole in it, an empty set, a disk along with another disk containing two holes and so forth. However, despite these apparent drawbacks, the fact that the two basic requirements (3.1.1) and (3.1.2) characterize the Euler characteristic indicate that we have chosen the correct, and *only*, functional for the task at hand.

The iterative formulation of Theorem 3.1.1 is both conceptually and computationally useful for computing Euler characteristics in practice. However, there is a serious problem with it from a probabilistic viewpoint, since if the set A is random it is not at all clear how to go about computing, for example, $E\{\varphi(A)\}$. To get around this, note that rather than going up the $x = t_2$ axis in Figure 3 we could simply count the arrowed turning points on ∂A, counting $+1$ when the boundary is convex, and -1 when it is concave.

If we now specialise to the case in which A is, in fact, an excursion set A_λ of some function F, note that at each point t^* where a +1 is counted, we have

$$F(t^*) = \lambda, \qquad F_1(t^*) = 0, \qquad F_2(t^*) > 0, \qquad F_{11}(t^*) < 0, \tag{3.1.6}$$

whereas the -1's correspond to points t^* for which

$$F(t^*) = \lambda, \qquad F_1(t^*) = 0, \qquad F_2(t^*) > 0, \qquad F_{11}(t^*) > 0, \tag{3.1.7}$$

where, in a notation that will remain fixed for the remainder of the review, first- and second-order partial derivatives are denoted by single and double subscripts, respectively.

It turns out that this technique of counting can actually be extended to higher dimensions. In fact, there is another way to compute the Euler characteristic that is due, in essence, to Morse and developed fully in [68]. For this, we need considerably more notation, but it will be worth the effort. Also, we can easily be more rigorous and complete than before.

Let $f(t)$, $t \in \Re^N$, be a real-valued function of class C^2 on an open subset of $\Re^N$. A *critical point* t^* [i.e., a point where $f_i(t^*) = 0$, $i = 1, \ldots, N$] will be called *ND* (*nondegenerate*) if the Jacobian

$$\begin{aligned} \frac{D(f_1, f_2, \ldots, f_N)}{D(t_1, t_2, \ldots, t_N)}(t^*) &= \det[\{f_{ij}(t^*)\}_{i,\,j=1}^N] \\ &= \det[\mathrm{Hess}_f(t^*)] \neq 0. \end{aligned} \tag{3.1.8}$$

If each critical value of f is ND, f itself will be termed ND. The *index* of a critical point t^* of f is the number of negative eigenvalues of the matrix $((f_{ij})(t^*))_{i,\,j=1,\ldots,N}$, counted with their multiplicities.

Now let $A \in \Re^N$ be a compact C^2 domain; that is, its boundary ∂A is an $(N-1)$-dimensional C^2 manifold. There is no need to assume that either ∂A or A is connected, but we do insist that they have a finite number of components. Then f is called *admissible* relative to A if f is of class C^2 on an open neighbourhood of A, if f has no critical points on ∂A and if the restrictions of f to A and ∂A, $f|_A$ and $f|_{\partial A}$, are both ND.

Finally, let ∂A^- denote the submanifold of points $t \in \partial A$ for which the directional derivative of f in the direction of the outward normal to ∂A at t is negative, and define the k-th *type numbers* of $f|_A$ and $f|_{\partial A^-}$, $\{m_k\}_{k=0,1,\ldots,N}$ and $\{m'_k\}_{k=0,\,1,\ldots,N-1}$, as the number of critical points of $f|_A$ and $f|_{\partial A^-}$, respectively, of index k. These are all finite if f is admissible relative to A. The following important result is due to Morse and Cairns ([68], Theorem 10.2′) and is of major importance both to differential topology and to us.

THEOREM 3.1.2 (Morse's theorem). *Let $f(t)$, $t \in \Re^N$, be a real-valued function of class C^2, admissible relative to a compact C^2 domain $A \in \Re^N$ with C^2 boundary and a finite number of components. Then the Euler characteristic*

of A is given by

$$\varphi(A) = \sum_{k=0}^{N} (-1)^k m_k + \sum_{k=0}^{N-1} (-1)^k m'_k. \tag{3.1.9}$$

The importance of this result for us lies in that it gives a local point set representation of a global topological variable, and expectations of the number of points in sets are often analytically accessible. (The impatient reader can jump ahead to Theorem 3.2.1 to see why.)

There are a number of ways that Theorem 3.1.2 can be immediately applied. For example, take X to be defined over a region T, take $A = A_\lambda(X, T)$ and assume that $A \cap \partial T = \varnothing$. Set the f of the theorem to be identical to X. Then, since f is constant on ∂A, the second sum in (3.1.9) disappears and we have a quite simple characterization of $\varphi(A)$ in terms of the critical points of X.

Under the same setup on X, T and A, and again assuming that $A \cap \partial T = \varnothing$, take f to be the Nth coordinate function, that is,

$$f(t) = f(t_1, \ldots, t_N) = t_N.$$

Then the next theorem follows easily from Theorem 3.1.2.

THEOREM 3.1.3. *Let $F\colon \Re^N \to \Re^1$ be C^2 over a compact $T \subset \Re^N$ and assume that $F(t) < \lambda$ for all $t \in \partial T$. Then, if the χ_k, defined later, are all finite, the Euler characteristic of the excursion set $A_u(F, T)$ is given by*

$$\varphi(A) = (-1)^{N-1} \sum_{k=0}^{N-1} (-1)^k \chi_k, \tag{3.1.10}$$

where χ_k is the number of points $t \in T$ satisfying the following conditions:

$$F(t) = u, \tag{3.1.11}$$

$$F_j(t) = 0, \qquad j = 1, \ldots, N-1, \tag{3.1.12}$$

$$F_N(t) > 0, \tag{3.1.13}$$

$$\text{the index of } D(t) \text{ equals } k, \tag{3.1.14}$$

where $D(t)$ is the $(N-1) \times (N-1)$ matrix of second derivatives with elements $X_{ij}(t)$, $i, j = 1, \ldots, N-1$.

Setting $N = 2$ in this result recovers (3.1.6) and (3.1.7), and so shows how the integral geometric and differential topological approaches to the Euler characteristic coincide.

In fact, (3.1.10) is such a neat and compact formula and, as we shall see later, has such an elegant form for its expectation for a wide variety of random $X \equiv F$ that one is tempted to neglect the fact that excursion sets may occa-

sionally intersect boundaries and use this relationship as an actual definition of a characteristic, as in the following.

DEFINITION 3.1.1. Let $F\colon \Re^N \to \Re^1$ be C^2 over a compact $T \subset \Re^N$. Then, if the χ_k of Theorem 3.1.3 are all finite, the DT (differential topology) characteristic of the excursion set $A_u(F, T)$ is given by

$$\chi(A) = (-1)^{N-1} \sum_{k=0}^{N-1} (-1)^k \chi_k. \tag{3.1.15}$$

The need for a new symbol here comes, of course, from the fact that the Euler and DT characteristics are identical in general only if the excursion set does not intersect ∂T.

Although this definition of an excursion characteristic is somewhat myopic (since it cannot see boundary events), it actually turns out to be quite useful. By small miracles of computation, its expected value can be explicitly computed for Gaussian and related fields and shown to be asymptotically (in λ) equivalent to both the expected number of local maxima of the field above the level λ and the expected Euler characteristic itself. Consequently, it turns out that one can use this formula to study the excursion probability $P\{\sup_{t\in T} X_t > \lambda\}$ in much the same way that Rice's formula was exploited in one dimension. All this was known 20 years ago, and was treated in detail in [1].

However, there was one major drawback to using the DT characteristic rather than the true Euler characteristic: when applied to estimate the excursion probability, it gave, in the notation of (1.1.10), only the first term of the asymptotic expansion. To go higher, we need to look at how to handle the true Euler characteristic, when the excursion set is allowed to intersect ∂T. To do this, consider Figure 4, which is a far more honest portrayal of excursion sets than the previous figures.

First, the parameter space T, in this case the surrounding dumbbell shape, appears. Second, of the three components of the excursion set (the hatched objects) two intersect ∂T. We already know how to characterize the small component in the top left: we count a +1 for each $\bullet$, -1 for for each $\circ$, and sum them. The problem is what to do with the remaining two components.

First, it is important to note that the iterative formulation of Theorem 3.1.1 still works here, without change. What fails is the alternative point set characterization via (3.1.6) and (3.1.7). Morse's Theorem 3.1.2 will also not apply here. The problem is that be ∂T as smooth as it likes, and the same for $\partial\{t\colon F(t) = \lambda\}$, these two boundaries, which together form ∂A, will, in general, intersect at a finite number of points to make ∂A only piecewise smooth.

Applying the approach of Theorem 3.1.1, Worsley [120] showed that the way to compute the Euler characteristic in a general two-dimensional situation with a smooth excursion set A over a set T with a piecewise smooth boundary ∂T was to add the number of $\bullet$'s in Figures 4 and 5 and subtract the number of $\circ$'s. (Figure 5 simply indicates those parts of ∂T which may be parallel, rather

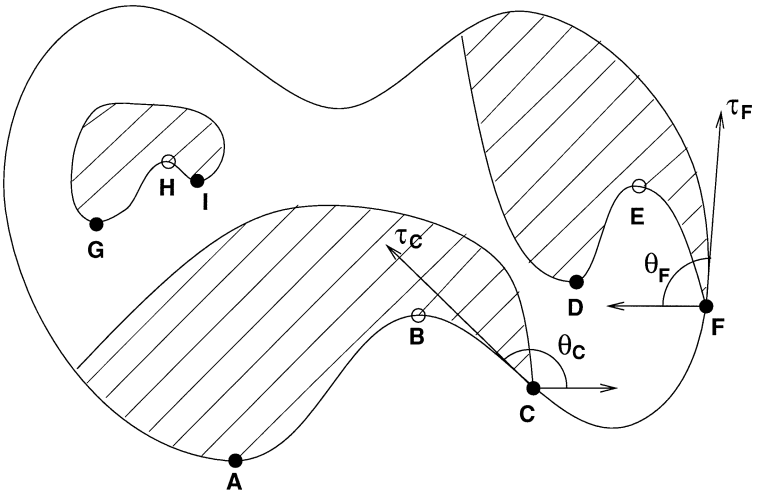

FIG. 4. *Computing the Euler characteristic honestly*: *with boundary.*

than merely tangent, to the horizontal axis.) More formally, these points are counted as follows:

(a) If t is in the interior of T, then apply the criteria (3.1.6) and (3.1.7) exactly as before.

(b) If $t \in \partial T \cap \partial A$ and the tangent to the boundary of ∂T is *not* parallel to the horizontal axis, then let $F_{\text{up}}(t)$ be the derivative of F in the direction of the tangent to ∂T pointing in the positive t_2 direction. (Two such vectors appear as τ_C and τ_F in Figure 4.) Furthermore, take the derivative of F with respect to t_1 in the direction pointing *into* T. Call this $F_\perp$. (It will equal either F_1 or $-F_1$, depending on whether the angles θ in Figure 4 from the horizontal to the τ develop in a counterclockwise or clockwise direction, respectively.) Now mark t as a $\bullet$ (and so count as +1) if $F_\perp(t) < 0$ and $F_{\text{up}}(t) > 0$. There are no $\circ$ points in this class.

(c) If $t \in \partial T \cap \partial A$ and the tangent to the boundary of ∂T is parallel to the horizontal axis, but is not included in an open interval all of which is parallel to this axis, then proceed as in (b), simply defining $F_\perp$ to be F_1 if the tangent is above ∂T and $-F_1$ otherwise.

(d) If $t \in \partial T \cap \partial A$ belongs to an open interval of ∂T that is parallel to the horizontal axis (as in Figure 5), then mark it as a $\bullet$ if T is above ∂T and $F_1(t) < 0$. (Thus, as in Figure 5, points such as B and C by which A "hangs" from ∂T will never count.)

(e) Finally, if $t \in \partial T \cap \partial A$ has not already been marked and coincides with one of the points that contribute to the Euler characteristic of T itself (e.g., A and B in Figure 5), then mark this point exactly as it was marked in computing $\varphi(T)$.

All told, this can be summarized as the following theorem.

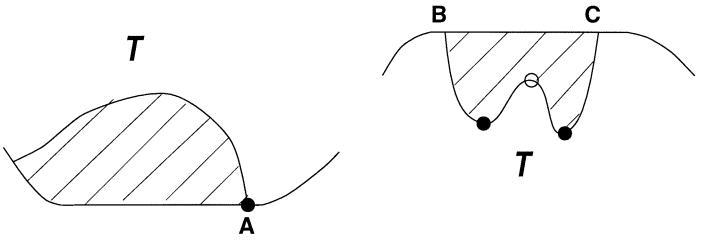

FIG. 5. *Even more honesty/detail.*

THEOREM 3.1.4 (Worsley [120]). *Let F be nice and let $T \subset \Re^2$ have a piecewise smooth boundary. Denote the DT characteristic of $A = A_\lambda(T)$ by $\chi(A)$, computed as in Definition* 3.1.1. *Denote the number of points satisfying conditions* (b)–(d) *as $\chi_{\partial T}$ and denote the sum of the contributions to $\varphi(T)$ of those points described in condition* (e) *by $\varphi_{\partial T}$. Then*

$$\varphi(A) = \chi(A) + \chi_{\partial T} + \varphi_{\partial T}. \tag{3.1.16}$$

Three questions now need to be asked (and answered):

1. Do we really need such a complicated characterization of $\varphi(A)$?
2. What can we do with it?
3. What happens in higher dimensions if we continue along this route?

The answer to the first question is, not surprisingly, "it depends." For example, under the type of conditions under which Theorem 3.1.4 holds, one can also apply the Gauss–Bonnet theorem to obtain

$$2\pi\varphi(A) = \int_{\partial A} \kappa(t)\, dt + \sum_n \beta_n, \tag{3.1.17}$$

where $\kappa(t)$ is the curvature of ∂A at t, dt is the element of arc length on ∂A and the β_n are the angles of rotation of the tangent to ∂A at its vertices.

While (3.1.17) may be no simpler to look at than (3.1.16), the definitions of its components are much briefer. *However*, if one had to compute an Euler characteristic on a *real* excursion set that came from some application, and so was probably derived from grid data, it is clear that the main component of (3.1.17), the integral of the curvature function against arc length, would be essentially uncomputatble, whereas the formula given by (3.1.16) is easily applied.

Perhaps more importantly, however, (3.1.16) leads to a simple approach for computing $E\{\varphi(A)\}$ when A is the excursion set of a nice random field.[10]

[10] I have computed the same expectation via (3.1.17), but the computations were led by knowing what happens in the other case and intrinsically more difficult.

Another reason to prefer the point set approach over a more global differential geometric approach is that it helps to keep separate the contributions to $\varphi(A)$ that come from points in the interior, T°, of T and on its boundary, ∂T. Overall, one would expect that for either large T or high λ (so that the components of the excursion set are small compared to T) there will be far more points within T° than on ∂T. This is, in fact, generally the case. From the point of view of the excursion probability power expansion (1.1.10) that we are seeking, it will turn out that the highest order term is given by the internal points, and the next highest by the boundary points.

The question of what we can do with results like Theorem 3.1.4 is the subject of the remainder of this section (after we see how to compute expectations in the random case) and in the following section where we apply these results to compute excursion probabilities.

Regarding higher dimensions, Theorem 3.1.4 has a version in $\Re^3$, also due to Worsley [120], in which the parameter space T is required to be smooth except for a set of smooth edges, or creases, of finite length and perhaps vertices where the edges meet. There is a formula corresponding to (3.1.16), although now with four rather than three terms on the right-hand side, and the definitions of the contributing points become more complex. They can be found in the original paper. Nevertheless, it is important to note that although it takes quite some time to write them out, the final list of things to check is simple enough to explain to a computer.

What happens in $\Re^N$, $N \geq 4$, is, in general, not as clear. There are no known formulas akin to (3.1.16). Nevertheless, the same procedure, based on the iterative definition of Theorem 3.1.1 and used to handle the cases $N = 2, 3$, should also work there. In practice, for higher dimensions the Euler characteristic is often replaced by the easier to compute DT characteristic. However, in a number of cases, in which the excursion sets are generated by isotropic Gaussian fields and some geometric restrictions are placed on T, it is nevertheless possible to compute the *expected* Euler characteristic (cf. Theorem 3.3.5). We shall return to this in the subsection after the next.

3.2. *Mean values*: *I—DT characteristic.* We will start with a meta-theorem about the expected number of points at which a vector-valued random field takes values in some set, and then describe how to use this to ultimately compute expected Euler characteristics. In this subsection, however, we will treat only the somewhat neater DT characteristic.

Fortunately, we are now finally in a situation where being rigorous is not too time consuming, so we start by setting out some notation and laying down some conditions.

For some $N, K \geq 1$, let $U = (U^1, \ldots, U^N)$ and $V = (V^1, \ldots, V^K)$, respectively, be $\Re^N$- and $\Re^K$- valued N-parameter random fields. We need two sets, $T \subset \Re^N$ and $B \subset \Re^K$. Here T is assumed compact, and ∂T most have zero N-dimensional measure. Furthermore, we require that every $t \in T^\circ$ have a convex neighbourhood completely contained in T. We write $\nabla U(t)$ to denote

the $N \times N$ matrix of first-order partial derivatives of U at t, that is,

$$\nabla U(t) = \left(\frac{\partial U^i(t)}{\partial t_j} \right)_{i,\, j=1,\ldots,N},$$

and denote the modulus of continuity of a real-valued function F by

$$\omega_F(h) := \sup_{\|t-s\| \le h} \left| F(t) - F(s) \right|, \qquad h > 0.$$

THEOREM 3.2.1. *Let U, V, T and B be as before and suppose that the following conditions are satisfied*:

(a) *All components of U, ∇U and V are continuous with probability* 1 *and have finite variances* (*over T*).

(b) *For all $t \in T$, the marginal densities $p_t(u)$ of $U(t)$* (*implicitly assumed to exist*) *are continuous in each of their N parameters.*

(c) *The conditional marginal densities $p_t(u|v)$ of $U(t)$ given $V(t) = v$* (*implicitly assumed to exist*) *are bounded above, uniformly in $t \in T$.*

(d) *The moduli of continuity of each of the components of U, ∇U and V satisfy*

$$P\{\, \omega(h) > \varepsilon \,\} = o\left(h^N\right) \quad as \quad h \downarrow 0 \tag{3.2.1}$$

for any $\varepsilon > 0$.

Then, if $N(U, V : B)$ denotes the number of points in T for which $U(t) = u \in \Re^N$ and $V(t) \in B$ and $p_t(u, \nabla u, v)$ denotes the joint density of $(U_t, \nabla U_t, V_t)$ [*with ∇u written as an $N(N-1)/2$-dimensional vector*], *we have, with $D = N(N-1)/2 + K$,*

$$E\{N(U, V : B)\} = \int_T \int_{\Re^D} \left| \det \nabla u \right| \cdot I_B(v)\, p_t(u, \nabla u, v)\, d(\nabla u)\, dv\, dt. \tag{3.2.2}$$

Alternatively, it is sometimes more convenient to write this as

$$\begin{aligned} &E\{N(U, V : B)\} \\ &\quad = \int_T E\left\{ \left| \det \nabla U(t) \right| \cdot I_B(V(t)) \,\middle|\, U(t) = u \right\} p_t(u)\, dt, \end{aligned} \tag{3.2.3}$$

where $p_t(u)$ now is the density of $U(t)$.

In the form stated, there is no published proof for this result. However, if one rewrites the proof of Theorem 5.2.2 of [1] in terms of the notation of Theorem 5.1.1 given there or in terms of the above, one can obtain a full proof.

To see how to apply this theorem, we start with the expected DT characteristic of (3.1.15). This leads to the following, in which we also introduce $|T|$ to denote the N-dimensional Lebesgue measure of T.

THEOREM 3.2.2. *Let $X(t)$ be a zero-mean, homogeneous Gaussian random field on $\Re^N$ and let $T \subset \Re^N$ have the properties described at the beginning of this subsection. Suppose that X has almost surely continuous partial derivatives of up to second order with finite variances, that the joint distribution of X and these partial derivatives is nondegenerate and that the moduli of continuity of the X_{ij} satisfy*

$$P\left\{\max_{i,\,j}\ \omega_{ij}(h) > \varepsilon\right\} \ = \ o(h^N) \quad \text{as } h \downarrow 0. \tag{3.2.4}$$

Then the mean value of the DT characteristic of the excursion set $A = A_\lambda(X, T)$ is given by

$$E\{\chi(A)\} = |T|\,\rho_N(\lambda), \tag{3.2.5}$$

where

$$\rho_N(\lambda) \ := \ \frac{\exp(-\lambda^2/2\sigma^2)(\det\Lambda)^{1/2}}{(2\pi)^{(N+1)/2}\sigma^N} H_{N-1}\left(\frac{\lambda}{\sigma}\right) \tag{3.2.6}$$

and

$$H_n(x) \ = \ n! \sum_{j=0}^{n} \frac{(-1)^j\, x^{n-2j}}{j!(n-2j)!2^j}$$

is the n-th Hermite polynomial, $\sigma^2 = E\{X^2(t)\}$ and Λ is the covariance matrix of the $X_j(t)$ (cf. Section 1.2).

Before I comment on the proof of Theorem 3.2.2, let us look at the theorem itself. The first, most noticeable fact is that (3.2.5) is a precise *equality* for all λ, and not an asymptotic result true only for large λ, which is all that is generally known for high-dimensional Gaussian fields.

It is instructive to see what (3.2.5) looks like in special cases. Setting $T = [0, 1]$ and $N = 1$, the formula reduces to

$$E\{\chi(t \in [0,1]\colon X(t) \geq \lambda)\} = \frac{1}{2\pi}\left(\frac{\lambda_2}{\sigma^2}\right)^{1/2} \exp\left(-\frac{\lambda^2}{2\sigma^2}\right), \tag{3.2.7}$$

where $\lambda_2 = E\{|dX(t)/dt|^2\}$. This is Rice's formula (1.1.5), and so we have, via the DT characteristic, a direct generalization of this basic result.

When $N = 2$ and $T = [0, 1]^2$ is the unit square, we have

$$E\left\{\chi\left(A_\lambda([0,1]^2)\right)\right\} \ = \ (2\pi)^{-3/2}|\det\Lambda|^{1/2}\sigma^{-3}\lambda \exp\left(-\frac{\lambda^2}{2\sigma^2}\right), \tag{3.2.8}$$

while if $N = 3$ and $T = [0, 1]^3$ the mean value of χ is given by

$$(2\pi)^{-2}|\det\Lambda|^{1/2}\sigma^{-5} \exp\left(-\frac{\lambda^2}{2\sigma^2}\right) \cdot (\lambda^2 - \sigma^2). \tag{3.2.9}$$

Note that in the two-dimensional case $E\{\chi(A_\lambda)\} = 0$ for $\lambda = 0$, while $\lambda = \sigma$ is the critical level in the three-dimensional case. The case $N = 2$ is easy to

understand: because of the distributional symmetry of Gaussian fields, at the mean level there are, on average, as many components to the excursion set as holes. Hence the mean of 0. The case $N = 3$ requires some more thought, because of the lack of a simple symmetry for three-dimensional objects.

In general, the mean value of χ is a polynomial in u with terms of order $(N-1), (N-3), (N-5), \ldots$, multiplied by a negative exponential in λ^2 and some dimension-dependent constants. I shall have more to say on the importance of this form soon.

Finally, a word about the conditions of the theorem. The existence of a.s. continuous, second-order partial derivatives is a real condition, which means that these results only work for smooth processes. The nondegeneracy is a very weak condition. The continuity condition (3.2.4) is easy to check, given that derivatives of the Gaussian fields are again Gaussian, with covariances given by derivatives of the original covariance, and bounds on the Gaussian moduli of continuity are easy to come by (cf., e.g., the examples in Chapter 1 of [2]).

For reasons that will become clear later, stationarity is a crucial assumption. Without it, we will not obtain, and cannot expect, as elegant a result as (3.2.5).

Now we can turn to the proof of Theorem 3.2.2.

PROOF OF THEOREM 3.2.2. I am not going to give details of the proof, particularly since they can all be found in [1]. Let it suffice to say that there is a lot of calculus involved and a lot of fancy Gaussian integration. However, I do want to describe what makes the proof work, so we can evaluate other results and have an idea of how far results like this can be extended.

First, one has to show that the conditions of the theorem suffice to make excursion sets "nice enough" that their DT characteristics are well defined. This is a time-consuming, but not inherently difficult job. See [1].

Next, note that the way to use Theorem 3.2.1 here is, in view of (3.1.11)–(3.1.15), to take $U = (X, X_1, \ldots, X_{N-1})$, $u = (\lambda, 0, \ldots, 0)$ and V to be the vector of length $K = 1 + (N-1)(N-2)/2$ whose first element is X_N and whose remaining elements are a lineal arrangement of the elements of $D = \{X_{ij}\}_{i,j=1}^{N-1}$. The set B is then $[0, \infty) \times \Re^{(N-1)(N-2)/2}$.

All of this should be obvious, except for the way in which B is defined, since the definition of $\chi(A)$ in (3.1.15) involves an alternating sum over subregions of $\Re^{(N-1)(N-2)/2}$. However, the absolute value of the determinant that appears there, together the alternating sign of the determinant over the different regions and the alternating factor of $(-1)^k$, gives the right-hand side of (3.2.2) as the expectation we are seeking.

Now compute, but from (3.2.3) rather than (3.2.2). What makes the computation work is, in part, the Gaussian nature of everything, and, in particular, the fact that because of this the X_i are independent of X and the X_{jk}, a fact, as explained in Section 1.2, that arises from the assumption that X is stationary. This has a *major* simplifying effect on all the computations. (Of course, there is a lot of hard computation here, but the fact is that it is doable.) □

3.3. *Mean values*: *II—Euler characteristics.* We now turn to objects of real interest to us, the mean Euler characteristics. First, we already know that formulas will not be easy to find in full generality, since we have a point set representation for the true Euler characteristic only in dimensions 2 and 3, and a priori we require such representations in order to apply the metatheorem (Theorem 3.2.1). (An exception to this situation will be Theorem 3.3.5 at the end of the subsection.) We start with $N = 2$, and the representation (3.1.16), which now reads as

$$E\{\varphi(A)\} \;=\; E\{\chi(A)\} \,+\, E\{\chi_{\partial T}\} \,+\, E\{\varphi_{\partial T}\}.$$

The first term on the right-hand side was computed in the previous subsection, so consider the second, which involves finding the number of points on ∂T satisfying a number of conditions involving, in general, how the first-order derivatives of X relate to the tangent to ∂T.

This seems at first like a simple computation. Since, in this case, ∂T is a simple (deterministic) curve in $\Re^2$, we can parameterise this curve somehow, say by arc length s, and restrict X to this curve to obtain a single parameter, process $U = X|_{\partial T}$ on ∂T. Now form the vector-valued process $V = (X_1, X_2)$, also restricted to ∂T, and the computation of $E\{\chi_{\partial T}\}$ becomes equivalent to finding the expected number of points on ∂T for which $U = \lambda$ and $V \in B$, where $B \subset \Re^2$ is determined by conditions (b)–(d) of Theorem 3.1.4.

This would seem to be tailor made for Theorem 3.2.1, and, conceptually, it is. In general, however, there are serious problems in performing the integrations, because, along ∂T, the process U (i.e., $X|_{\partial T}$) is not generally stationary. This means that it is no longer uncorrelated with its derivatives, and the integrations become, in general, either impossible or at least so unwieldy that the final answers lead not to closed-form formulas but rather to expressions ultimately requiring numerical integration for their evaluation (cf. Section 13.2 of [24] to see how bad even Rice's formula becomes for a nonstationary process).

There is, however, one simplifying assumption that will solve this. Assume X to be not only stationary, but also *isotropic*, so that $R(t)$ is a function of $\|t\|$ only. Then, in the notation of Figure 4, the angles θ, which are crucial in determining whether or not $V \in B$, are reasonably obviously uniformly distributed over $[0, \pi]$, *independently* of U. This is the key to making the computation possible.

Note also that, in the presence of isotropy, the matrix Λ of second-order spectral moments is of the form $\lambda_2 \mathbb{I}$, where $\mathbb{I}$ is the unit matrix and λ_2 is the variance of the derivative of X in any direction. Thus, with $\rho_N(\lambda)$ as in (3.2.6), but with $\det(\Lambda) = \lambda_2^N$, we obtain the following theorem from (3.2.5).

THEOREM 3.3.1 (Worsley [120]). *Under the conditions of Theorem* 3.2.2 *for* X, *along with isotropy*, *and assuming that the boundary of* $T \subset \Re^2$ *is continuously differentiable except at an at most finite number of points*,

$$E\{\varphi(A_\lambda(T))\} \;=\; |T|\,\rho_2(\lambda) \,+\, \frac{|\partial T|}{2}\,\rho_1(\lambda) \,+\, \varphi(T)\,\Psi\!\left(\frac{\lambda}{\sigma}\right). \tag{3.3.1}$$

The assumption of isotropy also allows one to compute in the three-dimensional situation and to employ Theorem 3.2.1 once again to, this time, obtain the following result.

THEOREM 3.3.2 (Worsley [120]). *Under the conditions of Theorem* 3.2.2 *for* X, *along with isotropy*, *and assuming that the boundary of* $T \subset \Re^3$ *is twice continuously differentiable except perhaps for a set of smooth edges*, *or creases*, *of finite length and also vertices where the edges meet*, *it follows that*

$$\begin{aligned} &E\{\varphi(A_\lambda(T))\} \\ &\quad = |T|\,\rho_3(\lambda) + \frac{|\partial T|}{2}\,\rho_2(\lambda) + \frac{H(\partial T)}{\pi}\,\rho_1(\lambda) + \varphi(T)\,\Psi\!\left(\frac{\lambda}{\sigma}\right), \end{aligned} \tag{3.3.2}$$

where $H(\partial T)$ *is a curvature integral that equals the mean curvature if* ∂T *is smooth*, *and involves all of the mean curvature*, *the lengths of the edges and the angles between them otherwise* (*cf.* [119] *and references therein for details*).

When T is a cube of length side ℓ, (3.3.2) leads to

$$E\{\varphi(A_\lambda(T))\} = \ell^3\rho_3(\lambda) + 3\ell^2\rho_2(\lambda) + 3\ell\rho_1(\lambda) + \Psi\!\left(\frac{\lambda}{\sigma}\right). \tag{3.3.3}$$

The last two theorems seem to indicate the existence of a pattern: a leading term in $|T|\rho_N$ followed by later terms in ρ_k, $k \le N-1$. This can be formalized in a slightly different setting, when the domain T is bounded by a C^2 manifold in $\Re^N$. In this case, we have the following elegant result for which we require a little notation.

For $1 \le k \le N$, let $X_{|k}(t_1, \ldots, t_k) = X(t_1, \ldots, t_k, 0, \ldots, 0)$ be the restriction of X to a k-dimensional affine subspace. For $1 \le k \le N$, let $\tilde\rho_k(\lambda) = \rho_k(X_{|k}, \lambda)$, so that $\tilde\rho_N \equiv \rho_N$, and set $\tilde\rho_0(\lambda) = P\{X(0) \ge \lambda\}$. Finally, for an $n \times n$ matrix M, let $\det_k(M)$ be the sum of all $k \times k$ principal minors of M, so that $\det_N(M) = \det(M)$, $\det_1(M) = \text{trace}(M)$, and define $\det_0(M) = 1$.

THEOREM 3.3.3 (Worsley [121]). *Under the conditions of Theorem* 3.3.2 *and assuming that* $T \subset \Re^N$ *is compact and bounded by a* C^2 *manifold*,

$$\begin{aligned} E\{\varphi(A_\lambda(X, T))\} &= |T|\,\tilde\rho_N(\lambda) \\ &\quad + \sum_{k=0}^{N-1} \left(\frac{\Gamma((N-k)/2)}{2\pi^{(N-k)/2}} \int_{\partial S} \det_{N-1-k}(C(t))\,dt \right) \tilde\rho_k(\lambda), \end{aligned} \tag{3.3.4}$$

where $C(t)$ *is the* $(N-1)\times(N-1)$ *inside curvature matrix of* ∂T *at the point* t.

The proof of the theorem relies on differential geometric and Morse theoretic considerations and results such as Theorems 3.1.2 and 3.1.3.

There is one case in which (3.3.4) is particularly simple to compute, for if $T = B_N(r)$, an N-ball of radius r, then the internal curvature matrix is

constant, and so the integral reduces to $\binom{N-1}{k} r^k$. With $\Lambda = \mathrm{diag}(\lambda_2, \ldots, \lambda_2)$,

$$\omega_k := \frac{\pi^{k/2}}{\Gamma(1+k/2)} \tag{3.3.5}$$

and (3.2.6), a little rearrangement leads to

$$\begin{aligned} &E\{\varphi(A_\lambda(X,T))\} \\ &\quad = \exp\left(\frac{-\lambda^2}{2\sigma^2}\right) \sum_{k=0}^{N-1} \frac{\binom{N}{k} \lambda_2^{(N-k)/2} H_{N-k-1}(\lambda/\sigma)\omega_N r^{N-k}}{\omega_k (2\pi)^{(N-k+1)/2}\sigma^{N-k}} + \Psi\left(\frac{\lambda}{\sigma}\right) \\ &\quad = \exp\left(\frac{-\lambda^2}{2\sigma^2}\right) \sum_{k=0}^{N-1} \frac{\binom{N}{k} \lambda_2^{(N-k)/2} H_{N-k-1}(\lambda/\sigma) W_k(B_N(r))}{\omega_k (2\pi)^{(N-k+1)/2}\sigma^{N-k}} + \Psi\left(\frac{\lambda}{\sigma}\right) \end{aligned} \tag{3.3.6}$$

where W_k is the kth "Minkowski functional." In a moment, we shall see that this special case generates a stepping stone to extensions of Theorem 3.3.3 to a much wider class of parameter sets T.

A word of explanation about Minkowski functionals is now in order. These are functionals originally defined over $\mathscr{K}^N$, the class of convex sets in $\Re^N$, and can be extended in a quite natural fashion to unions and some intersections. There are $N+1$ such functionals, and they essentially characterize the topological properties of sets in $\mathscr{K}^N$. The term $W_0(A)$ denotes the N-dimensional Lebesgue measure (volume) of A. If ∂A is smooth, then $NW_1(A)/2$ is the $(N-1)$-dimensional measure of ∂A (i.e., surface area), and $W_N(A) = \omega_N \varphi(A)$. The intermediate functionals are well defined via various Crofton integrals, or integrals of curvature functions, but have somewhat less immediate interpretations.

If A is a sphere or parallelogram, then the Minkowski functionals are particularly simple. In particular, $W_k(B_N(r)) = \omega_N r^{n-k}$, and so the second line in (3.3.6) follows trivially from the first.

The importance of rewriting (3.3.6) in terms of Minkowski functionals lies in the following result, which is due to Hadwiger [37].[11]

THEOREM 3.3.4 (Hadwiger [37]). *If $F\colon \mathscr{K}^N \to \Re$ is a functional on convex sets in $\Re^N$, invariant under rotation and translation, additive* [*in the sense of* (3.1.2)] *and monotone nonincreasing* [$A \subset B \Rightarrow F(A) \le F(B)$], *then*

$$F(T) = \sum_{k=0}^{N} \alpha_k W_k(T), \tag{3.3.7}$$

where $\alpha_0, \ldots, \alpha_N$ are nonnegative, real constants. If F is monotone nondecreasing, then the α_k are nonpositive.

[11]The only place to find a proof of this result seems to be in the original book, but, unless you have a weakness for classical German, you should turn to Schneider [91] to read about it.

Here is the first "new" result of this paper.[12]

THEOREM 3.3.5. *Let T be obtained as the union of a finite number of convex sets in $\Re^N$. Let X be as in Theorem* 3.2.2, *isotropic, and let* $\Lambda = \operatorname{diag}(\lambda_2, \ldots, \lambda_2)$. *Then*

$$E\{\varphi(A_\lambda(X, T))\}$$

$$= \exp\left(\frac{-\lambda^2}{2\sigma^2}\right) \sum_{k=0}^{N-1} \frac{\binom{N}{k} \lambda_2^{(N-k)/2} H_{N-k-1}(\lambda/\sigma) W_k(T)}{\omega_k (2\pi)^{(N-k+1)/2} \sigma^{N-k}} + \Psi(\lambda/\sigma). \tag{3.3.8}$$

Before turning to the proof, it is worthwhile to note that this result is, of course, identical to those in Theorems 3.3.1–3.3.4, when all the conditions hold. The difference in form is that in those examples the Minkowski functionals have been explicitly computed.

PROOF. Let T be as in the theorem, but convex. Fix λ and set $F(T) = E\{\varphi(A_\lambda(X, T))\}$. The homogeneity of X makes F translation invariant, and invariance under rotation follows from isotropy. Since the Euler characteristic itself is additive, the same is true of its expected value, F. Furthermore, for any fixed λ, $E\{\varphi(A_\lambda(X, T))\}$ is either always positive or always negative (or 0), and so monotonicity follows from homogeneity. Consequently, F has a representation of the form (3.3.7). All that remains is to identify the constants α_k.

By Theorem 3.3.6 the theorem is true in the specific case of T a unit ball in $\Re^N$. Comparing (3.3.6) with (3.3.7) immediately identifies the α_k as the coefficients of the Minkowski functionals there, and so (3.3.9) and the theorem follow for simple convex sets T. The extension to unions of convex sets follows from the fact that X is stationary implies that $F(T)$ is additive as a function of T. □

These five results—the mean DT characteristic in any dimension for stationary Gaussian processes, the mean Euler characteristic in dimensions 2 and 3, and the preceding elegant extensions of Theorems 3.3.3 and 3.3.5 (assuming isotropy)—are essentially all that is known. However, since most examples are either two or three dimensional, they more or less cover all the important cases.

In a more general situation, it is usually possible, given enough time, energy and, often, numerical integration, to carry out this program for any specific Gaussian process, even the nonstationary, and to compute a mean Euler characteristic. However, it is doubtful that there exist elegant results beyond those just described.

[12]It is really implicit in [121], although it was not recognized as such at the time.

3.4. *Mean Minkowski functionals.* We came to the Euler characteristic of excursion sets via the two requirements (3.1.2) (additivity) and (3.1.1) [the normalization that $\varphi(\text{sphere-like set}) = 1$]. In fact, additivity holds for all the Minkowski functionals, and, since these, together with the Euler characteristic, characterize the basic topological properties of excursion sets, a natural question is why not generalise all that we have done so far for mean Euler characteristic values to mean Minkowski functionals.

In fact, some (nonrigorous) work has been done in this area, at least for a DT-like version of Minkowski functionals[13] and appears in [114].

There is an inherent difficulty, however, in explicitly evaluating the expectations of Minkowski functionals in dimensions greater than 2. This fact, combined with (or perhaps leading to) a lack of immediate applications, has made the study of Minkowski functionals in the random field setting a rather academic question, and there has not been as much attention paid to this problem as one might have initially imagined.

3.5. *Non-Gaussian processes*: *I.* The computation of expected Euler characteristics has also been carried out for a number of non-Gaussian random fields as well, although all have some sort of relation to the Gaussian. An archetypical example is the so-called χ^2 random field, which is built as follows.

Let $X^1(t), \dots, X^n(t)$ be n independent, zero-mean, homogeneous, real-valued, Gaussian fields on $\Re^N$ with identical covariance function, $R(t)$, and with variance $\sigma^2 = R(0)$. From these we define a process $Y(t)$ by setting

$$Y(t) = [X^1(t)]^2 + \cdots + [X^n(t)]^2.$$

The univariate density for $Y(t)$ is therefore that of a scaled χ^2 random variable with n degrees of freedom, from whence the name of the process. Under the usual sort of smoothness and nondegeneracy conditions (see [1] for details), the mean value of the DT characteristic of the excursion set $A_u = A_u(Y, T)$ in two dimensions is given by

$$\frac{\lambda^{(n-2)/2}|\Lambda|^{1/2}}{2^{n/2}\pi\sigma^n\Gamma(\frac{1}{2}n)}\left[\frac{\lambda}{\sigma^2} - (n-1)\right]\exp\left(\frac{-\lambda}{2\sigma^2}\right), \tag{3.5.1}$$

where Λ is the usual covariance matrix of the first-order derivatives of the X^i.

The basis for the passage from the general integral representations (3.3.2) or (3.3.4) to (3.5.1) is not that different from the Gaussian case, and relies on the fact that the partial derivatives of Y are given by

$$Y_j(t) = 2\sum_{i=1}^{n} X^i(t)X^i_j(t), \tag{3.5.2}$$

$$Y_{jk}(t) = 2\sum_{i=1}^{n} X^i_j(t)X^i_k(t) + 2\sum_{i=1}^{n} X^i(t)X^i_{jk}(t). \tag{3.5.3}$$

[13]This work has been done with astrophysical applications in mind, and so the DT approach is reasonable, given that the volume of the Universe is large compared to its boundary.

Because of (1.2.5) and (1.2.6) and their implications of independence in the Gaussian case, there are all kinds of interesting conditional (in)dependencies hidden in these formula, many very reminiscient of the Gaussian case. For example, conditioning on $X^i = x^i$, $i = 1, \ldots, n$, gives the Y_j a zero-mean normal distribution with variance $4\lambda_{jj}\sum(x^i)^2 = 4\lambda_{jj}Y$. Conditioning cleverly, computing by efficiently exploiting simpler conditional distributions and then ultimately lifting the conditioning is what makes the computation possible.

Similar kinds of arguments have been used in a number of other cases as well. The basic approach is generally the same, but the devil is in the details.

Mean Euler characteristics for general χ^2 fields, as well as F and t fields, have been computed by Worsley in [119]. Cao and Worsley study Hotelling's T^2 and "correlation" fields in [18, 19]. Detailed proofs are given in all of these papers.

I will have more to say about the non-Gaussian situation in Section 7.2.

3.6. *Back to applications.* I want to look at only one example here, that of the COBE radiation data described in the Introduction. All others that I know of are similar in spirit, although they occasionally vary a great deal in the details.

To recap, a null hypothesis in this case would be that background data are essentially unstructured noise. The measured noise is not generally spatially "white," if only because the measuring mechanism involves a smoothing/filtering operation that induces a dependence structure. Assuming Gaussianity, this leaves two parameters to be estimated: the expectation (estimated via a sample mean) and the covariance function (either estimated from the data via one of a number of quite standard techniques or computed by assuming that true radiation is, spatially, purely white and then computing the covariance function from known properties of the filter).

A one-statistic test to determine whether or not this null hypothesis is feasible would now be, as suggested earlier, to compute $\sup_{t\in\text{galaxy}}\text{Intensity}(t)$, as in (1.3.1). Sections 4 and 6 will tell us how to determine excellent approximations to the distribution of this statistic, and so give the critical value for the hypothesis test.

None of the above uses Euler characteristics, beyond the applications, to come, of using mean Euler characteristics to compute excursion probabilities, and so critical values. However, there is a rather interesting way to use them, suggested close to two decades ago in [4].[14]

Figure 6 shows a graph (wiggly line) of the empirical Euler characteristic curve for the COBE data, and its expectation, assuming a Gaussian field with

[14]Another piece of history to reinforce the "wisdom" of Section 1.4: the paper [4], which is probably my most (only?) influential paper (although far, I think, from my best) was published in a "nonprobability" journal by coincidence. If we had published it in one of the standard probability journals, I am certain that the Euler characteristic would never have reached the astrophysics community, and so the brain mapping community, and the impact of the ideas in it and everything else I did in a similar vein but published in the probability literature would have been minimal.

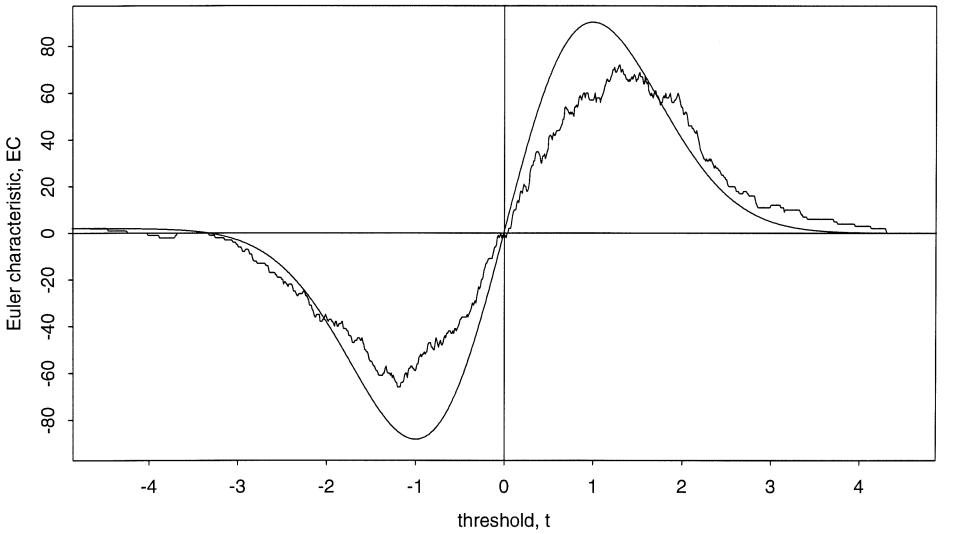

FIG. 6. *Empirical versus theoretical Euler characteristic for the COBE galactic radiation data of Figure* 1.

spectral parameters estimated from the data. Note that they do not really match. While the general shape is similar, this means little, since virtually any model will give more "components" than "holes" at levels above the mean, and the opposite below. Also, since at high and low levels all activity disappears, both ends of the curve tend to 0.

The lack of matching here is then based on two factors: the empirical curve is too low in the central regions, and too high at large, positive levels. One should note that these represent two different phenomena, and both point to the fact that it is unlikely that the data are consistent with a Gaussian model.

It would be nice to have a proper statistical test here, something that would give the distribution of

$$\sup_{-\infty<\lambda<\infty} \frac{\varphi(A_\lambda(X,T)) - E\{\varphi(A_\lambda(X,T))\}}{n(\lambda,T)}, \tag{3.6.1}$$

where n is some appropriate normalizing function, but I know of nothing of this kind beyond one dimension, where the Euler characteristic reduces to the number of level crossings.

While all of the above may seem rather ad hoc, it is a reasonably faithfull description of what is done in practice. The need for further theory is discussed in Section 7.5.

3.7. *Discrete data.* A short digression to keep myself honest. Throughout this paper, random fields are defined *continuously* over subsets of $\Re^N$. Data,

however, are almost always discrete, and generally come on a lattice. That being the case, virtually all the theory has a problem when it comes to applications.

The way around this, already noted in [1], is to use the fact that the Euler characteristic can also be computed via triangulations and alternating sums as in (3.1.3) and to apply this type of definition to the discrete data. One then uses the fact that in a large number of situations the Euler characteristic is continuous in the Hausdorff metric (cf. [37], [38]), and so computation on a fine lattice should yield the Euler characteristic of the true excursion set of the underlying, continuous field.

There are far fewer theorems to justify this procedure than one would like, but the approach seems to work well in practice.

4. Back to maxima. In the preceding section, we spent a lot of time looking at Euler characteristics, which arose as multidimensional generalizations of level crossings. In one dimension, these had a simple relationship with extremal probabilities. However, we have not yet attempted to relate Euler characteristics directly to extremal probabilities.

4.1. *Maxima and the Euler characteristic.* The basic equation

$$P\left\{\sup_{t\in[0,T]} X(t) \geq \lambda\right\} \leq P\{X(0) \geq \lambda\} + TE\{N_\lambda\} \tag{4.1.1}$$

[cf. (1.1.4)] that related extremal probabilities to level crossing rates will not work in higher dimensions, in that the basic inequality that would result from replacing N_λ there by $\varphi(A_\lambda)$ no longer holds. [For example, it is no longer true, in higher dimensions, that $\varphi(A_\lambda) \geq 0$.]

However, an argument that would generalise nicely to higher dimensions would involve the random variable

$$M_\lambda(T) \equiv \#\{t \in T : t \text{ is a local maximum of } X \text{ and } X(t) \geq \lambda\}. \tag{4.1.2}$$

In this case, the same argument that led to (1.1.4) would give

$$P\left\{\sup_{t\in T} X(t) \geq \lambda\right\} \leq E\{M_\lambda(T)\} + P\left\{\sup_{t\in\partial T} X(t) \geq \lambda\right\}. \tag{4.1.3}$$

One expects that, in general, the expectation here is of a higher order (by a power of λ) than the probability of the boundary term, which seems, a priori, to be the harder to calculate. However, if T has a simple structure, then the right-hand side here can be replaced by what might eventually lead to the kind of power series expansions that we are seeking. For example, if T is a rectangle, so that $\partial^k(T) := \partial(\ldots(\partial T))$ is the set of $(N-k)$-dimensional "edges of edges" of T, (4.1.3) leads, by induction, to

$$P\left\{\sup_{t\in T} X(t) \geq \lambda\right\} \leq \sum_{k=0}^{N} E\{M_\lambda(\partial^k T)\}, \tag{4.1.4}$$

where, obviously, $\partial^0 T = T$ and $\partial^N T$ is composed of the 2^N vertices of T.

There is also an easily obtained lower bound for $P\{\sup_{t\in T} X(t) \geq \lambda\}$, due, as far as I know, to Piterbarg. Argue as follows:

$$\left\{\sup_{t\in T} X(t) \geq \lambda\right\} \Leftrightarrow \left\{\sup_{t\in T} X(t) \geq \lambda, \sup_{t\in \partial T} X(t) < \lambda\right\} \cup \left\{\sup_{t\in \partial T} X(t) \geq \lambda\right\}. \tag{4.1.5}$$

But

$$\begin{aligned}
&\left\{\sup_{t\in T} X(t) \geq \lambda, \sup_{t\in \partial T} X(t) < \lambda\right\} \\
&\quad \Leftrightarrow \left\{M_\lambda(T) \geq 1, \sup_{t\in \partial T} X(t) < \lambda\right\} \\
&\quad \Leftrightarrow \left\{M_\lambda(T) = 1, \sup_{t\in \partial T} X(t) < \lambda\right\} \cup \left\{M_\lambda(T) \geq 2, \sup_{t\in \partial T} X(t) < \lambda\right\}.
\end{aligned}$$

To compute the probabilities of the preceding two events, set $p_k = P\{M_\lambda(T) = k\}$, and note that $E\{M_\lambda(T)\} = P\{M_\lambda(T) = 1\} + \sum_{k=2}^\infty k p_k$, so that

$$\begin{aligned}
&P\{M_\lambda(T) = 1, \sup_{t\in \partial T} X(t) < \lambda\} \\
&\quad = P\{M_\lambda(T) = 1\} - P\{M_\lambda(T) = 1, \sup_{t\in \partial T} X(t) > \lambda\} \\
&\quad = E\{M_\lambda(T)\} - \sum_{k=2}^\infty k p_k - P\{M_\lambda(T) = 1, \sup_{t\in \partial T} X(t) > \lambda\},
\end{aligned}$$

while, in a similar vein,

$$P\{M_\lambda(T) \geq 2, \sup_{t\in \partial T} X(t) < \lambda\} = \sum_{k=2}^\infty p_k - P\{M_\lambda(T) \geq 2, \sup_{t\in \partial T} X(t) > \lambda\}.$$

Putting the last two equalities together with (4.1.5) immediately gives

$$\begin{aligned}
P\{\sup_{t\in T} X(t) \geq \lambda\} &= E\{M_\lambda(T)\} - \sum_{k=2}^\infty (k-1) p_k \\
&\quad + P\{M_\lambda(T) = 0, \sup_{t\in \partial T} X(t) > \lambda\} \\
&> E\{M_\lambda(T)\} - E\{M_\lambda(T)[M_\lambda(T) - 1]\}/2,
\end{aligned} \tag{4.1.6}$$

on dropping the boundary term and noting that $k - 1 < k(k-1)/2$ for $k \geq 2$.

We will return to the second factorial moment here later on, but now note that it is reasonable to expect that it will be of lower order than $E\{M_\lambda(T)\}$. Hence everything will now hinge on our ability to compute this expectation for general T. Theorem 3.2.2 implies that, under a simple translation of the conditions there to our current situation,

$$\begin{aligned}
\rho_{\max}(\lambda, T) &:= E\{M_\lambda(T)\} \\
&= \int_T dt \int_\lambda^\infty dx \int_{\Re^{N(N+1)/2}} 1_{\ddot{x}\prec 0} \, |\det \ddot{x}| \, p_t(x,\, 0,\, \ddot{x}) \, d\ddot{x},
\end{aligned} \tag{4.1.7}$$

where, if $\ddot{X}(t)$ is the $N(N+1)/2$-dimensional vector of elements of the matrix $\nabla(\nabla X(t)) = \{X_{ij}\}_{i,j=1,\dots,N}$, then $p_t(x,\ y,\ \ddot{x})$ is the joint density of $(X(t),$ $\nabla X(t),\ \ddot{X}(t))$, and $\ddot{x} \prec 0$ indicates that the matrix corresponding to $\ddot{x}$ is negative definite.

It turns out that there is no way to explicitly evaluate the integral in (4.1.7), but the following rather curious, and extremely useful, result, due to Worsley [121], will help us to derive some asymptotics.

THEOREM 4.1.1. *Let X be a stationary random field, satisfying the conditions of Theorem* 3.2.2. *Set*

$$\rho(\lambda, T) \ = \ (-1)^N \int_T dt \int_\lambda^\infty dx \int_{\Re^{N(N+1)/2}} (\det \ddot{x})\, p_t(x, 0, \ddot{x})\, d\ddot{x}. \tag{4.1.8}$$

[*Note that the indicator and the absolute value sign around the determinant have been dropped from* (4.1.7.)] *Then*

$$\rho(\lambda, T) \ \equiv \ E\{\chi(A_\lambda(T))\}, \tag{4.1.9}$$

where $E\{\chi(A_\lambda(T))\}$ is given by (3.2.5).

The proof of this result is pure (but clever) calculus. There is no sample path analysis involved. A related result, also with a calculus proof, is the following.

THEOREM 4.1.2. *Let X be a stationary random field, satisfying the conditions of Theorem* 3.2.2. *Then, with the notation of the preceding theorem,*

$$\rho_{\max}(\lambda, T) = \rho(\lambda, T) \ + \ o\left(\exp\left(\frac{-\lambda^2}{2\sigma^2}\right) u^{-\alpha}\right) \tag{4.1.10}$$

for all $\alpha > 0$ as $\lambda \to \infty$.

As stated, but with $\alpha = 2 - N$, this result is prehistoric, and, for example, is essentially Theorem 6.3.1 of [1]. It was extended by Breitung [15] to $\alpha = 4 - N$, and, in the much more powerful form that it is given here, for all $\alpha > 0$, it is due to Delmas [26]. The proof, given in full in Delmas' doctoral thesis, remains one of basic calculus, but this time wielded somewhat more cleverly than in the past. [In fact, Delmas has (4.1.10) in more generality than stated here, for she does not even require stationarity. In the more general scenario, however, the coefficients in the expansion of $\rho(\lambda, T)$ are far more complicated, and not always explicitly computable.]

The previous two theorems give us that, up to a surprising level of agreement, there is no difference among $\rho(\lambda, T)$, $\rho_{\max}(\lambda, T)$ and $E\{\chi(A_\lambda(T))\}$, and so each could be used to provide the first term in the upper bound (4.1.4).

Despite the title of this subsection, we have still not mentioned the mean Euler characteristic, $E\{\varphi(A_\lambda(T))\}$. In fact, in those situations in which we know how to compute $E\varphi(A_\lambda(T))$, it always follows from either the same

calculus that leads to Theorem 4.1.2 or from comparison of explicit expressions that

$$E\{\varphi(A_\lambda(T))\} \approx E\{\chi(A_\lambda(T))\} \tag{4.1.11}$$

in the sense that the leading term in λ in each of these expressions is the same. From this it also follows that $E\{\varphi(A_\lambda(T))\}$ is also an approximant to $\rho(\lambda, T)$ and $\rho_{\max}(\lambda, T)$, but now only at the λ^{-1} level of accuracy.

Let us investigate these equivalences by looking a little more closely at $E\{\varphi(A_\lambda(T))\}$ and $E\{\chi(A_\lambda(T))\}$. Write these as

$$\begin{aligned} E\{\varphi(A_\lambda(T))\} &= \left(\sum_{k=0}^{N} a_k^\varphi \lambda^{N-k-1}\right) \exp\left(\frac{-\lambda^2}{2\sigma^2}\right), \\ E\{\chi(A_\lambda(T))\} &= \left(\sum_{k=0}^{N} a_k^\chi \lambda^{N-k-1}\right) \exp\left(\frac{-\lambda^2}{2\sigma^2}\right). \end{aligned} \tag{4.1.12}$$

The equivalences that we discussed previously all imply that $a_0^\varphi = a_0^\chi$, and nothing beyond this. In fact, it is clear that they can say nothing about higher order terms, due to the simple fact that, at least for the odd-ordered terms, the coefficients *must* be different. In particular, we have from (3.2.5) that $a_{2k+1}^\chi \equiv 0$ for all k, while the specific cases in Theorems 3.3.1–3.3.4 all have nonzero coefficients a_k^φ for all k. We shall return to this point later.

Nevertheless, why all these expectations should be so closely related, which is the first clue to understanding why the upper bound can be replaced by an asymptotic equivalence, is the subject of the next subsection.

4.2. *Slepian model processes.* To understand the preceding results, we need to consider how a smooth Gaussian field behaves in the neighbourhood of its local maxima. To do so, we shall limit ourselves, for this subsection, to stationary Gaussian fields.

Suppose we search around for a local maximum (e.g., the one closest to some arbitrary, fixed point) and then affinely translate the coordinate system so that the new origin lies under this maximum.

Then, using an argument that dates back to Slepian (cf. [48], [100]) for Gaussian processes on $\Re$ and to Lindgren [57] for Gaussian fields, it is possible to determine the distribution of the field conditioned on a local maximum (chosen as just described) of height λ at the origin. The distribution takes on a particularly simple form for large λ, and under the assumptions of Theorem 3.2.2, one has that, with probability approaching 1 as $\lambda \to \infty$, the random field has the following representation in the neighborhood of the origin:

$$X(t) = \lambda - \frac{\lambda}{2\sigma} t\Lambda t' + O(1). \tag{4.2.1}$$

[For a more formal and precise version of this result, see either the original paper of Lindgren, [57] or Sections 6.7 and 6.8 of [1]. Chapter 10 of [53] gives a detailed treatment of the one-dimensional case.)

A more general version of (4.2.1) conditions on X taking the value λ at $t = 0$, and with first-order derivatives $\nabla X(0) = (X_1(0), \ldots, X_N(0)) = (d_1, \ldots, d_N) = d$, in which case (4.2.1) becomes

$$X(t) = \lambda + d \cdot t' - \frac{\lambda}{2\sigma} t\Lambda t' + O(1). \tag{4.2.2}$$

In fact, more precise information than either (4.2.1) or (4.2.2) is available, so that one can make some (but not too many) informative statements about the statistical properties of the $O(1)$ term. However, even without this additional information, it is immediately clear from (4.2.1) and (4.2.2) where the equivalences of the previous subsection come from: in the neighbourhood of a high maximum, the random field looks roughly parabolic. Hence each local maximum is associated with a single excursion, and the excursion set that it generates is, up to order λ^{-1}, an N-dimensional ellipse, with axis lengths determined by (4.2.1) and (4.2.2). The full distribution of the axis lengths is not, however, immediately obvious, since this is hidden in the $O(1)$ terms in (4.2.1) and (4.2.2). In the following section, we shall use an approximation technique to determine the expected size of one of these ellipses.

Overall, however, the excursion set A_λ is made up of a number of approximate ellipses, each one of which has Euler characteristic 1 and each one of which is associated with a single maximum above the level λ. Furthermore, since each component of an excursion set is small if λ is large, they are unlikely to be any intersecting ∂T. This explains that asymptotic equivalence of $E\{\chi(A_\lambda(T))\}$ and $E\{M_\lambda(T)\} = \rho_{\max}(\lambda, T)$.

To understand why $\rho_{\max}(\lambda)$ and $\rho(\lambda)$ should be equivalent, compare the integrals (4.1.7) and (4.1.8) defining them. The difference lies in the extra integration over the region for which the Hessian of X is *not* negative definite. However, the previous sample path approximations indicate that, above excursion sets at high levels, there is very little probability associated with this event. Thus the extra region adds little to the integral and hence the asymptotic equivalence.

Why $\rho(\lambda)$ and $E\{\chi(A_\lambda)\}$ should be *identical* seems to be a lot harder to explain geometrically. At this point, it still seems to be a fortunate coincidence.

All of the preceding point to a simple-minded approximation of the structure of a general (high-level, Gaussian) excursion set as a combination of a small number of approximate ellipses. This is the basis behind the Poisson clumping heuristic of David Aldous.

4.3. *The Poisson clumping heuristic.* The philosophy behind the Poisson clumping heuristic (which for the remainder of this section we shall call simply "the heuristic") is the approximate modeling of excursion sets as "mosaic processes." Mosaic processes have two components: a random subset, or "clump," $B \subset \Re^N$, and a Poisson point process on $\Re^N$ with mean measure μ. The mosaic process, A, is then determined by choosing a sequence $B_1, B_2, \ldots$ of iid copies of the random set and a numbering $x_1, x_2, \ldots$ of the points of the Poisson

process, and defining the random set

$$A = \bigcup_k (x_k \oplus B_k), \tag{4.3.1}$$

where $x \oplus B =: \{y \in \Re^N : y = x + z \text{ for some } z \in B\}$. These processes have been studied in detail by Hall [39].

The basic argument behind the heuristic, in our setting, is that excursion sets, above a high level λ, can be modeled as mosaic processes in which the Poisson mean measure μ_λ is given by

$$\mu_\lambda(T) = E\{M_\lambda(T)\} = \rho_{\max}(\lambda, T), \tag{4.3.2}$$

and the basic random set B is a "typical" local excursion. This is not a well-defined object, and so, a fortiori, its precise distribution is not avaliable.

Nevertheless, let us assume that the random sets $B_1, B_2, \ldots$ are small enough, or the points of the Poisson process sparse enough, that the B_j rarely, if ever, overlap. Writing, as usual, $|\cdot|$ for the Lebesgue measure, we then have that, for the random mosaic A,

$$E\{|A \cap T|\} \approx \mu(T) E\{|B|\}. \tag{4.3.3}$$

In essence, this equation is the heuristic. It assumes sparce, Poisson points, small random clumps and lots of independence. Here is a simple example of how to use (4.3.3) in our case.

If X is a stationary Gaussian field and A is the excursion set $A_\lambda([0,1]^N)$, then

$$E\{|A_\lambda|\} = P\{X(0) \geq \lambda\} \approx \frac{\sigma}{\sqrt{2\pi}\lambda} \exp\left(\frac{-\lambda^2}{2\sigma^2}\right). \tag{4.3.4}$$

Since $\mu(S)$ is given by (4.3.2) and (4.1.10) leads to the expression (3.2.5), substituting into (4.3.3) gives the mean size of a single clump as

$$E\{|B|\} \approx \frac{(2\pi)^{N/2}\sigma^{2N}}{\lambda^N \det \Lambda}, \tag{4.3.5}$$

a result which is not directly derivable from either (4.2.1) or (4.2.2), although the heuristic discussion about these results indicates that a result such as (4.3.5) should hold.

Another use of the heuristic, which is not necessary for us at the moment, would be to find, via (4.3.3)–(4.3.5), the mean rate of clumps, or mean number of local maxima, if we knew the mean clump size. We shall return to this more common use soon.

Now we relate all this to extremal probabilities. Let A be the (approximate) random mosaic made up of the clumps in an excursion set. Then, assuming rareness and smallness of clumps, the Poisson approximation gives us, with

μ_λ the Poisson rate measure of excursion clumps above the level λ,

$$\begin{aligned} P\{\sup_T X(t) \geq \lambda\} &= 1 \ - \ P\{A_\lambda(T) = \varnothing\} \\ &= 1 \ - \ P\{\text{no excursion clumps in } T\} \\ &\approx 1 \ - \ \exp\bigl(-\mu_\lambda(T)\bigr) \\ &= 1 \ - \ \exp\bigl(-\rho_{\max}(\lambda, T)\bigr) \\ &\approx \rho_{\max}(\lambda, T) \\ &\approx E\{\chi(A_\lambda(T))\} \\ &\approx E\{\varphi(A_\lambda(T))\}. \end{aligned} \tag{4.3.6}$$

The first approximation follows from the heuristic, and the second from the fact that $e^{-x} \approx x$ for small x. Theorems 4.1.1 and 4.1.2 and (4.1.12) give the last two. Note that while $\rho_{\max}(\lambda, T)$ and $E\{\chi(A_\lambda(T)\}$ are identical up to $o(\lambda^{-\alpha})$ for all α (in terms of the powers of λ) the approximation to $E\{\varphi(A_\lambda(T)\}$ is good only to $o(\lambda^{-1})$, and so, a priori, it hardly seems worthwhile adding this last step. This is particularly true since both of the last final two expectations in (4.3.6) are explicitly computable, that for $E\{\chi(A_\lambda(T)\}$ in general and that for $E\{\varphi(A_\lambda(T)\}$ in a number of cases. Nevertheless, as we shall see soon, the last line generally leads to the best approximation to the first!

Before leaving the heuristic, here is an example, from Cao [17], of how to leapfrog with it.

One can use the Slepian processes of the preceding subsection to get some decent probabilistic information of the volume of a single connected component of the excursion set $A_\lambda(T)$. Suppose we denote the volume of the i-th component by V_i, and let the number of such components be $L = L(\lambda, T)$. Then an interesting statistic in spatial pattern analysis, particularly in problems of medical scanning, is

$$V_{\max} \ =: \ \max_{1 \leq i \leq L} V_i.$$

If V denotes a generic component volume, then the mosaic process underlying the heuristic leads to the approximation in

$$\begin{aligned} P\{V_{\max} \leq v \big| L \geq 1\} &= \sum_{\ell=1}^{\infty} P\{V_i \leq v, \ 1 \leq i \leq \ell, \ L = \ell \ \big| \ L \geq 1\} \\ &= \sum_{\ell=1}^{\infty} P\{L = \ell \ \big| \ L \geq 1\} \ P\{V_i \leq v, \ 1 \leq i \leq \ell \ \big| \ L = \ell\} \\ &\approx \frac{1}{P\{L \geq 1\}} \sum_{\ell=1}^{\infty} \frac{(E\{L\})^\ell}{\ell!} \exp(-E\{L\}) \ (P\{V \leq v\})^\ell \\ &= \frac{\exp(-E\{L\}P\{V \leq v\}) \ - \ \exp(-E\{L\})}{1 - \exp(-E\{L\})}, \end{aligned} \tag{4.3.7}$$

where, as usual, $E\{L\}$ needs to be replaced by one's favourite estimate based on an excursion characteristic of some kind. How reliable are the numbers that finally come out of this kind of leapfrogging is anyone's guess, but they are the best (i.e., basically the only) ones available for this problem.

4.4. *Nonstationary and nonsmooth processes.* Up until now, we have concentrated on stationary processes with sample paths smooth enough to generate excursion sets for which Euler characteristics and their ilk are well defined and have finite expectation.

If we drop the requirement of stationarity, nothing changes very much, *in principle*. However, *in practice*, the computation of expectations often becomes very much more complicated. Consider, for example, the expression (3.2.3) for computing a generic expectation of the number of points in a random set. The first observation is that the spatial, t, integral cannot be ignored and replaced, as in the stationary case, by $|T|$. However, the addition of one more level of integration is a small part of the change.

The major change goes back to (1.2.5) and its consequences (1.2.5) and (1.2.6), which gave that various partial derivatives of stationary processes are uncorrelated. Since, in dealing with Gaussian processes, lack of correlation implies independence, this means that, for example, in any computation involving the joint density $p_t(x, \nabla x, \ddot{x})$ of $(X, \nabla X, \ddot{X})$, we can write, in a self-explanatory notation,

$$\begin{aligned} p_t(x, \nabla x, \ddot{x}) &= p_t(\nabla x)\ p_t(x, \ddot{x}) \\ &= p_t(x)\ p_t(\nabla x)\ p_t(\ddot{x}|x), \end{aligned} \tag{4.4.1}$$

and then even drop the t dependence on each of the densities. Without this simplification, which also occurs, in one form or another, in what we called "Gaussian-related" fields, computations of expectations are usually prohibitively complicated, and, if doable, then only in very special cases. One way around this difficulty is to handle the integration via symbolic computation, and this has been done for a number of cases, using the Maple package, by Shafie in a McGill thesis [93] (cf. also [92]). Shafie gives some interesting examples of rather complicated integrations that actually yield rather neat, compact results, in which the general structure of a result like (3.3.6) can still be seen.

Nevertheless, the passage from mean Euler characteristic to excursion probability, while probably correct, has never been formally justified. Hence, in practice, if not in principle, much of what we have been discussing is safely applicable only for stationary fields.

The second assumption, of sample path smoothness, is even more crucial. In the Gaussian case, for example, mean square differentiability of up to second order ($\Leftrightarrow$ the components of Λ are finite) is a minimal condition for all the expectations we have considered to be finite. Nondifferentiable processes such as the Brownian sheet are even worse, since, then, excursion sets are made up of uncountably many components, and the excursion characteristics with which we have been working cannot be defined.

Nevertheless, the Poisson clumping heuristic still works in this scenario, although in a somewhat different format than we have applied it. Details can be found in Aldous [6], but here is a quick summary of what happens.

Basically, one has to be a little more creative in deciding how to identify the "clumps" of the heuristic, and this is done by allowing a clump to be composed of any number (even uncountably infinite) of components, characterized by the fact that they occur in a small region and the distance to other clumps is comparatively large. One can then develop a theory somewhat parallel to that of the Slepian models of Section 4.2 to obtain information on the size of such a clump. In particular, the mean area $E\{|B|\}$ of a generic clump is computable.[15] Still assuming stationarity and putting this together as in (4.3.3) and (4.3.4), we get

$$\mu_\lambda(T) \approx \frac{P\{X(0) \geq \lambda\}}{E\{|B_\lambda|\}}, \tag{4.4.2}$$

where B_λ is a typical clump of the excursion set A_λ and μ_λ is the corresponding clump rate.

Having established μ_λ in this way, following the first three lines of (4.3.6) gives an approximate excursion probability. Again, the reader should turn to [6] for details.

Note also that if we further complicate the situation by allowing nonstationarity as well as nonsmoothness, the application of the heuristic, while not changing in principle, becomes somewhat more complicated in practice. For example, $E\{|B_\lambda|\}$ is now position dependent, and the simple relationship (4.3.4) becomes

$$E\{|A_\lambda(T)|\} = \int_T P\{X(t) \geq \lambda\}\, dt,$$

which makes the computation of $\mu_\lambda(T)$ in (4.4.2) somewhat less straightforward than in the stationary situation.

4.5. *How accurate are the approximations*: *I*. Throughout all of the discussions we have had so far on approximations, very little has been said on their accuracy.

Results such as Theorem 4.1.2, which relate the various expectations to one another, indicate precision of order $\lambda^{-1}\exp(-\lambda^2/2\sigma^2)$ between them. Since we freely interchanged many of these at various stages, it is hard to believe that the final level of accuracy will be higher than this.

Another compounding factor is the fact that in applying the heuristic we are using a single parameter distribution, Poisson, to approximate something far more complex. Thus, again, it is hard to believe that the level of accuracy involved here would be very high, either.

In general, there is little that one can say beyond this. After all, if one could make precise computations to enable comparisons, then approximate,

[15]There are also many other ways of determining $E\{|B|\}$ for situations similar in spirit but extremely different in detail to the Gaussian situation, in which the Slepian model approach needs to be either significantly changed or avoided altogether. It is the development of these techniques in a wide variety of quite different situations that makes [6] such interesting reading.

heuristic arguments would not be necessary. However, there is one case where more information is available. Slightly extending a result of Pickands [71] and Piterbarg [74], Kratz and Rootzén [50] showed the following.

THEOREM 4.5.1. *Let X be a zero-mean, stationary Gaussian process on $\Re$ with covariance function R satisfying*:

$$\begin{aligned} R(t) &= 1 - \tfrac{1}{2}t^2 + o(t^2), && t \to 0, \\ E\{[X'(t) - X'(0)]^2\} &= 2(R''(t) - R''(0)) \ \le \ C_1 t^2, && t \ge 0, \\ |R(t)| \ + \ (R'(t))^2 &\le C_2 t^{-\alpha}, && t \ge 0, \end{aligned}$$

for some $\alpha > 2$ and finite constants C_j. Then, with $N_\lambda(1)$ indicating the mean number of level crossings of λ in $[0, 1]$,

$$\begin{aligned} 0 &\ge P\{\sup_{[0,T]} X(t) \ge \lambda\} \ - \ (TE\{N_\lambda(1)\} \ - \ P\{X(0) \ge \lambda\}) \\ &\ge -K\left(\lambda^{2(1+1/\alpha)} T \exp\left(\frac{-\lambda^2(1+\delta)}{2\sigma^2}\right) \ + \ \lambda^2 T^2 \exp\left(\frac{-\lambda^2}{\sigma^2}\right)\right), \end{aligned} \tag{4.5.1}$$

where K is an R-dependent constant and

$$\delta \ := \ \inf_{t \ge 0} \left(\frac{1 - R^2(t)}{1 - R^2(t) + R'(t)|R'(t)|} \right).$$

The first inequality in (4.5.1) is merely our original inequality (1.1.4), the proof of which was trivial. The lower bound requires rigorous arguments, at the core of which lies the heuristic.

What is interesting here, however, is the level of accuracy of the approximation. Since $E\{N_\lambda(1)\} = O(\exp(-\lambda^2/2\sigma^2))$ and the boundary term $P\{X(0) \ge \lambda\}$ is of order $\lambda^{-1}\exp(-\lambda^2/2\sigma^2)$, the main implication is that the basic heuristic is correct to $O(\lambda^{-1}\exp(-\lambda^2/2\sigma^2))$, as expected. Adding the boundary correction leads to an extra term of $O(\lambda^{-2}\exp(-\lambda^2/2\sigma^2)$.

Now consider the error term. Both of its components are $o(\exp(-\lambda^2/2\sigma^2))$, which means that the expansion to $O(\lambda^{-2}\exp(-\lambda^2/2\sigma^2))$ is accurate up to an exponential, and not power, level. This is a result of striking practical importance and theoretical elegance, which one would never guess, on the basis of the heuristic, to be true.

Recalling the fact that in one dimension the Euler characteristic of the excursion set $A_\lambda(X, T)$ is simply the number of upcrossings of λ plus 1, if $X(0) \ge \lambda$, and recalling Rice's formula (1.1.5), one can rewrite (4.5.1) in the rather tantalizing form

$$\begin{aligned} 0 &\ge P\{\sup_{[0,T]} X_t \ge \lambda\} \ - \ E\{\phi(A_\lambda(T))\} \\ &\ge -O\left(\lambda^{-2}\exp\left(\frac{-\lambda^2}{\sigma^2}\right)\right), \end{aligned} \tag{4.5.2}$$

which rather makes one wonder what happens in general.

To obtain a general result, for all dimensions, a little more has to be assumed. The conditions are of two forms. The random field is required to be close to isotropic, in that the matrix Λ of second-order spectral moments must be diagonal, a condition automatically satisfied by isotropic fields. Beyond this, we require homogeneity, continuous mean square derivatives of up to third order and mild nondegeneracy conditions akin to those of Theorem 3.2.2. Furthermore, the parameter space T must be convex and rather smooth, in that it looks like a convex polyhedron (up to transformations of class C^3) with neither large nor small angles between adjacent faces. Because of the centrality of Theorem 4.5.2, we will now spell out the details.

Some notation and terminology: a pair $(\mathscr{U}, \alpha)$, where $\mathscr{U} \subset \Re^N$ and $\dim \mathscr{U} = L$, $L \le N$, $\alpha: \mathscr{U} \to \Re^L$ is a diffeomorphism, is called a *map of class* C^ν if:

1. α belongs to C^ν,
2. the set $\alpha\mathscr{U} \subset \Re^L$ is open and bounded,
3. the inverse mapping $\alpha^{-1}: \alpha\mathscr{U} \to \Re^N$ exists and belongs to C^ν.

A one-point subset of $\Re^N$ is also called a map.

A closed set $T \subset \Re^N$ is called a *simple cell complex* of class C^ν if there exists a finite collection of maps $(\mathscr{T}_i, \alpha_i)$, $i = 1, \dots, m$, $\mathscr{T}_i \subset \Re^N$, each of class C^ν, such that $\mathscr{T}_i$ are mutually disjoint and $T = \bigcup_{i=1}^m \mathscr{T}_i$. Any such collection is called a *stratification* of T and denoted by $S(T)$.

Let $\Pi \subset \Re^N$ denote a convex solid angle which is a finite intersection of half-spaces:

$$\Pi = \bigcap \{x: (a_i, x) \ge 0\}, \qquad a_i \in \Re^N.$$

Note that Π is called a *ψ-angle*, $0 < \psi < 1$, if for any k, ℓ either

$$|(a_k, a_\ell)| \le \psi |a_k|\,|a_\ell| \quad \text{or} \quad |(a_k, a_\ell)| = |a_k|\,|a_\ell|,$$

that is the vectors a_k, a_ℓ are either collinear, or the angle between them is bounded by $\arccos(\psi)$. Note also that the dimension of a ψ-angle may be less than N.

DEFINITION 4.5.1. *A simple cell complex $T \subset \Re^N$ is of class $C^3(M, K, \psi, \varepsilon)$, $K > 0$, $M > 0$, $1 > \psi > 0$, $\varepsilon > 0$, if*:

1. *There exists a stratification $S(T)$ such that $\#(S(T)) \le K$, and $\alpha \in C^3(\mathscr{U})$ with norm and derivatives up to order 3 bounded by M for all $(\mathscr{U}, \alpha) \in S(T)$.*
2. *For any $s, t \in T$ such that $|t - s| \le \varepsilon$, there exist a diffeomorphism β of class $C^3(M, B_{2\varepsilon}(t))$, $B_{2\varepsilon}(t) = \{v: |v - t| \le 2\varepsilon\}$, a neighborhood V of the segment $(h\beta(t) + (1-h)\beta(s),\ 0 \le h \le 1)$ and a ψ-angle Π such that $\beta(B_{2\varepsilon}(t) \cap T) \cap V = \Pi \cap V$.*

For example, finitely connected sets with nondegenerate smooth boundaries such as balls or tori, their finite intersections (with some conditions on nondegeneracy at the points of intersections of boundaries) and closed convex

polyhedra with finite numbers of vertices all belong to $C^3(M, K, \psi, \varepsilon)$ for appropriate M, K, ψ, ε.

THEOREM 4.5.2 (Piterbarg [74], [76]). *Let $T \in \Re^N$ be a convex simple cell complex of class $C^3(M, K, \psi, \varepsilon)$ for some M, K, ψ and ε. Let X satisfy regularity conditions of the kind spelled out after* (4.5.2). *In particular, the matrix Λ of second-order spectral moments should be diagonal. Then there exists a constant $\alpha > 1$ such that, for all $\lambda > 0$,*

$$\left| P\left\{\sup_{t\in T} X(t) \geq \lambda\right\} - E\{\varphi(A_\lambda(X, T))\} \right| \leq O\left(\exp\left(\frac{-\alpha\lambda^2}{2}\right)\right), \tag{4.5.3}$$

where, by $E\{\varphi(A_\lambda(X, T))\}$, we refer to the sum on the right hand side of (3.3.8), *which, under the additional conditions on T of Theorem* 3.3.5 *is also the expected Euler characteristic.*

This result is *the* key result in justifying all the discussion of this paper as far as using the expected Euler characteristic to approximate excursion probabilities is concerned. I have found no rigorous way to justify the Poisson clumping heuristic, but for the fact that, in general, it "seems to work" very well. In the cases covered by Theorem 4.5.2, we now know that it works superbly.[16]

In fact, in these cases, we have a better result than we were expecting when we started looking at the asymptotic expansion (1.1.10). For here, not only do we have an expansion of $N + 1$ terms in N dimensions, but we have that it is correct to *exponential* rather than power accuracy. There is at least one additional setting in which something similar can be shown to occur, for which the reader is referred to the following section.

It is reasonably clear that the condition of convexity is rather important for practical applications of this theorem. To see why this should be so, consider an example that Keith Worsley described to me: let T be a "hairy disk," that is, a disk of radius R with a large number of densely sited very thin rectangles, of uniform length L, growing out of it. Then $\sup_T X_t$ over the hairy disk should behave much like $\sup_T X_t$ over a disk of radius $R + L$. However, the Euler characteristic of the excursion sets in the two cases will be very different, since one expects the hairs of the disk to significantly add to it by shattering into smaller pieces (each of Euler characteristic 1) components of the excursion set over $\Re^2$. It would be interesting to see how far Theorem 4.5.2 could be extended.

Given the importance of Theorem 4.5.2, and the fact that, as stated, it both differs somewhat from its formulation in [76] and seems somewhat more general, a few words about its proof are in order.

(ABOUT THE) PROOF. First, note that Piterbarg's Theorem 5.1 in [76] discusses only isotropic processes for which the first derivatives have a diagonal

[16] A fact which, of course, means that in these cases no heuristics are needed.

covariance matrix Λ with elements $1/N$. It is trivial to see that this can be extended to a general matrix by an orthogonal transformation of the parameter space.

Furthermore, Piterbarg's result is not expressed in terms of the expected Euler characteristic as in (4.5.3), but rather via a sum involving Hermite polynomials. A little algebra, and our crucial Theorem 3.3.5, enables the translation.

Now let me describe how Piterbarg proves his result.

He starts with the so-called "cosine" random field

$$X_N(t) \equiv X_N(t_1, \ldots, t_N) := N^{-1/2} \sum_{k=1}^{N} (Z_k \cos t_k + Z'_k \sin t_k), \tag{4.5.4}$$

where the Z_k and Z'_k are independent, standard Gaussians. (Note that the "Λ matrix" in this case is diagonal with elements $1/N$, which is where the assumption mentioned previously comes from.)

When $N = 1$ one can, quite simply, compute the precise distribution of $\sup_{[0,T]} X_1(t)$ and check that it satisfies (4.5.1). (This was actually done in both [74] and [11].) When $N > 1$ the computation works in an iterative fashion, and some simple, but clever, calculus gives

$$\begin{aligned} &P\left\{\sup_{t\in T} X_N(t) \geq \lambda\right\} \\ &\quad = E\{\varphi(A_\lambda(X_N, T))\} + O\left(\exp\left(\frac{-\lambda^2(1+1/N)}{2}\right)\right) \end{aligned} \tag{4.5.5}$$

from which (4.5.3) now follows for this case.

To go from the cosine field to more general fields, we need the following rather powerful Gaussian tool known as Slepian's inequality.

THEOREM 4.5.3 (Slepian [100]). *Let X and Y be two zero-mean Gaussian processes on a general space T such that $EX_t^2 = EY_t^2$ for all $t \in T$ and*

$$E(X_t - X_s)^2 \leq E(Y_t - Y_s)^2. \tag{4.5.6}$$

Then, for all λ,

$$P\left\{\sup_{t\in T} X_t > \lambda\right\} \leq P\left\{\sup_{t\in T} Y_t > \lambda\right\}. \tag{4.5.7}$$

[Note that despite the existence of a trivial heuristic "proof" of Slepian's inequality along the lines that processes that are intrinsically less correlated—as in (4.5.6)—"must" move around more—as in (4.5.7)—the result is false if $\sup_T X$ is replaced by $\sup_T |X|$.]

Returning to our discussion of the proof of Theorem 4.5.2, we now argue as follows: since we want to prove (4.5.3) for smooth processes, and the covariance functions of the X_N are smooth, we can use the X_N to approximate them, at least over small regions. The fact that (4.5.5) holds for X_N allows one to apply

Slepian's inequality[17] to derive (4.5.3) for the processes of the theorem, at least for small enough sets T. Lifting the result to more general T now involves the same kind of integral geometric argument that we used to justify deriving Theorem 3.3.5 from a similar result for spheres. For further details, see [76].

This concludes the discussion on the accuracy of the Poisson approximation and the role that the expected Euler characteristic has to play in it, at least for the moment. Some further comments on the accuracy issue in more general situations, including nonstationary processes, can be found in Section 6.6. While the results there are not as sharp as here, I believe this to be due to the lack of powerful enough techniques for proving sharpness, rather than being due to their inapplicability.

5. Some rigor for the heuristic. First, I should point out once again that I know of no theoretical justification for the heuristic not only in the generality in which Aldous applies it in [6], but even in the Gaussian setup of this section. The main justification is that it works.

Nevertheless, it does have an historical precedent in earlier work on Gaussian (and other) extremes, although the setting is, at least at first glance, somewhat different.

Results dating back to Cramér [22, 23] considered the distribution of the normalized expression

$$\frac{\sup_{[0,T]} X(t) \; - \; b(T)}{a(T)} \tag{5.0.1}$$

as $T \to \infty$, for a smooth Gaussian process X for which $R(t) \to 0$ as $t \to \infty$, generally at some given rate, and appropriate functions a and b. This formulation was in the spirit of Gnedenko's original approach [35] to extreme value theory.

The proofs of these theorems were based on a Poisson argument, that, roughly, went as follows: divide $[0, \; T]$ into $O(1/\alpha_T)$ alternating intervals, long ones of size $O(\alpha_T)$ and short ones of size $o(\alpha_T)$, but such that both their lengths go to ∞ with T. The short ones make up little of the total, so they can be ignored. On the other hand, since they become large with increasing T and since $R(t) \to 0$ as $t \to \infty$, the behavior of X in the intervals of length $O(\alpha_T)$ becomes independent. With b in (5.0.1) also going to ∞ with T, we are in the situation of a large number of almost independent rare events, which is tailor made for a Poisson approximation. Writing the precise formulas for a and b in terms of level crossing rates is what makes the argument look even more familiar.

There is one very big difference, however, between the preceding scenario and that of the heuristic: the heuristic works for all T, and not just for arbi-

[17] Slepian's inequality is, of course, not quite enough, since it is crucial to keep track of the precision of estimates and bounds. This requires more delicate, and rather long, technical arguments, of a type dating back to Berman [9–11].

trarily large parameter sets. Thus there is no asymptotic dependence coming from the decay of correlations.

In the late 1970's and early 1980's, Leadbetter and coworkers[18] realized that it was not necessary to require asymptotic independence of the values of X at distant points. It sufficed to require only that high local maxima tended to occur in an independent fashion. This, in itself, would lead to asymptotic Poisson arguments.

Once one realises this, it is a small, but very significant, conceptual step to develop the Poisson heuristic for Gaussian processes on parameter spaces of fixed size. Aldous' contribution, therefore, was not so much the introduction of a new idea, but the realisation that there were efficient ways of using an old one in a wide variety of different situations, well beyond the Gaussian, many of which were too complicated to analyse rigorously.

5.1. *The "double sum" method.* Since we have spent so much time on heuristics, it is only reasonable that we also briefly describe at least one rigorous way to obtain asymptotic exceedence probabilities.

The description will be very brief, since the topic is treated in detail in both [53] and [76]. In particular, Piterbarg's monograph [76] contains an excellent and remarkably readable introduction on the rigorous approach, which Piterbarg has dubbed the "double-sum" method.

In fact, the technique goes back to Pickand's original 1969 papers [71, 72] on Gaussian exceedence probabilities and has changed little since, although details often vary significantly from case to case.

Motivated by the technique just described for establishing (5.0.1), the basic idea is to break up the parameter space T into a finite union of small sets T_k, where the size of the T_k generally depends on the exceedence level λ. The T_k need not be disjoint, although any overlap should be small in relation to their sizes. There is no need to assume any particular structure for T, so that we are back in the general setting of Section 2. Consequently, the number of T_k required for a fine covering of T is related to its metric entropy in the canonical metric.

It is then elementary that

$$\begin{aligned}\sum_k P\Big\{\sup_{t\in T_k} X(t) \ge \lambda\Big\} &\ge P\Big\{\sup_{t\in T} X(t) \ge \lambda\Big\} \\ &\ge \sum_k P\Big\{\sup_{t\in T_k} X(t) \ge \lambda\Big\} \\ &\quad - \sum\sum_{j\ne k} P\Big\{\sup_{t\in T_j} X(t) \ge \lambda, \sup_{t\in T_k} X(t) \ge \lambda\Big\}. \end{aligned} \tag{5.1.1}$$

[18]See [53] for references and details, in particular, the discussion on the conditions D and D' there.

The upper bound here is treated by looking at each summand, choosing a point $t_k \in T_k$ and writing

$$P\left\{\sup_{t\in T_k} X(t) \geq \lambda\right\} = \int_{-\infty}^{\infty} P\left\{\sup_{t\in T_k} X(t) \geq \lambda \middle| X(t_k) = u\right\} p_{t_k}(u)\, du. \tag{5.1.2}$$

There are a number of ways of treating the probability in the integrand. In particular, since T_k is small, and if X is smooth, it is quite simple to build a linearized version of $\{X\}_{t\in T_k}$ [conditioned on $X(t_k) = u$] and bound the probability in question by the probability that the linearized version exceeds λ somewhere in T_k. In fact, an appropriate approximation can be found even in the nonsmooth case.

In the Gaussian case, the bound on the linearized process is usually taken either from Borell's inequality (2.1.5) or some version of this. Connecting this bound to one on the original process is based on Slepian's inequality, Theorem 4.5.3. Optimising the sizes of the sets T_k with respect to λ gives the best results. Note that at this stage one has an upper bound for the exceedence probability, not much different, in principle, from the elementary bound (1.1.4) that we obtained at the very beginning of the paper.

The first term on the right-hand side of (5.1.1) is handled identically. What remains is to show that the negative term here, the "double sum," is of lower order, and this is the hard part of the argument. This is what turns bounds into asymptotics, and here is where the same considerations that arose in the clumping characteristic arise again.

If we could write the joint probability $P\{\sup_{t\in T_j} X(t) \geq \lambda, \sup_{t\in T_k} X(t) \geq \lambda\}$ as a product of probabilities of the individual events, then we would be done, since, then, the double-sum term is easily seen to be of lower order than the single sum. Such independence obviously does not hold, but if we choose the sizes of the T_k in such a fashion that a "typical" excursion set, or clump, is considerably smaller than this size, and note somehow that high extrema, or clumps, are independent of one another, then we are well on the way to a proof.

The details, which are heavy, are all in Piterbarg [76]. Justifiably, there is no mention there of the clumping heuristic, since Poisson considerations never enter into the rigorous argument. However, it should now be clear how the two are related.

Before turning to a new approach, it is worth noting that Piterbarg's monograph is also an excellent source of worked examples, and includes a number of rigorous computations of excursion probabilities for many interesting examples of processes and fields.[19] With the exception of his version of our Theorem

[19]There is an error in the result of Fatalov [33, 34] reported there for the distribution of the supremum of the N-dimensional Brownian bridge on $[0, 1]^N$ ([76], page 139), which should read $(4 \ln 2)^{N-1} \lambda^{2(N-1)} \exp(-2\lambda^2)/(N-1)!$. [The factor $(N-1)!$ is missing.] The correct result was derived via the Poisson heuristic in a Cornell Ph.D. dissertation by Michael Turmon [115]. In the final analysis, both the heuristic and Fatalov's rigorous argument lead to the same penultimate result, requiring the evaluation of an elementary but tricky N-dimensional integral to obtain a

4.5.2 and (4.5.5) for the cosine field, virtually all of Piterbarg's results deal only with the first (and occasionally second) term of the expansions of excursion probabilities.

5.2. *Processes "without boundary."* There is one other case in which rigorous results have recently been obtained via explicit computation.

Suppose the parameter set T has no boundary, either in a purely geometric sense or in that the random field is periodic in some fashion. In that case, putting together (4.1.4) and (4.1.6), we have

$$\begin{aligned} &E\{M_\lambda(T)\} - E\{M_\lambda(T)[M_\lambda(T)-1]\}/2 \\ &\le P\{\sup_T X(t) \ge \lambda\} \le E\{M_\lambda(T)\}. \end{aligned} \tag{5.2.1}$$

What has changed between this and the previous inequalities of a similar kind is that expectations related to boundary phenomena have now disappeared, and the "correction" so obtained is always in a good direction, in that it lowers the upper bound and raises the lower one.

It is possible, under quite general conditions, requiring neither isotropy nor even stationarity, to bound the factorial moment in (5.2.2) (cf. [26], [77]) and obtain, in this case, that

$$\left| P\{\sup_T X(t) \ge \lambda\} - E\{M_\lambda(T)\} \right| = o\left(\lambda^{-\alpha} \exp\left(\frac{-\lambda^2}{2\sigma_T^2}\right)\right) \tag{5.2.2}$$

for any $\alpha > 0$.

Although we still cannot explicitly compute $E\{M_\lambda(T)\}$, even in this case, it is not hard to see from the formulas in [26] and [77] that it is also true that

$$\left| E\{\phi(A\lambda(T))\} - E\{M_\lambda(T)\} \right| = o\left(\lambda^{-\alpha} \exp\left(\frac{-\lambda^2}{2\sigma_T^2}\right)\right)$$

for any $\alpha > 0$, so that we also have

$$\left| P\{\sup_T X(t) \ge \lambda\} - E\{\phi(A_\lambda(T))\} \right| = o\left(\lambda^{-\alpha} \exp\left(\frac{-\lambda^2}{2\sigma_T^2}\right)\right), \tag{5.2.3}$$

which tells us that the Euler chracteristic is the right route to take in this setting.

---

constant and complete the derivation. Turmon got the integral right, as has since been verified by Fatalov himself.

6. Tubes.

6.1. *Higher order expansions.* As outlined in the Introduction, what we would really like to do would be to identify as many terms as possible of the expansion

$$P\left\{\sup_T X(t) \geq \lambda\right\} = \lambda^\alpha \exp\left(\frac{-\lambda^2}{2\sigma_T^2}\right)\left(C_0 + \sum_{n=1}^{n'} C_n \lambda^{-n}\right) \tag{6.1.1}$$

for $\lambda \geq \lambda_0$ and n' (optimally) finite, or, if need be, infinite.

We have nice, rigorous expansions up to order $n' = N$ for smooth, N-parameter, isotropic Gaussian fields in which the expected Euler characteristic gives the expansion (Theorem 4.5.2), and, as described in the previous subsection, in cases when the parameter set T has no boundary. All indications are that these formulas also hold under far weaker conditions, essentially those that permit computation of the expectation, but there is no proof for this. In any case, such an approach will not work for nonsmooth processes, for which Euler characteristics are not well defined.

In this section, I will describe another approach to the computation of extremal probabilities that is both old (dating back to Hotelling [41] in 1939) and new, having seen a significant regrowth of interest in the past decade or so.

This approach, based on the so-called "tube formulas," seems to have nothing to do with the Poisson approximations of the preceding sections, and, as opposed to them, is highly Gaussian in its basic approach. It does, however, *rigorously* yield information not only on C_0 and C_1 in (6.1.1), but also, in principle, on all the C_n in a wide variety of situations. The qualifier "in principle" is relevant here for two reasons: on the one hand, it will turn out that the C_n for $n \geq 2$ are close to uncomputable in specific examples; on the other hand, truncating the sum in (6.1.1) at $n' = 1$ generally gives an approximation that is both much better than taking only the first term and also adequate for most practical purposes.

To set this up, we shall first of all require a little extra theoretical background on Gaussian processes on Euclidean spaces. Note again the qualifiers: whereas the Poisson clumping heuristic, and similar techniques, yield expansions not only for Gaussian processes on quite general state spaces, but also for *non*-Gaussian processes, the expansions that we are going to derive in this section are obtainable for only Gaussian processes on Euclidean spaces. In principle, these techniques could be extended to more general parameter spaces. However, their essence is Gaussian, and while some extension to other spherically symmetric cases should be possible to date this has only been done in some special cases (cf. Section 7.2).

6.2. *Karhunen–Loève expansions.* Let $T = [0, 1]^N$ and let X be a centered Gaussian process on T with continuous covariance function $R(s, t)$. For the moment, we require neither stationarity nor regularity of R beyond continuity.

The restriction to $[0, 1]^N$ as a parameter space is not, as we shall see later, all that important in principle. However, if T is not a rectangle, the methodology that follows is very hard to put into practice.

Let $\lambda_1 \geq \lambda_2 \geq \cdots$ and $\psi_1, \psi_2, \ldots$ be, respectively, the eigenvalues and normalized eigenfunctions of the operator $\mathscr{R}: L_2(T) \to L_2(T)$ defined by $\mathscr{R}\psi(t) = \int_T R(s,t)\psi(s)\,ds$. That is, the λ_n and ψ_n solve the integral equation

$$\int_T R(s,\, t)\psi(s)\,ds = \lambda\psi(t) \quad \text{for all } t \in T, \tag{6.2.1}$$

with

$$\int_T \psi_n(t)\psi_m(t)\,dt = \begin{cases} 1, & \text{if } n = m, \\ 0, & \text{if } n \neq m. \end{cases} \tag{6.2.2}$$

The following result can be found, for example, in Riesz and Sz-Nagy [85] when $N = 1$ or Zaanen [125] for general N.

THEOREM 6.2.1 (Mercer). *Let R, T, $\{\lambda_n\}_{n\geq 1}$ and $\{\psi_n\}_{n\geq 1}$ be as before. Then*

$$R(s,\, t) = \sum_{n=1}^{\infty} \lambda_n \psi_n(s)\psi_n(t)\,, \tag{6.2.3}$$

where the series converges absolutely and uniformly on $T \times T$.

Mercer's expansion leads to the so-called "Karhunen–Loève" expansion, which is the source of many good things in Gaussian process theory. Two references for this on my desk are Adler [1, 2], but the result is essentially prehistoric and available in any good book on stochastic processes.

THEOREM 6.2.2 (Karhunen–Loève). *Under the conditions of Mercer's theorem, the sum*

$$X_t = \sum_{n=1}^{\infty} \lambda_n^{1/2} \xi_n \psi_n(t), \tag{6.2.4}$$

where the ξ_n are orthonormal Gaussian, converges, uniformly in $t \in T$, in mean square. The sum will also converge, uniformly, with probability 1, if, and only if, X is a.s. continuous.

The mean square convergence of (6.2.4) is rather simple to show. The equivalence between a.s. convergence and continuity is deep.

As a theoretical tool, Theorem 6.2.2 is extremely powerful. As an applied tool, it can be equally so, as long as one has a powerful computer on hand and knows how to numerically solve the eigenvalue equation (6.2.1).

Before going into the numerics, however, we shall give one classic example of the Karhunen–Loève expansion for standard Brownian motion W_t on $[0, 1]$.

Unfortunately, this is one of the very few examples for which a specific solution is available. In the Brownian case,

$$\psi_n(t) = \sqrt{2}\sin\left(\frac{1}{2}(2n+1)\pi t\right), \qquad \lambda_n = \left(\frac{2}{(2n+1)\pi}\right)^2, \tag{6.2.5}$$

so that, with probability 1,

$$W_t = \frac{\sqrt{2}}{\pi}\sum_{n=0}^{\infty}\xi_n\left(\frac{2}{2n+1}\right)\sin\left(\frac{1}{2}(2n+1)\pi t\right), \tag{6.2.6}$$

where $\{\xi_n\}_{n\geq 1}$ is an orthonormal Gaussian sequence.

Now let us see how to exploit the Karhunen–Loève expansion for our purposes.

6.3. *Maxima and areas.* We start by truncating (if necessary) the Karhunen-Loève expansion (6.2.4) after $1 \leq K < \infty$ terms, to define

$$X_t^K = \sum_{n=1}^{K}\lambda_n^{1/2}\xi_n\psi_n(t), \tag{6.3.1}$$

and rearranging as follows:

$$\begin{aligned} X_t^K &= \left(\sum_{k=1}^{K}\xi_k^2\right)^{1/2}\sum_{n=1}^{K}\lambda_n^{1/2}\,\frac{\xi_n}{(\sum_{k=1}^{K}\xi_k^2)^{1/2}}\,\psi_n(t) \\ &\equiv \sqrt{\chi_K^2}\cdot\sum_{n=1}^{K}U_n\phi_n(t), \end{aligned} \tag{6.3.2}$$

where $\chi_K^2 = \sum_{k=1}^{K}\xi_k^2$ and $\phi_n(t) = \psi_n(t)/\chi_K$.

As a first example, consider a unit variance, zero-mean Gaussian process, although not necessarily stationary. In this case, in view of (6.2.3), $\sum_n \phi^2(t) = \sum_n \lambda_n\psi^2(t) = \sigma^2 = 1$ for all $t \in [0, 1]$.

Consequently, the curve $\gamma\colon [0, 1] \to \Re^K$ defined by

$$\gamma(t) \equiv (\phi_1(t), \ldots, \phi_K(t)) \tag{6.3.3}$$

is a curve in the $(K-1)$-dimensional unit sphere, S^{K-1}.

Now note that the K-dimensional random variable

$$U \equiv (U_1, \ldots, U_k) = \left(\frac{\xi_1}{(\sum_{k=1}^{K}\xi_k^2)^{1/2}}, \ldots, \frac{\xi_K}{(\sum_{k=1}^{K}\xi_k^2)^{1/2}}\right) \tag{6.3.4}$$

is actually uniformly distributed on S^{K-1}, so that we can write

$$\begin{aligned} P\left\{\sup_{t\in T}X^K(t) > \lambda\right\} &= \int_0^{\infty} P\left\{\sup_{t\in T}X^K(t) > \lambda \,\middle|\, \chi_K^2 = x\right\} f_K(x)\,dx \\ &= \int_0^{\infty} P\left\{\sup_{t\in T}\langle U, \gamma(t)\rangle > \lambda/\sqrt{x}\right\} f_K(x)\,dx, \end{aligned}$$

where $f_K(x)$ is the (χ^2, with K degrees of freedom) density of χ_K^2.

Equation (6.3.5) is really the key to developing the higher order extremal asymptotics we are looking for. Note that the main contribution of (6.3.5) is that it has translated a high-dimensional Gaussian computation to one involving one χ^2 variate and random variables distributed uniformly over the surfaces of spheres. This uniformity means that the probability computation in the final integrand of (6.3.5) reduces to a computation of surface area on S^{K-1}.

To exploit this approach, we need first to understand the geometric meaning of the object $\{U \in S^{K-1}\colon \sup_{t\in T}\langle U, \gamma(t)\rangle > x\}$ for a general curve γ on S^{K-1}.

6.4. *Weyl's formula.* We are now going to need concepts and terminology from differential geometry, the basics of which we shall use freely and the rest we shall define. Good references here are Gray [36] and Millman and Parker [67].

Suppose $\mathscr{M}$ is an N-dimensional manifold embedded in S^{K-1}. We shall think of $\mathscr{M}$ as the image of a mapping γ, so that

$$\mathscr{M} \;=\; \{\gamma(t) = (\gamma_1(t), \ldots, \gamma_K(t))\colon t \in T \subset \Re^N\}, \tag{6.4.1}$$

where, for the moment, we gain nothing by restricting to the case $T = [0,\ 1]^N$, and so shall not do so.

(For more complete generality, we should really construct $\mathscr{M}$ out of patches, in the usual differential geometric fashion, but will have no need of this for our applications.)

DEFINITION 6.4.1. The tube of geodesic (angular) radius θ of a manifold $\mathscr{M} = \{\gamma(t)\colon t \in T\}$ embedded in S^{K-1} is the set

$$\begin{aligned} T_\theta \;\equiv\; T_\theta(\gamma) \;\equiv\; T_\theta(\mathscr{M}) &= \Big\{x \in S^{K-1}\colon \sup_{t\in T}\langle x, \gamma(t)\rangle \geq \cos\,\theta\Big\} \\ &= \big\{x \in S^{K-1}\colon d(x, \gamma) \leq (2(1-\omega))^{1/2}\big\}, \end{aligned} \tag{6.4.2}$$

where $\omega = \cos\,\theta$ and

$$d(x,\ \gamma) \;:=\; \inf_{t\in T}\|x - \gamma(t)\| \;=\; 2\left(1 - \sup_{t\in T}\langle u,\ \gamma(t)\rangle\right). \tag{6.4.3}$$

The cross-sectional geometry of Figure 7a shows the relationship between the geodesic and Euclidean radii of a tube.

An example ($N = 1,\ K = 3$) is given in Figure 7b. Note that manifolds may have boundaries, as is the case in this example (where the boundaries are the two endpoints of the curve), or could be closed. This difference will be of major importance soon.

Note that the curve in Figure 7b does not curve around to intersect itself in any way. Since some manifolds will do this, however, and this will be a "bad" property, we are going to need some definitions to handle self-intersections.

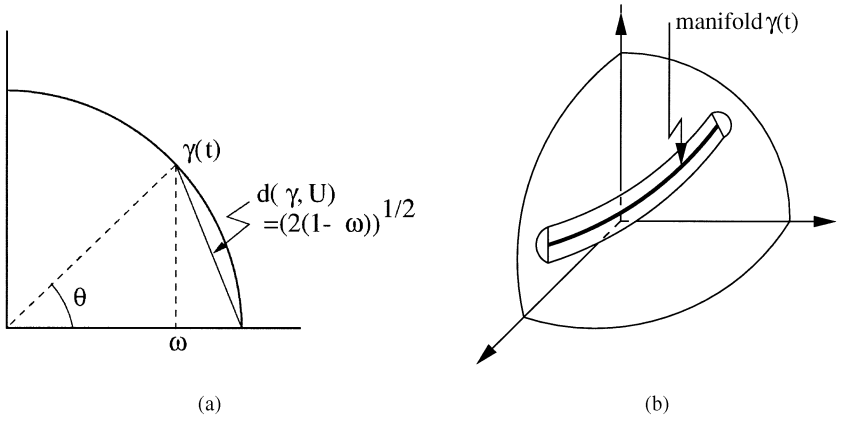

FIG. 7. *Tubes*: (a) *Cross section showing the relationship between geodesic and Euclidean radii*; (b) *a one-dimensional tube on the* 2*-sphere, based on a manifold* (*curve*) *with boundary.*

Thus define the *cross section* of the tube T_θ at the point $\gamma(t)$ on $\mathscr{M}$ as

$$C_\theta(\gamma(t)) = \{x \in T_\theta : (x - \gamma(t)) \perp \nabla\gamma(t)\}. \tag{6.4.4}$$

Each point $x \in T_\theta$ lies in, at least, one cross section, $C_\theta(\gamma(x^*))$, where x^* is chosen from the set $\{y \in \gamma : \|x - y\| = \min_\gamma d(x - y)\}$, and so it is clear that

$$T_\theta = \bigcup_{t \in T} C_\theta(\gamma(t)). \tag{6.4.5}$$

No self-overlap is said to occur if the union in (6.4.5) is disjoint. There are two ways that self-overlapping *can* occur. The cause can be "global," in the sense that $\mathscr{M}$ can "turn around" after a while and come back to close to where it was earlier. [Of course, we are using the terminology of one (i.e., $N = 1$) dimension here, but our meaning, for the general case, should be clear.] This kind of behavior is not determined by local differential properties. Alternatively, the tube radius θ may be large with respect to the local curvature of $\mathscr{M}$ on S^{K-1}. For example, if $\mathscr{M}$ is a great circle on S^2, then the cross sections $C_\theta(x)$ will also lie on great circles, perpendicular to $\mathscr{M}$. They will be disjoint for all $\theta < \pi/2$ will all meet at the poles if $\theta = \pi/2$ and will intersect in the remaining case.

If $\mathscr{M}$ is a curve, then there will be no self-overlap if the geodesic radius of the tube does not exceed the maximum radius of geodesic curvature of $\mathscr{M}$ (cf. [41]).

This leads to the following definition.

DEFINITION 6.4.2. The critical radius of first overlap of a tube $T_\theta(\gamma)$ [which is really an intrinsic property of γ and not $T_\theta(\gamma)$] is

$$\theta_c = \theta_c(\gamma) = \inf\{\theta > 0 : T_\theta(\gamma) \text{ has no self-overlap}\}. \tag{6.4.6}$$

Here is the first most useful and simplest result on tube volumes. It is due to Hotelling [41]. We call a curve $\gamma : [0, T] \to S^{K-1}$ *regular* if it is continuously

differentiable with nowhere vanishing derivative and $a \neq b \Rightarrow \gamma(a) \neq \gamma(b)$, except possibly for the case $a = 0,\ b = T$. In this case, we call the curve *closed* and also demand that $\nabla\gamma(a+) = \nabla\gamma(b-)$.

Denote the volume of the unit ball in $\Re^d$ by ω_d [as in (3.3.5)] and the $(d-1)$-dimensional volume of $\mathcal{S}^{d-1}$ by $\bar{\omega}_{d-1}$, so that

$$\omega_d = \frac{\pi^{d/2}}{\Gamma((d+2)/2)}, \qquad \bar{\omega}_{d-1} = \frac{2\pi^{d/2}}{\Gamma(\frac{1}{2}d)}. \tag{6.4.7}$$

THEOREM 6.4.1 (Hotelling [41]). *Let γ be a regular, closed curve in S^{K-1} of length $|\gamma|$. If $\theta < \theta_c$, then*

$$\begin{aligned}\mathrm{Vol}_{K-1}(T_\theta(\gamma)) &= |\gamma|\,\omega_{K-2}\sin^{K-2}\theta \\ &\quad + (1 - 1_{\{\gamma \text{ closed}\}})\bar{\omega}_{K-2}\int_{\cos\theta}^{1}(1-z^2)^{(K-3)/2}\,dz.\end{aligned} \tag{6.4.8}$$

The integral here is, modulo multiplicative constants, that of a $\beta(1/2, 1/2(K-1))$ density, and thus, while it does not have a simple, explicit form, is easily computed. Note that if the curve γ is closed, the volume is merely the arc length of γ times the volume of a cross-sectional $(K-2)$-dimensional ball of radius $\sin\theta$. When γ has endpoints, the additional volume of two hemispherical caps of dimension $K-1$, subtending angles of θ, must also be incorporated.

Similar results hold also for higher dimensional manifolds as well. However, the only really simple remaining case is for a two-dimensional manifold.

Thus let $\mathcal{M}$ now be a regular, oriented surface embedded in S^{K-1}, which is given locally by a C^3 function $\gamma = \gamma(t) = \gamma(t_1, t_2)$ defined on an open set in $\Re^2$. Assume that the boundary $\partial\mathcal{M}$ is given by a piecewise regular, positively oriented curve parameterized by arc length. Let $|\mathcal{M}|$ and $|\partial\mathcal{M}|$, respectively, denote the surface area of $\mathcal{M}$ and length of $\partial\mathcal{M}$. Then we have the following result.

THEOREM 6.4.2 (Knowles and Siegmund [49]). *With $\mathcal{M}$ as described, assume that the exterior angles at the vertices of $\partial\mathcal{M}$ (if there are any) are positive in the sense that the tangent to $\partial\mathcal{M}$ rotates through a positive angle at each vertex. If $\theta < \theta_c$, then*

$$\begin{aligned}\mathrm{Vol}_{K-1}(T_\theta(\mathcal{M})) &= \frac{\bar{\omega}_{K-4}}{(K-3)}\Big[|\mathcal{M}|\,\cos\theta(\sin\theta)^{K-3} + 2\pi\varphi(\mathcal{M})\int_0^\theta \sin^{K-2}(x)\,dx\Big] \\ &\quad + \frac{\bar{\omega}_{K-3}}{2(K-2)}\,|\partial\mathcal{M}|\,(\sin\theta)^{K-2},\end{aligned} \tag{6.4.9}$$

where $\varphi(\mathcal{M})$ is the Euler characteristic of $\mathcal{M}$.

If $\mathcal{M}$ is a simple, boundaryless manifold of the form of a "belt" around S^{K-1}, then $\varphi(\mathcal{M}) = 0$, and the only difficult term in (6.4.9) disappears.

If one wants to avoid the global topological considerations required to deduce the value of $\varphi(\mathcal{M})$, then the Gauss–Bonnet theorem (cf. [49]) allows one to write

$$2\pi\varphi(\mathcal{M}) = \int_{\mathcal{M}} \kappa \, dA + \int_{\partial\mathcal{M}} k_g \, ds + \sum_i \alpha_i, \tag{6.4.10}$$

where κ is the Gaussian curvature, dA is the element of surface area on $\mathcal{M}$, k_g is the geodesic curvature, ds is the element of arc length on $\partial\mathcal{M}$, and the α_i are the angles of rotation of the tangent to $\partial\mathcal{M}$ at its vertices.

While (6.4.10) may look like an unreasonably complicated way to compute an integral multiple of 2π, there are occasions, as we shall soon see, where such an approach will make a lot of sense. Now, however, consider the general version of the last two results. It is as old as Theorem 6.4.1, and appeared in a companion paper [117] by Weyl to that of Hotelling. Although it has only recently been reapplied to the statistical problem that led to it, it has long been regarded as an important result in differential topology. It holds, however, only for manifolds without boundary.

THEOREM 6.4.3 (Weyl [117]). *Let $\mathcal{M}$ be a C^3, N-dimensional manifold, without boundary, embedded in S^{K-1}. If $\theta < \theta_c$, then*

$$\mathrm{Vol}_{k-1}(T_\theta(\mathcal{M})) = \frac{2\pi^{m/2}}{\Gamma(m/2)} \sum_{n=0}^{2[N/2]} \kappa_{2n} J_{2n}(\theta), \tag{6.4.11}$$

where $[x]$ is the integer part of x, $m = K - N - 1$,

$$J_0(\theta) = \int_0^\theta \sin^{m-1}(x)\cos^N(x)\,dx, \tag{6.4.12}$$

$$J_n(\theta) = \frac{\int_0^\theta \sin^{m+n-1}(x)\cos^{N-n}(x)\,dx}{m(m+2)\cdots(m+n-2)}, \qquad n = 2, 4, \ldots \le N, \tag{6.4.13}$$

and the $\kappa_n = \kappa_n(\mathcal{M})$ are functions of the manifold to be discussed later. If $\theta \ge \theta_c$, then (6.4.13) *still holds, but with "$\le$" replacing the equality.*

It is important to note that Weyl's theorem holds only for manifolds without boundary. The special one- and two-dimensional cases that we have seen in Theorems 6.4.1 and 6.4.2 show that the volume formula is considerably simpler in this case. Presumably, there exists a general formulation of Weyl's theorem for manifolds with boundary, but, as far as I have been able to find, only special cases are known. The comparatively recent text of Gray [36] avoids the issue by defining tubes in such a way that Weyl's formula will be correct even when the manifold has a boundary, essentially by leaving out of the tube sections akin to the semicircular "caps" of Figure 7b. General *upper bounds* have been found by Naiman [69] and Naiman and Wynn [70].

The constants κ_n that appear in (6.4.11) have a variety of geometric interpretations, although Weyl's original description of them as "certain integral

invariants of the surface determined only by ... intrinsic curvature" is, in general, as illuminating as any other. We have already seen the following special cases.

If $N = 1$, then comparing (6.4.11) with (6.4.8) shows that in this simple case of a one-dimensional manifold $\mathscr{M} = \gamma$ we have $\kappa_0 = |\gamma|$.

Similarly, if $N = 2$, then comparing (6.4.11) with (6.4.8) shows that in this case $\kappa_0 = |\mathscr{M}|$ while κ_2 is related to both $|\partial\mathscr{M}|$ and the Euler characteristic $\varphi(\mathscr{M})$. The reasons for this will become clearer later.

As N increases, the geometric significance of the κ_n becomes less and less clear. However, the general pattern remains, in that κ_0 is related to a boundary length, and κ_2 to a "certain integral invariant" of $\mathscr{M}$. The higher order coefficients become somewhat more complex. If the manifold has a boundary, then, as mentioned previously, one can, in principle, compute various correction terms to Weyl's formula. It is important to note, however, that these correction terms can be of quite a significant size, and so should not be thought of as minor perturbations of an otherwise exact result.

Before we leave general tube volume computations, recall why we started them. In the probabilistic motivation of the previous subsection, the manifold $\mathscr{M}$ arose by mapping the parameter set T of a random field onto S^{K-1} via the mapping $t \to (\phi_1(t), \ldots, \phi_K(t))$, where the ϕ_n were the normalized eigenfunctions appearing in the finite Mercer expansion of the covariance kernel R, as in (6.2.3). Since it therefore follows that once R is determined the same is true of $\mathscr{M}$, one might hope that Weyl's theorem could be reformulated in terms of R alone, alleviating the need for intermediate geometric calculations. In part, this is true, and we have the following.

THEOREM 6.4.4 (Sun [107]). *Suppose the manifold $\mathscr{M}$ is defined as in the previous paragraph, where $R: T \subset \Re^N \to \Re$ is smooth (in a sense to be described later), and has constant variance $R(t, t) = 1$. Suppose, furthermore, that $\mathscr{M}$ has no boundary. Then*

$$\kappa_0 = \int_T |\det G(t)|^{1/2}\, dt_1 \cdots dt_N, \tag{6.4.14}$$

$$\kappa_2 = \int_T \frac{1}{2}\left(-S(t) - N(N-1)\right) |\det G(t)|^{1/2}\, dt_1 \cdots dt_N, \tag{6.4.15}$$

where $S(t)$ is the intrinsic scalar curvature of the manifold whose metric tensor matrix is $G = (g_{ij})_{N\times N}$, where

$$g_{ij}(t) = \partial^2 R(s, t)/\partial s_i \partial t_j|_{s=t}. \tag{6.4.16}$$

(See the following discussion for details.) If the underlying random field is stationary, then $S \equiv 0$ and G is independent of t and is actually the matrix of second spectral moments Λ, so that the integrals in (6.4.14) *and* (6.4.15) *reduce to $|\Lambda|^{1/2} \cdot |T|$ and $-\frac{1}{2}N(N-1)|\Lambda|^{1/2} \cdot |T|$, respectively.*

The main point of Theorem 6.4.4 is that one does not need to know any geometry to compute κ_0 and κ_2 when $\mathscr{M}$ does not have a boundary. All one needs to know is how to differentiate and integrate[20]. On the other hand, the values of the κ_n for $n > 2$ seem to be so difficult to compute, that nobody ever does so!

Before we tie all of this back to the distributions of Gaussian suprema, here is how the intrinsic scalar curvature function S is computed. For extra details, see either [51] or [107].

Write $G^{-1}(t) = (g^{ij}(t))_{N\times N}$, where G is the metric tensor matrix (6.4.16). Then the *Christoffel symbols* on $\mathscr{M}$ are defined by

$$\Gamma_{ij}^k = \frac{1}{2}\sum_{l=1}^N g^{lk}(t)\left(\frac{\partial g_{lj}(t)}{\partial t_i} - \frac{\partial g_{ij}(t)}{\partial t_l} + \frac{\partial g_{il}(t)}{\partial t_j}\right). \tag{6.4.17}$$

The *Riemannian curvature tensor* of $\mathscr{M}$ is then defined to be the tensor with components

$$R_{ijk}^l = \frac{\partial \Gamma_{ik}^l(t)}{\partial t_j} - \frac{\partial \Gamma_{ij}^l(t)}{\partial t_k} + \sum_{p=1}^N \left(\Gamma_{ik}^p(t)\Gamma_{pj}^l(t) - \Gamma_{ij}^p(t)\Gamma_{pk}^l(t)\right), \tag{6.4.18}$$

and the *Ricci curvature tensor* is defined to be the tensor with components

$$G_{ij}(t) = \sum_k R_{ijk}^k. \tag{6.4.19}$$

Finally, the scalar curvature function of $\mathscr{M}$ is

$$S(t) = \sum_{ij=1}^N g^{ij}(t)R_{ij}(t). \tag{6.4.20}$$

6.5. *Application to maxima distributions.* We can now return to our main interest, the computation of Gaussian tails, and recall the basic formula (6.3.5), which gave

$$P\left\{\sup_{t\in T} X^K(t) > \lambda\right\} = \int_0^\infty P\left\{\sup_{t\in T}\langle U, \gamma(t)\rangle > \lambda/\sqrt{x}\right\} f_K(x)\,dx, \tag{6.5.1}$$

where $f_K(x)$ was the density of a χ^2 variable (with K degrees of freedom), as the tail probability for a Gaussian process over T with a K-term Karhunen–Loève expansion whose eigenfunctions define γ, as in (6.3.3).

In view of the preceding subsection, we now know how to relate the probability in the integrand in (6.5.1) to a volume computation on the surface of a unit sphere, under the condition that $|\gamma(t)| = R(t, t) = 1$ for all t. [If $|\gamma(t)|$ is some other constant, then a normalization solves everything, and we can work as before. If $|\gamma(t)|$ is actually t-varying, the situation is somewhat more

[20]Or, alternatively, have a computer (or graduate student) for which (whom) symbolic differentiation and numerical integration are straightforward operations.

complex, but still not intractable.] Thus a volume computation, followed by integration over x, yields, in principle, the following theorem.

THEOREM 6.5.1 (Sun [107]). *Let $X(t)$ be a zero-mean, unit variance Gaussian random field over a Borel set $T \subset \Re^N$ with a K-order Karhunen–Loève expansion. If, in the notation of* (6.3.1)–(6.3.3), *the manifold defined by γ has no boundary and if the covariance function $R \in C^3$, then, as $\lambda \to \infty$,*

$$P\left\{\sup_{t\in T} X^K(t) > \lambda\right\} = \kappa_0\psi_0(\lambda) + \kappa_2\psi_2(\lambda) + \cdots + \kappa_{\widehat{N}}\psi_{\widehat{N}}(\lambda) + o(\psi_{\widehat{N}}(\lambda)), \tag{6.5.2}$$

where $\widehat{N} = 2[N/2]$, the κ_n are the same constants as in Weyl's formula and

$$\psi_n(\lambda) = \frac{1}{2^{1+n/2}\pi^{(N+1)/2}} \int_{\lambda^2/2}^{\infty} u^{(N-1-n)/2} \exp(-u)\, du, \qquad n = 0, \ldots, \widehat{N}. \tag{6.5.3}$$

As Sun remarks in her paper, there are a number of difficulties when it comes to applying this theorem. Perhaps the main one is the assumption of the lack of boundary for the manifold generated by γ. We will come back to this point later. Another difficulty arises from the assumed finiteness of the Karhunen–Loève expansion. The following result gets around this, although, in general, at the cost of replacing the equality in (6.5.2) with an upper bound.

THEOREM 6.5.2 (Sun [107]). *Let X be as in Theorem* 6.5.1, *and assume that the set T is an N-dimensional rectangle. Assume that the covariance function $R(s, t)$ of X satisfies Conditions* 1 *(for $m = 6$),* 2 *and* 3 *below. Then, as $\lambda \to \infty$,*

$$P\{\sup_{t\in T} X(t) > \lambda\} \le \kappa_0\psi_0(\lambda) + \kappa_2\psi_2(\lambda) + o(\psi_2(\lambda)), \tag{6.5.4}$$

where the ψ_n are as in (6.5.3) *and κ_0 and κ_2 are computed either as in Weyl's formula when the Karhunen–Loève expansion is finite or via* (6.4.14)–(6.4.16) *in the general case.*

If, furthermore, Condition 4 *is satisfied, then the inequality in* (6.5.4) *can be replaced by an equality.*

The four regularity conditions are as follows.

Condition 1. *One of the following is true for some $m > 0$:*

(i) *There exist functions f and g such that*

$$R(s, t) = g(f(s) - f(t)), \tag{6.5.5}$$

where f, $g \in C^{mN^2}(T)$, g is even and real valued in each of its coordinate(s) and f is a real vector function of full rank.

(ii) *There exist integers* $N_1 < N$, $N_2 = N - N_1$ *and* $N_3 < \infty$ *and functions* f, H_{ij}, h_i *for* $i,\ j = 1, \dots, N_3$ *such that*

$$R(s,t) = \sum_{i,j=1}^{N_3} h_i(s^{(1)})\, h_j(t^{(1)})\, h_{ij}(f(s^{(2)}) - f(t^{(2)})), \tag{6.5.6}$$

where $f, h_{ij} \in C^{mN_2(N_+1)}(T)$, $h_i \in C^4$, *the* h_{ij} *are even and real valued in each of their coordinate*(*s*), f *is a real vector function of full rank and* $s^{(1)} = (s_1, \dots, s_{N_1})$, $s^{(2)} = (s_{N_1+1}, \dots, s_N)$.

Condition 2. *The* $N \times N$ *matrix with components* $R_{ij}(t) = \partial^2 R(s,t)/\partial s_i\, \partial t_j\,|_{s=t}$ *is nonsingular on* T.

Condition 3. *Truncate the Karhunen–Loève expansion of* X *to* K *terms, as in* (6.3.1), *and form the corresponding functions* γ^K *as in* (6.3.3) *and manifolds* $\mathscr{M}^K$ *as in* (6.4.1). *The* $\mathscr{M}^K$ *must have no boundary for all* $K > N$.

Condition 4. *For some* $\theta_0 > 0$, *the critical radii* θ_c^K *of the tubes* $T_\theta(\gamma^K)$ *in* $\mathscr{M}^K$ *satisfy* $\theta_c^K \geq \theta_0$ *for all* $K > N$.

Some words on the conditions of Theorem 6.5.2 are called for. First, one should remember that there is an implicit assumption behind all the conditions that the sample paths are a.s. continuous, so that the Karhunen–Loève expansion converges uniformly, with probability 1.

Condition 1 is easily checked. Stationary processes trivially satisfy (6.5.5), and (6.5.6) covers many other cases. The various levels of differentiability are also easy to check. The real problems arise in checking the assumptions on the manifolds $\mathscr{M}^K$, and this seems, in most cases, to be an impossible task. There are, in fact, two different problems here. The easier one is the issue of lack of boundary. Whether or not this holds is a question of luck, and may not be much harder to check in the "$K = \infty$" case than in the finite case. As an example, consider the case of Brownian motion, W_t, on $[0,\ T]$, for which the Karhunen–Loève expansion is given by (6.2.6) Here the sinusoidal eigenfunctions begin at 0 when $t = 0$ and all end at 1 when $t = 1$, and so the manifolds γ^K have a boundary for all K. This is despite the fact that the full distribution of $\sup W_t$ is comparatively easy to compute (via the reflection principle) from first principles.

The second, often more serious, problem lies in the assumption that $\theta_c^K \geq \theta_0$ for all $K > N$, insofar that one cannot generally expect this to happen. To see why this should be so, think of the simple case of $T = [0,\ 1]$ and note that, under this assumption, there is a precise upper bound, depending only on θ_0, on the number of times $t \in T$ for which $X(t) = X(0)$. However, this random variable is, for most interesting processes X, in principle, unbounded, and so having it bounded (uniformly) for each approximation X^K is unreasonable.

Nevertheless, we shall, in general, now treat the result of Theorem 6.5.2 as if it were an approximation, regardless of our (in)ability to check Conditions 3 and 4.

One comment on the mechanics of the proof of Theorem 6.5.2 is in order, since it is rather important to understand how the application of precise tube formulas leads to the asymptotic version of (6.5.4). The issue lies in the integrand in (6.5.1) for large values of x. In this case, the equivalent tube problem involves a tube of geodesic radius $\cos^{-1}(\lambda/\sqrt{x})$ which, of course, is close to $\pi/2$. This being the case, one cannot expect the tube about $\mathscr{M}$ to have no self-overlap, and so blindly applying a volume formula to estimate the integrand yields to the overestimate of (6.5.4).

On the other hand, for large x, the density $f_K(x)$ is small, and so one could hope that the overestimate, after integration against f_K, would be small, and an approximate equality obtained.

There is also a version of Theorems 6.5.1 and 6.5.2 in the case that the manifold $\mathscr{M}$ does possess a boundary. This will be of major importance to us in what follows.

THEOREM 6.5.3 (Sun [109]). *Under regularity conditions similar to those in the preceding two theorems, as* $\lambda \to \infty$,

$$P\Big\{\sup_{t\in T} X(t) > \lambda\Big\} = \kappa_0\psi_0(\lambda) + \kappa_1\widehat{\psi}_1(\lambda) + (\kappa_2 + C_1 + \kappa_{11})\psi_2(\lambda) + o(\psi_2(\lambda)), \tag{6.5.7}$$

where ψ_0, ψ_2 *and* κ_0, κ_2 *are as before,* C_1 *is related to the curvature of* $\partial\mathscr{M}$ *in the same way that* κ_2 *is related to that of* $\mathscr{M}$, κ_{11} *is related to the rotation angles in the regions of* $\partial^2\mathscr{M}$ *(the boundary of the boundary of* $\mathscr{M}$*) where two faces meet,* $\kappa_1 = |\partial\mathscr{M}|$ *and*

$$\widehat{\psi}_1(\lambda) = \frac{1}{4\pi^{N/2}} \int_{\lambda^2/2}^{\infty} u^{(N-2)/2} \exp(-u)\, du. \tag{6.5.8}$$

The main difference between the boundary case and the nonboundary case is clearly the extra term in $\widehat{\psi}_1$, which leads to an order in the expansion that was not there earlier. It is rather interesting, in fact, that in Theorems 6.5.1 and 6.5.2 the powers of λ in the expansion always dropped by two at a time, much as in the expression (3.2.5) for the mean DT characteristic, and not by one at a time, as for the mean Euler characteristic in, for example, (3.3.8). We shall have more to say about this soon.

It is not always easy to apply Theorem 6.5.3 without knowing a lot of geometry, but if $N = 2$ then $\kappa_0 + \kappa_2 + \kappa_1 + \kappa_{11} = 2\pi\varphi(\mathscr{M})$. Thus, for example, if T is a rectangle, then we recover the result of Knowles and Siegmund[21] [49] (cf. the related Theorem 6.4.2) that

$$P\Big\{\sup_{t\in T} X^K(t) > \lambda\Big\} = \left(\frac{|\mathscr{M}|\lambda}{(2\pi)^{3/2}} + \frac{|\partial\mathscr{M}|}{4\pi} + \frac{1}{\lambda} + o\left(\frac{1}{\lambda}\right)\right)\exp\left(\frac{-\lambda^2}{2}\right)$$

as $\lambda \to \infty$.

[21] See the correction in [98], pages 621–622.

Of course, this has taken us back to geometric computations on $\mathcal{M}$. However, with the usual notation for spectral moments, it is easy to check that, for example, for a stationary process this is equivalent to

$$P\left\{\sup_{t\in T} X^K(t) > \lambda\right\}$$

$$= \left(\frac{|\Lambda|^{1/2}|T|}{\sigma^2(2\pi)^{3/2}}\lambda + \frac{(\lambda_{11}+\lambda_{22})|\partial T|}{8\pi\sigma} + \frac{\sigma}{\lambda} + o\left(\frac{1}{\lambda}\right)\right)\exp\left(\frac{-\lambda^2}{2\sigma^2}\right) \tag{6.5.9}$$

as $\lambda \to \infty$.

6.6. *Tubes versus Euler, or How accurate are the approximations*: *II.* The time has now come to attempt to tie together our two different approaches to Gaussian extrema.

The last formula of the preceding subsection, (6.5.9), gave an expansion for a Gaussian excursion probability which, by Theorem 6.5.3, and in terms of our initial expansion (1.1.10) is correct up to "third order" in λ. This, of course, is far more accurate than we deserve to get from the Poisson clumping heuristic and expected Euler characteristics.

However, (6.5.9) should be reminiscent of Theorem 4.5.2, in which we used the expected Euler characteristic to approximate excursion probabilities for certain isotropic fields. In fact, it is not difficult to compute $E\{\varphi(A_\lambda(X, T))\}$ in the setting of (6.5.9) and to discover that it is identical to the first three terms on the right-hand side. In other words, for this quite common situation, we have a proof that, under the conditions of Theorem 6.5.2, the Euler characteristic/Poisson clumping argument gives an answer far beyond its expected level of precision, without the requirement of isotropy, as in Theorem 4.5.2. On the other hand, we do not have the fine superexponential bounds on the error term that we had in the isotropic case,[22] although I imagine that they still hold, implying that the remainder term in (6.5.7) is really smaller than indicated there.

How common is this situation? Here is a very specific, and rather complicated, example, in which the details have been carefully worked out.

Consider the space–time, zero-mean Gaussian field defined by

$$X(x,\, t) = t^{-N/2} \int_{\Re^N} g\left(\frac{y-x}{t}\right)\, W(dy), \tag{6.6.1}$$

where $x \in C \subset \Re^N$, $0 < t_0 \le t \le t_1 < \infty$, W is a Gaussian white noise on $\Re^N$ [so that $E\{W(A)W(B)\} = |A \cap B|$ for $A,\ B \in \Re^N$] and g is a smooth kernel with $\int g^2(x)\,dx = 1$. This particular $(N+1)$-parameter random field arises in an N-parameter signal testing problem with variable resolution filter[23] studied

[22]Note that using the Euler characteristic, and not something like the DT characteristic, is crucial here, since the higher order terms are quite different for the two. The additional precision comes from the boundary terms inherent in the Euler formulation.

[23]Which is where the extra parameter comes from.

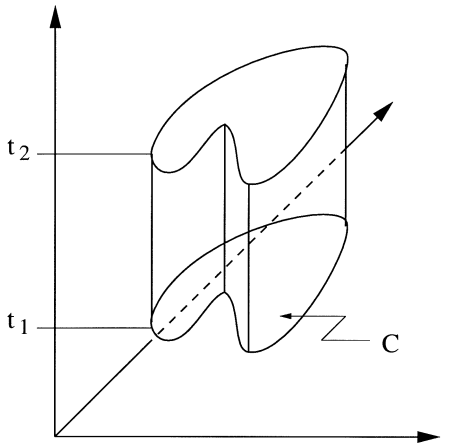

FIG. 8. *The parameter set for the process of* (6.6.1) ($N = 2$).

in Siegmund and Worsley [98], from where all the following calculations are taken and to where the reader should turn for details. The parameter space is shown in Figure 8. Note that C can be quite arbitrary, and, in the general case, need only have piecewise smooth boundary.

It is straightforward to see that the covariance function of X is given by

$$R((x,t),(y,s)) = (s/t)^{-N/2} \int g(z + (x-y)/t) \cdot g((zs)/t)\, dz, \qquad t_0 \le s \le t \le t_1, \tag{6.6.2}$$

so that all the various spectral moments that were required in Section 3 for computed expected Euler characteristics can now be derived with a little calculus and a lot of patience. [To see why "spectral moments" is a sensible term here, rewrite (6.6.3) in terms of a transformed time parameter $\tau = \log t$, and time "differences" will appear in R. If you prefer not to think in terms of spectral moments, think in terms of derivatives of R only.]

With notation a little different from that of the preceding sections, let

$$\Lambda = \int \nabla g(z) \cdot (\nabla g(z))'\, dz, \qquad \kappa = \int \left[z' \cdot \nabla g(z) + \frac{1}{2} N g(z) \right]^2 dz, \tag{6.6.3}$$

so that Λ is, loosely speaking, made up of the spatial second-order spectral moments of X and κ measures the covariance between $X(x, t)$ and $\partial X(x, t)/\partial t$.

With this notation and applying the arguments of Section 3, we have the following result.

THEOREM 6.6.1 (Siegmund and Worsley [98]). *Let $\varphi(A_\lambda)$ be the Euler characteristic of the excursion set of the random field X, defined by* (6.6.1), *over the level λ. Assume that $N = 3$ and that the set C defining the parameter*

space has a twice differentiable boundary ∂C. Assume, furthermore, that X is spatially isotropic, so there exists a positive constant ν for which $\mathbf{\Lambda} = \nu\mathbf{I}$ in (6.6.3). *Then*

$$
\begin{aligned}
E\{\varphi(A_\lambda)\} &= E\{\varphi\left((x,t) \in C \times [t_1,t_2] : X(x,t) \geq \lambda\right)\} \\
&= |C|(t_1^{-3} - t_2^{-3})\,\nu^{3/2}\kappa^{1/2}(\lambda^3/3 - \lambda(1-1/\kappa))\phi(\lambda)/(2\pi)^2 \\
&\quad + (|C|/2)\left(t_1^{-3} + t_2^{-3}\right)\nu^{3/2}\,(\lambda^2-1)\phi(\lambda)/(2\pi)^{3/2} \\
&\quad + (|\partial C|/4)\left(t_1^{-2} - t_2^{-2}\right)\nu\kappa^{1/2}\,(\lambda^2 - 1 + 1/\kappa)\phi(\lambda)/(2\pi)^{3/2} \\
&\quad + (|\partial C|/4)(t_1^{-2} + t_2^{-2})\nu\lambda\phi(\lambda)/(2\pi) \\
&\quad + (H(\partial C)/\pi)(t_1^{-1} - t_2^{-1})(\nu\kappa)^{1/2}\lambda\phi(\lambda)/(2\pi) \\
&\quad + (H(\partial C)/(2\pi))(t_1^{-1} + t_2^{-1})\nu^{1/2}\phi(\lambda)/(2\pi)^{1/2} \\
&\quad + \varphi(C)\log(t_2/t_1)\kappa^{1/2}\phi(\lambda)/(2\pi)^{1/2} \\
&\quad + \varphi(C)\Psi(\lambda),
\end{aligned}
\tag{6.6.4}
$$

where $H(\partial C)$ is the mean curvature of ∂C (e.g., Santaló [90]), *and ϕ and Ψ are, respectively,the standard Gaussian density and tail functions.*

A similar result also holds for piecewise smooth ∂C, but then $H(\partial C)$ also includes terms related to the angles at joins. See [98] *for details, as well as the corresponding formula when $N = 2$, and* [123] *for generalizations to arbitrary dimensions.*

This is a rather long expression, and the only reason I have included it to show how involved it is. Note that the powers of λ here run from λ^3 to λ^{-1}, that is, a five-term expansion. As usual, we could use $E\{\varphi(A_\lambda)\}$, via the Poisson clumping heuristic, to approximate the excursion probability $P\{\sup_{(x,t)\in C\times[t_1,t_2]} X(x,t) \geq \lambda\}$, but without a great deal of information as to how accurate the expansion would be as regards terms beyond the first.

So far there are no surprises. The big surprise comes in that Siegmund and Worsley also made a careful tube formula–based computation of the previous excursion probability, which they found to be identical to $E\{\varphi(A_\lambda)\}$ up to order $o(\lambda^{-1})$. Thus the Euler characteristic/Possion clumping approach is, in this case, accurate up to five $(= N+2)$ terms, one term beyond what we have come to expect.[24]

It is very hard to say how widespread this phenomenon is. Some calculations by Slava Sigal, a Stanford Ph.D. student of Siegmund's, indicate that it may be true for all "nice enough" Gaussian fields over star-shaped domains. However,

[24]And, in the isotropic case, also proven to hold.

no general theory, or even explanation, exists at this point. Nevertheless, we do have the:

PUNCHLINE OF THIS PAPER. *In computing excursion probabilities for smooth enough Gaussian random fields over reasonable enough regions, the expected Euler characteristic of the corresponding excursion sets gives an approximation, for large levels, that is accurate to as many terms as there are in its expansion.*

"PROOF." Theorem 4.5.2 gives this in the smooth, homogeneous, isotropic case. The discussion of Section 5.2 shows that it is true for fields over parameter sets without boundary. The discussion at the beginning of this subsection proves the "Punchline" for smooth processes on a rectangle. Theorem 6.6.1 and the discussion following it provides a proof for a specific process in $\Re^3$. Star-shaped sets have been looked at by Sigal. For the general case, see the challenge of Section 7.1.

6.7. *Statistical applications.* Throughout the discussion on tube formulas, I have had very little to say about their general application out of the Gaussian setting or as a general tool of statistical analysis, which, after all, is where they began in Hotelling's original paper [41]. However, the interested reader can turn to the following papers, which cover various problems, primarily in hypothesis-testing situations, to see these tools in action in a wider class of problems: [45] by Johansen and Johnstone for an excellent introduction to the area, [70, 71] by Naiman and Wynn, [49, 96, 98, 99] by Siegmund and his coworkers, and [106, 108] by Sun.

6.8. *Karhunen–Loève and the density of the maximum.* I would be amiss if I did not point out somewhere that the Karhunen–Loève approach of (6.3.1)–(6.3.4) can also be used (without the need for tube formulas, but with other surface area computations) to study the existence and properties of the density of $\sup_T X$ in quite general situations. Hence this subsection.

For details, see Cirelśon [20], Diebolt and Posse [27–29] and the references therein.

7. Directions for research. I feel a little uncertain about adding this section, since Joe Gani taught me, over two decades ago, that the most interesting directions for new research in probability are usually generated by modeling needs rather than by pontificating probabilists. Nevertheless, here are some thoughts:

7.1. *Deeper connections and issues of accuracy.* One of the most intriguing themes of this review has been how well the expected Euler characteristic matches up with the excursion probability. In all the cases where this could be shown, the proof was extremely indirect, in that both expressions were computed and seen to be the same. It would be very nice to somehow establish,

in a more formal sense, that the two (under appropriate conditions) *must* be the same, thereby alleviating the need to compute them both.

The Slepian model processes of Section 4.2 do indicate, via sample paths, that the two should be similar, and this is the justification behind the Poisson clumping heuristic. However, these do not show why the agreement should be to as high a degree as it is.

At this point, I cannot see any way in which to do this, but have the feeling (and have had for some time) that we are all missing some "obvious" deeper connection here.

One immediate application of such a result would be the lifting of the assumption of the diagonality[25] of Λ in Theorem 4.5.2, without having to reprove the theorem from scratch.

If such a deeper connection exists, then it may also be possible to exploit it to extend the basically Gaussian set of results that we have to non-Gaussian cases and so see how accurate are the mean Euler characteristic expressions as approximations to excursion probabilities there. The importance of this lies in the fact that while for many non-Gaussian cases we do know how to compute long expressions for the mean Euler characteristic (cf. Section 3.5) we do not know how to compute decent expansions for excursion probabilities.

7.2. *Non-Gaussian processes*: *II*. As just mentioned, it also would be nice to develop expansions like (1.1.10) for the excursion probabilities of non-Gaussian processes and fields from their mean Euler characteristic in many cases. In fact, there is no a priori reason not to follow the logic of the Gaussian case and simply identify the two expressions. However, I know of no rigorous justification for this for any non-Gaussian process.

Alternatively, one can try to extend the tube approach to the non-Gaussian scenario, an approach that should work only in special cases, but then should work well.

For example, go back to the Karhunen–Loève expansion (6.2.4), but now with non-Gaussian ξ_n. The tube approach will still work as before, as long as after the normalization in (6.3.3) we still have that the $U_k = \xi_k/(\sum_j \xi_j^2)^{1/2}$ are uniformly distributed on the sphere. That is, we require some sort of spherical symmetry for the ξ_k.

McCormick [65] has adopted this approach when the "Karhunen Loève" expansion is finite, and $\xi \equiv (\xi_1, \ldots, \xi_n)$, where n is the order of the expansion, has a spherically symmetric distribution, and the distribution of $\|\xi\|$ either has compact support or is in the (extremal) domain of attraction of $\exp(-e^{-x})$. For this situation, he has found the first two terms in an expansion like (1.1.10).

An interesting example of a non-Gaussian computation for a smoothed Poisson field can be found in [82], but this is a very different class of processes from those we have considered.

[25]Which I am quite certain is an artificial condition resulting only from the style of the proof.

It also would be rather interesting to extend this approach to the heavy-tailed scenario, and so improve on first-order results for the suprema of stable processes as in Marcus and coworkers [59, 61], Samorodnitsky and Taqqu [89] and related papers. As an aside, it is worth noting that it is not at all clear that the Euler characteristic approach has much to offer in this scenario, since no examples of mean Euler characteristics for stable fields have yet been computed. In fact, even for stable processes on the line, only first-order approximations to the mean level crossing rates are known; cf. [5, 60, 62].

7.3. *Nonstationary fields.* In Section 4.4, we noted how difficult, in practice, it can be to do computations in the nonstationary case. Nevertheless, nonstationary processes really do arise in many practical cases. This is clearly an area that requires more attention. As mentioned there, Shafie's approach [92, 93] via computer algebra seems extremely promising.

7.4. *Random fields on manifolds.* Although it may seem so at first sight, this is not a suggestion about generalisation for its own sake. Examples of fields on manifolds, and related excursion problems, arise in both of the examples I used to motivate this paper. In the brain mapping situation, the natural manifold of study is the surface of the cortex. In the astrophysical problem, manifolds arise naturally in two ways. The first is a sampling issue. As described in the Introduction, the COBE data, along with similarly obtained data sets, are really on a sphere. The second, more fundamental way is due to the basic structure of galaxies themselves, which tend to self-organise on two-dimensional manifolds in space.

Piterbarg and coworkers [66, 71] already have some basic results about excursion probabilities for Gaussian processes on manifolds. Worsley [121] adopts the Euler characteristic approach to the manifold setting by first "thickening" the manifold, so that it becomes a "nice" set in its ambient space. One can then apply the usual Euler characteristic approach, and ultimately "thin" the results of that analysis to obtain results on the manifold.

Both of these papers, however, seem to represent only the tip of a very interesting iceberg.

7.5. *The distribution of the empirical Euler characteristic.* Throughout this paper, we have concentrated on the expectation of the Euler characteristic and have had nothing to say about its distribution. In fact, based on experience with level crossings, $N_\lambda(T)$, in one dimension, it is unlikely that this is a nut we shall ever be able to crack.

Nevertheless, in one dimension there do exist central limit theorem–style results for

$$\frac{N_\lambda(T) - E\{N_\lambda(T)\}}{\sqrt{T}}$$

as $T \to \infty$ (cf. [25], [58]). While this is not really the scenario that interests us (since we have always dealt with T of a fixed size), it is nevertheless of interest and far more tractable than the nonasymptotic situation.

In view of the above, one assumes that CLTs also exist for the empirical Euler characteristic $\varphi(A_\lambda(X, T))$ as $|T| \to \infty$. In fact, functional CLTs should also be provable, so that the test statistic

$$\sup_{-\infty<\lambda<\infty} \frac{\varphi(A_\lambda(X, T)) - E\{\varphi(A_\lambda(X, T))\}}{n(\lambda, T)}$$

(with n a normalizing function) of (3.6.1) should converge to a nice limit as well, as $|T| \to \infty$. Of course, since one assumes that the ratio here, considered as a function of λ, should converge weakly to a Gaussian process, finding the distribution of supremum takes us back to (1.1.1), and so we have come full circle.

This *has* to be a good place to stop, and perhaps turn to filling in all the missing details in the last 69 pages by getting back to work on [3].

Acknowledgments. The reason for my renewed interest in random fields after a hiatus of over a decade was due, initially, to the two applications described briefly in the Introduction, but, ultimately, to the infectious enthusiasm of Keith Worsley that went along with them. My thanks to him for reinfecting me. Jiayang Sun, some of whose work is described in Section 6 on tube formulas, showed amazing patience in helping me begin to understand what was going on there.

Much of this paper was written at the Mathematical Sciences Research Institute in Berkeley, during its special year on stochastic analysis. I thank the organisers for allowing me to work on Gaussian processes in their Markov midst.

Notes added in proof. In the 18 months between posting this paper on my Web page, and its reaching proof stage, three important developments have taken place.

1. A. Takemura and S. Kuriki, in "Maximum of Gaussian field on piecewise smooth domain: Equivalence of tube method and Euler characteristic method" available at `www.e.u-tokyo.ac.jp/~takemura`, have studied processes with finite Karhunen-Loève expansion of the form $\sum_1^K t_k \xi_k$, where the ξ_k are iid standard Gaussian, and $t = (t_1, \ldots, t_K)$ ranges over a manifold in S^{K-1} with piecewise smooth boundary. They have shown that also for these processes the "Punchline" of §6.6 is correct: The Euler characteristic and tube approaches to computing excursion probabilities give the same asymptotic result. Their proof is more "intrinsic" than those in this paper, and may contain the kernel of what is needed to establish the Punchline in general, without resorting to actually computing the answer by each technique and then comparing results. In order to prove their result, they also establish a version of Morse's Theorem (Theorem 3.1.2) that is valid for manifolds with piecewise smooth boundary, a result of independent interest.

See also "Distribution of the maximum of Gaussian random field: tube method and Euler characteristic method" (in Japanese), *Proc. Inst. Stat. Math.* **47** 201–221 (1999).

2. The same authors, in "Tail probability via tube formula and Euler characteristic method when critical radius is zero," available at the same web site, have also found some examples (for which the critical radius of (6.4.6) is zero) where the Euler characteristic and tube approaches disagree, and *both* give the wrong answer. The fields they study here are all *non*-Gaussian.
3. In work under progress, Jonathan Taylor (a doctoral student at McGill and the Technion) has found explicit expressions for the expected Euler characteristic of a constant variance Gaussian process defined over quite general (abstract) manifolds, with boundaries. His work also includes a more general version of Morse's Theorem, again of independent interest.

REFERENCES

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

[2] ADLER, R. J. (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. IMS, Hayward, CA.

[3] ADLER, R. J. (2001). *Random Fields and Their Geometry*. Birkhäuser, Boston.

[4] ADLER, R. J. and FIRMAN, D. (1981). A non-Gaussian model for random surfaces. *Philos. Trans. Roy. Soc. London Ser. A* **303** 433–462.

[5] ADLER, R. J. and SAMORODNITSKY, G. (1997). Level crossings of absolutely continuous stationary symmetric α-stable processes. *Ann. Appl. Probab.* **6** 460–493.

[6] ALDOUS, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic*. Springer, New York.

[7] AZAÏS, J.-M. and WSCHEBOR, M. (1999). The distribution of the maximum of a stochastic process by the Rice method. Preprint.

[8] BENNETT ET AL. (1994). Morphology of the interstellar cooling lines detected by COBE. *Astrophys. J.* **434** 587–598.

[9] BERMAN, S. M. (1964). Limit theorems for the maximum term in stationary sequences. *Ann. Math. Statist.* **35** 502–516.

[10] BERMAN, S. M. (1971). Maxima and high level excursions of stationary Gaussian processes. *Trans. Amer. Math. Soc.* **160** 65–85.

[11] BERMAN, S. M. (1971). Asymptotic independence of the numbers of high and low level crossings of stationary Gaussian processes. *Ann. Math. Statist.* **42** 927–945.

[12] BERMAN, S. M. (1992). *Sojourns and Extremes of Stationary Processes*. Wadsworth and Brooks/Cole, Pacific Grove, CA.

[13] BICKEL, P. and ROSENBLATT, M. (1973). Two-dimensional random fields. In *Multivariate Analysis III* (P. R. Krishnaiah, ed.) 3–15. Academic, New York.

[14] BORELL, C. (1975). The Brunn–Minkowski inequality in Gauss space. *Invent. Math.* **30** 205–216.

[15] BREITUNG, K. (1996). Higher order approximations for maxima of random fields. In *IUTAM Symposium on Advances in Nonlinear Stochastic Mechanics* (A. Naess and S. Krenk, eds.) 79–88. Kluwer, Dordrecht.

[16] BULINSKAYA, E. V. (1961). On the mean number of crossings of a level by a stationary Gaussian process. *Theory Probab. Appl.* **6** 435–438.

[17] CAO, J. (2000). The size of the connected components of excursion sets of χ^2, t and F fields. *Adv. in Appl. Probab.* **31** 579–595.

[18] CAO, J. and WORSLEY, K. J. (1999). The detection of local shape changes via the geometry of Hotelling's T^2 fields. *Ann. Statist.* **27** 925–942.

[19] CAO, J. and WORSLEY, K. J. (1999). The geometry of correlation fields with an application to functional connectivity of the brain. *Ann. Appl. Prob.* **9** 1021–1057.

[20] CIRELSON, B. S. (1975). Density of the distribution of the maximum of a Gaussian process (in Russian). *Theory Probab. Appl.* **20** 847–856.
[21] CIRELSON, B. S., IBRAGIMOV, I. A. and SUDAKOV, V. N. (1976). Norms of Gaussian sample functions. In *Proceedings of the Third Japan–USSR Symposium on Probability Theory*. *Lecture Notes in Math.* **550** 20–41. Springer, New York.
[22] CRAMÉR, H. (1965). A limit theorem for the maximum values of certain stochastic processes. *Theory Probab. Appl.* **10** 137.
[23] CRAMÉR, H. (1996). On the intersections between the trajectories of a normal stationary stochastic process and a high level. *Ark. Mat.* **6** 337–349.
[24] CRAMÉR, H. and LEADBETTER, M. R. (1967). *Stationary and Related Stochastic Processes*. Wiley, New York.
[25] CUZIK, J. (1976). A central limit theorem for the number of zeros of a stationary Gaussian process. *Ann. Probab.* **4** 547–556.
[26] DELMAS, C. (1998). An asymptotic expansion for the distribution of the maximum of a class of Gaussian fields. *C. R. Acad. Sci. Paris Sér. I Math.* **327** 393–397.
[27] DIEBOLT, J. (1981). Sur la loi du maximum de certains processus gaussiens sur le tore. *Ann. Inst. H. Poincaré Probab. Statist.* **17** 165–179.
[28] DIEBOLT, J. and POSSE, C. (1995). A nonasymptotic approach to the density of the maximum of smooth Gaussian processes. *C. R. Acad. Sci. Paris Sér. I Math.* **321** 933–938.
[29] DIEBOLT, J. and POSSE, C. (1996). On the density of the maximum of smooth Gaussian processes. *Ann. Probab.* **24** 1104–1129.
[30] DUDLEY, R. M. (1967). The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Funct. Anal.* **1** 290–330.
[31] DUDLEY, R. M. (1973). Sample functions of the Gaussian process. *Ann. Probab.* **1** 66–103.
[32] EILENBERG, S. and STEENROD, N. (1952). *Foundations of Algebraic Topology*. Princeton Univ. Press.
[33] FATALOV, V. R. (1992). Exact asymptotics of large deviations of Gaussian measures in a Hilbert space (in Russian). *Izv. Nats. Akad. Nauk Armenii Mat.* **27** 43–61.
[34] FATALOV, V. R. (1993). Asymptotics of the probabilities of large deviations of Gaussian fields: applications (in Russian). *Izv. Nats. Akad. Nauk Armenii Mat.* **28** 25–51.
[35] GNEDENKO, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44** 423–453.
[36] GRAY, A. (1990). *Tubes*. Addison–Wesley, Redwood City, CA.
[37] HADWIGER, H. (1957). *Vorlesüngen Über Inhalt, Oberfläche und Isoperimetrie*. Springer, Berlin.
[38] HADWIGER, H. (1971). Normale Körper im euklidischen Raum und ihre topologischen und metrischen Eigenschaften. *Math. Z.* **71** 124–140.
[39] HALL, P. (1988). *Introduction to the Theory of Coverage Processes*. Wiley, New York.
[40] HOGAN, M. L. and SIEGMUND, D. (1986). Large deviations for the maxima of some random fields. *Adv. in Appl. Math.* **7** 2–22.
[41] HOTELLING, H. (1939). Tubes and spheres in n-spaces and a class of statistical problems. *Amer. J. Math.* **61** 440–460.
[42] ITÔ, K. (1964). The expected number of zeros of continuous stationary Gaussian processes. *J. Math. Kyoto Univ.* **3** 206–216.
[43] JAMES, B., JAMES, K. L. and SIEGMUND, D. (1988). Conditional boundary crossing probabilities, with applications to change-point problems. *Ann. Probab.* **16** 825–839.
[44] JENNEN, C. (1985). Second-order approximations to the density, mean and variance of Brownian first-exit times. *Ann. Probab.* **13** 126–144.
[45] JOHANSEN, S. and JOHNSTONE, I. M. (1990). Hotelling's theorem on the volume of tubes: some illustrations in simultaneous inference and data analysis. *Ann. Statist.* **18** 652–684.
[46] JOHNSTONE, I. and SIEGMUND, D. (1989). On Hotelling's formula for the volume of tubes and Naiman's inequality. *Ann. Statist.* **17** 184–194.
[47] KAC, M. (1943). On the average number of real roots of a random algebraic equation. *Bull. Amer. Math. Soc.* **43** 314–320.

[48] KAC, M. and SLEPIAN, D. (1959). Large excursions of Gaussian processes. *Ann. Math. Statist.* **30** 1215–1228.
[49] KNOWLES, M. and SIEGMUND, D. (1989). On Hotelling's approach to testing for a nonlinear parameter in a regression. *Internat. Statist. Rev.* **57** 205–220.
[50] KRATZ, M. and ROOTZÉN, H. (1997). On the rate of convergence for extremes of mean square differentiable stationary normal processes. *J. Appl. Probab.* **34** 908–923.
[51] KREYSZIG, E. (1968). *Introduction to Differential Geometry and Reimannian Geometry.* Univ. Toronto Press. Toronto.
[52] LANDAU, H. and SHEPP, L. A. (1970). On the supremum of a Gaussian process. *Sankhyā Ser. A* **32** 369–378.
[53] LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes.* Springer, New York.
[54] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces. Isoperimetry and Processes.* Springer, Berlin.
[55] LERCHE, H. R. and SIEGMUND, D. (1989). Approximate exit probabilities for a Brownian bridge on a short time interval, and applications. *Adv. in Appl. Probab.* **21** 1–19.
[56] LIFSHITS, M. A. (1994). Tail probabilities of Gaussian extrema and Laplace transform. *Ann. Inst. H. Poincaré Probab. Statist.* **30** 163–179.
[57] LINDGREN, G. (1972). Local maxima of Gaussian fields. *Ark. Mat.* **10** 195–218.
[58] MALEVICH, T. L. (1969). Asymptotic normality of the number of crossings of level zero by a Gaussian process. *Theory Probab. Appl.* **14** 287–295.
[59] MARCUS, M. B. (1987). *ξ-Radial Processes and Random Fourier Series.* Amer. Math. Soc., Providence, RI.
[60] MARCUS, M. B. (1989). Some bounds for the expected number of level crossings of symmetric harmonizable p-stable processes. *Stochastic Process. Appl.* **33** 217–231.
[61] MARCUS, M. B. and PISIER, G. (1981). *Random Fourier Series with Applications to Harmonic Analysis.* Princeton Univ. Press.
[62] MARCUS, M. B. and SHEN, K. (1997). Bounds for the expected number of level crossings of certain harmonizable infinitely divisible processes. Preprint.
[63] MARCUS, M. B. and SHEPP, L. A. (1971). Sample behavior of Gaussian processes. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2** 423–442. Univ. California Press, Berkeley.
[64] MATHERON, G. (1975). *Random Sets and Integral Geometry.* Wiley, New York.
[65] MCCORMICK, W. P. (1997). A geometric approach to obtaining the distribution of the maximum for a class of random fields. Preprint.
[66] MIKHALEVA, T. L. and PITERBARG, V. I. (1996). On the distribution of the maximum of a Gaussian field with constant variance on a smooth manifold. *Theory Probab. Appl.* **41** 367–379.
[67] MILLMAN, R. S. and PARKER, G. D. (1977). *Elements of Differential Geometry.* Prentice–Hall, Englewood Cliffs, NJ.
[68] MORSE, M. and CAIRNS, S. (1969). *Critical Point Theory in Global Analysis and Differential Topology.* Academic, New York.
[69] NAIMAN, D. Q. (1990). Volumes of tubular neighborhoods of spherical polyhedra and statistical inference. *Ann. Statist.* **18** 685–716.
[70] NAIMAN, D. Q. and WYNN, H. P. (1992). Inclusion–exclusion–Bonferroni identities and inequalities for discrete tube-like problems via Euler characteristics. *Ann. Statist.* **20** 43–76.
[71] PICKANDS, J., III (1969). Upcrossing probabilities for stationary Gaussian processes. *Trans. Amer. Math. Soc.* **145** 51–74.
[72] PICKANDS, J., III (1969). Asymptotic properties of the maximum in a stationary Gaussian process. *Trans. Amer. Math. Soc.* **145** 75–86.
[73] PISIER, G. (1986). Probabilistic methods in the geometry of Banach space. *Lecture Notes in Math.* **1206** 167–241. Springer, New York.
[74] PITERBARG, V. I. (1981). Comparison of distribution functions for maxima of Gaussian processes. *Theory Probab. Appl.* **26** 687–705.
[75] PITERBARG, V. I. (1982). Large deviations of random processes close to Gaussian ones. *Theory Probab. Appl.* **27** 504–524.

[76] PITERBARG, V. I. (1996). *Asymptotic Methods in the Theory of Gaussian Processes and Fields*. Amer. Math. Soc., Providence, RI.
[77] PITERBARG, V. I. (1997). Rice's method for large excursions of Gaussian random fields. Technical Report 478, Center for Stochastic Processes, Univ. North Carolina. [Translation of Russian version in *Fund. Appl. Math.* **2** (1996) 187–204.]
[78] PITERBARG, V. I. and STAMATOVICH, S. (1998). On the maximum of Gaussian non-centered fields indexed on smooth manifolds. Preprint.
[79] PITERBARG, V. I. and WEBER, M. (1997). Tail distribution results for Gaussian suprema—standard methods. Technical Report 491, Center for Stochastic Processes, Univ. North Carolina.
[80] POWELL, C. S. (1992). The golden age of cosmology. *Scientific American* July 17–22.
[81] QUALLS, C. and WATANABE, H. (1973). Asymptotic properties of Gaussian random fields. *Trans. Amer. Math. Soc.* **177** 155–171.
[82] RABINOWITZ, D. and SIEGMUND, D. (1997). The approximate distribution of the maximum of a smoothed Poisson random field. *Statist. Sinica* **7** 167–180.
[83] RESNICK, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer, New York.
[84] RICE, S. O. (1945). Mathematical analysis of random noise. *Bell. System Tech. J.* **24** 46–156. [Reprinted in (1954). *Selected Papers on Noise and Stochastic Processes* (N. Wax, ed.) Dover, New York.]
[85] RIESZ, R. and SZ-NAGY, B. (1955). *Functional Analysis*. Ungar, New York.
[86] SAMORODNITSKY, G. (1987). Bounds on the supremum distribution of Gaussian processes—polynomial entropy case. Preprint.
[87] SAMORODNITSKY, G. (1987). Bounds on the supremum distribution of Gaussian processes—exponential entropy case. Preprint.
[88] SAMORODNITSKY, G. (1991). Probability tails of Gaussian suprema. *Stochastic. Process. Appl.* **38** 55–84.
[89] SAMORODNITSKY, G. and TAQQU, M. S. (1994). *Stable Non-Gaussian Random Processes*. Chapman and Hall, New York.
[90] SANTALÓ, L. A. (1976). *Integral Geometry and Geometric Probability. Encyclopedia of Mathematics and Its Applications* **1.** Addison–Wesley, Reading, MA.
[91] SCHNEIDER, R. (1993). *Convex Bodies: The Brunn–Minkowski Theory*. Cambridge Univ. Press.
[92] SHAFIE, KH., WORSLEY, K. J., WOLFORTH, M. and EVANS A. C. (1998). Rotation space: detecting functional activation by searching over rotated and scaled filters. *NeuroImage* **7** S755.
[93] SHAFIE, KH. (1998). Ph.D. dissertation, Dept. Math. Statist., McGill Univ.
[94] SIEGMUND, D. (1982). Large deviations for boundary crossing probabilities. *Ann. Probab.* **10** 581–588.
[95] SIEGMUND, D. (1986). Boundary crossing probabilities and statistical applications. *Ann. Statist.* **14** 361–404.
[96] SIEGMUND, D. (1988). Approximate tail probabilities for the maxima of some random fields. *Ann. Probab.* **16** 487–501.
[97] SIEGMUND, D. (1992). Tail approximations for maxima of random fields. In *Probability Theory*. 147–158. de Gruyter, Berlin.
[98] SIEGMUND, D. O. and WORSLEY, K. J. (1995). Testing for a signal with unknown location and scale in a stationary Gaussian random field. *Ann. Statist.* **23** 608–639.
[99] SIEGMUND, D. and ZHANG, H. (1993). The expected number of local maxima of a random field and the volume of tubes. *Ann. Statist.* **21** 1948–1966.
[100] SLEPIAN, D. (1962). The one-sided barrier problem for Gaussian noise. *Bell System Tech. J.* **42** 463–501.
[101] SLEPIAN, D. (1963). On the zeroes of Gaussian noise. In *Time Series Analysis* (M. Rosenblatt, ed.) 104–115. Wiley, New York.
[102] SMOOT, G. F., BENNETT, C. L., KOGUT, A., WRIGHT, E. L., AYMON, J., BOGGESS, N. W., CHENG, E. S., DE AMICI, G., GULKIS, S., HAUSER, M. G., HINSHAW, G., JACKSON, P. D., JANSSEN,

M., KAITA, E., KELSALL, T., KEEGSTRA, P., LINEWEAVER, C., LOWENSTEIN, K., LUBIN, P., MATHER, J., MEYER, S. S., MOSELEY, S. H., MURDOCK, T., ROKKE, L., SILVERBERG, R. F., TENORIO, L., WEISS, R. and WILKINSON, D. T. (1992). Structure in the *COBE* differential microwave radiometer first-year maps. *Astrophys. J.* **396** L1–L5.

[103] SPIVAK, M. (1979). *A Comprehensive Introduction to Differential Geometry* **1**, 2nd ed. Publish or Perish, Wilmington, DE.

[104] SUDAKOV, N. V. (1969). Gaussian and Cauchy measures and ε-entropy. *Soviet Math. Dokl.* **10** 310–313.

[105] SUDAKOV, N. V. (1971). Gaussian random processes, and measures of solid angles in Hilbert space. *Soviet Math. Dokl.* **12** 412–415.

[106] SUN, J. (1991). Significance levels in exploratory projection pursuit. *Biometrika* **78** 759–769.

[107] SUN, J. (1993). Tail probabilities of the maxima of Gaussian random fields. *Ann. Probab.* **21** 34–71.

[108] SUN, J. (1993). Some practical aspects of exploratory projection pursuit. *SIAM J. Sci. Comput.* **14** 68–80.

[109] SUN, J. (1997). Personal communication.

[110] TALAGRAND, M. (1987). Regularity of Gaussian processes. *Acta Math.* **159** 99–149.

[111] TALAGRAND, M. (1992). A simple proof of the majorizing measure theorem. *Geom. Funct. Anal.* **2** 118–125.

[112] TALAGRAND, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.* **22** 28–76.

[113] TALAGRAND, M. (1994). The supremum of some canonical processes. *Amer. J. Math.* **116** 283–325.

[114] TOMITA, H. (1990). In *Formation, Dynamics and Statistics of Patterns* (K. Kawasaki, M. Suzuki and A. Onuki, eds.) **1** 113–157. World Scientific, Singapore.

[115] TURMON, M. (1995). *Assessing generalization of feedforward neural networks.* Ph.D. dissertation, Cornell Univ.

[116] WALLACE, A. H. (1968). *Differential Topology: First Steps.* Benjamin, New York.

[117] WEYL, H. (1939). On the volume of tubes. *Amer. J. Math.* **61** 461–472.

[118] WOODROOFE, M. and TAKAHASHI, H. (1982). Asymptotic expansions for the error probabilities of some repeated significance tests. *Ann. Statist.* **10** 895–908. [Correction: (1985) **13** 837.]

[119] WORSLEY, K. J. (1994). Local maxima and the expected Euler characteristic of excursion sets of χ^2, F and t fields. *Adv. in Appl. Probab.* **26** 13–42.

[120] WORSLEY, K. J. (1995). Estimating the number of peaks in a random field using the Hadwiger characteristic of excursion sets, with applications to medical images. *Ann. Statist.* **23** 640–669.

[121] WORSLEY, K. J. (1995). Boundary corrections for the expected Euler characteristic of excursion sets of random fields, with an application to astrophysics. *Adv. in Appl. Probab.* **27** 943–959.

[122] WORSLEY, K. J. (1997). The geometry of random images. *Chance* **9** 27–40.

[123] WORSLEY, K. J. (1998). Testing for signals with unknown location and scale in a χ^2 random field, with an application to fMRI. Preprint.

[124] YLVISAKER, N. D. (1965). The expected number of zeros of a stationary Gaussian process. *Ann. Math. Statist.* **36** 1043–1046.

[125] ZAANEN, A. C. (1956). *Linear Analysis.* North-Holland, Amsterdam.

FACULTY OF INDUSTRIAL ENGINEERING AND MANAGEMENT
TECHNION—ISRAEL INSTITUTE OF TECHNOLOGY
HAIFA 32000
ISRAEL
E-MAIL: robert@ieadler.technion.ac.il