1

Introduction

1.1 Distribution of extremes in random fields

The aim of this book is to present a method for analyzing the tail distribution of extreme values in random fields. A random field can be considered as a collection of random variables $\{X_t : t \in T\}$, indexed by a set of parameters T. The index set T may be quite complex. However, in the applications that we will analyze in this book it will typically turn out that T is a 'nice' subset of $\mathbb{R}^d$, the d-dimensional space of real numbers.

In some statistical applications one is interested in probabilities such as:

$$\mathrm{P}\left(\sup_{t \in T} X_t \geq x\right) ,$$

the probability that the maximum of the random field exceeds a threshold x, for large values of x. There are only a few special cases in which the problem of computing such probabilities has an exact solution. In all other cases one is forced to use numerical methods, such as simulations, or to apply asymptotic approximations in order to evaluate the probability. This book concentrates on the application of the proposed method for producing asymptotic analytical expansions of the probability. Nonetheless, some elements in the method may, and have been, applied in order to simulate numerical evaluations more efficiently. An application that illustrates the usefulness of the method in the context of simulations is presented in the second part of the book.

As a motivating example consider scanning statistics. Scanning statistics are used in order to detect rare signals in an environment contaminated by random noise. For example, let us assume measurements that are taken in a one-dimensional environment. Each measurement is associated with a point in the environment and the points are equally spaced. For the most part, the expected values of the observations are fixed at some baseline level throughout

the environment. However, at some unknown locations the expected value is different from the baseline. Such a shift of the expectation extends over an interval of unknown length. An interval of shifted expectations is the signal we seek to identify. Such a signal is parameterized by the location of the interval, by the length of the interval, and perhaps also by the magnitude of the shift.

The expectations of the observations correspond to signals (or lack thereof). A complication in fulfilling the task at hand is the fact that the observations are subject also to random noise, which may be parameterized by the variance of the observations. Frequently, this random noise is taken to be normally distributed and independent among observations. In such a case, the expectation structure and the variance specifies completely the distribution of the observations.

Say that our goal is to decide whether or not there is any signal in the environment. A reasonable approach, which has statistical merits to it, is to associate with each potential signal a statistic that summarizes the information in the data regarding that signal. For example, if signals are all of the form of an interval with a fixed level of the expectation above the baseline then an appropriate statistic is the standardized sample average of the observations that belong to the interval, with standardization conducted with respect to the baseline expectation and variance. The presence of a signal in the environment is announced if there exists a statistic with a value above a previously determined threshold. False detection occurs when all observations share the same background expectation level but, due to random fluctuations, the threshold is crossed. The preliminary task of the statistician, in order to limit the probability of false alarms, is to determine the value of the threshold.

In the current example the parameter set T corresponds to the collection of all potential signal intervals. An element in T, an interval, is composed of a pair of numbers – the central location of the interval on the line and the length of the interval. Hence, T is a subset of the two-dimensional plain of real numbers. The statistic X_t that is associated with $t \in T$ is the standardized sample average based on the observations that belong to the interval t. Note that although the original observations were assumed to be independent, the collection of statistics $\{X_t : t \in T\}$ are not since two overlapping intervals share some of the observations. On the other hand, if the noise is normal then the distribution of the statistics is also normal. A graphical illustration of the situation is given in Figure 2.5 in Section 2.3.[1] Denote the threshold by x. It follows that the probability of false detection is of the form:

$$P\left(\sup_{t \in T} X_t \geq x\right),$$

which is the form we declared to be of interest for us. Approximations that relate the probability to the value of the threshold x enables one to select the value of the threshold to meet tolerated levels of the probability.

---

[1] In Figure 2.5, our t is denoted there by $\theta = (t, h)$, with t the central location and h the length of the interval. The statistic that we denote here by X_t is denoted there by Z_θ.

The text is devoted to the task of analyzing the tail probability of extremes. We ignore completely more fundamental issues of sample-path properties of the random field and questions of measurability of the random variable $\sup_{t \in T} X_t$. The parameter set T in many of our applications is either finite or countable. In such a case measurability of the supremum follows readily. In other cases one may rely on the separability of the random field in order to establish the measurability requirement. As for us in this book, we just ignore the issue.

The analysis of the probability that the maximal value of the field exceeds the threshold x in the situation where T is such that this probability is vanishingly small will occupy a central part of the discussion. This type of analysis, the analysis of vanishingly small probabilities, is frequently referred to as *large deviations*. Typically, a statement of a theorem in large deviations establishes the exponential rate by which the probability converges to zero. This first-order approximation of the probability will not be sufficient for our needs. The aim of our analysis will be to produce refined approximations, approximations that include polynomial terms and associated constants. These refined expansions open the door for the production of approximations to probability of events that involve maximization of a random field in settings where probabilities do not converge to zero or the computation of other functionals that are associated with such events.

For example, in the case of a scanning statistic with normal noise we have that the marginal distribution of each element in the random field is standard normal. Consequently,

$$P(X_t \geq x) = 1 - \Phi(x) \approx \frac{1}{x\sqrt{2\pi}} e^{-\frac{1}{2}x^2} ,$$

where Φ is the cumulative distribution function of the standard normal distribution and the approximation is valid for large values of x. It can be shown that

$$\lim_{x \to \infty} x^{-2} \log P \left(\sup_{t \in T} X_t \geq x \right) = -\frac{1}{2} ,$$

which is the content of the large deviation statement in this case. However, in Chapters 2 and 3 we prove the more detailed approximation:

$$P \left(\max_{t \in T} X_t \geq x \right) \sim x^3 e^{-\frac{1}{2}x^2} (2\pi)^{-\frac{1}{2}} \cdot (0.5)^2 (t_1 - t_0)(1/h_0 - 1/h_1) ,$$

when $T = [t_0, t_1] \times [h_0, h_1] \subset \mathbb{R}^2$. This more detailed description involves both the large deviation rate (appearing in the approximation in the form of the element $e^{-\frac{1}{2}}$) but also polynomial terms and constants.

In this book we will selfishly concentrate on a specific approach for dealing with the problem at hand and thus portray the false image that the method that we present is the best method, not to say the only method, for producing asymptotic

approximation of the probability that a random field obtains an extremely high value. The special situation where the index T is a subset of the real line, in which case the random field is actually a random process, has a long history and many tools for solving. Some of the alternative methods of solution in this case will be presented briefly in the next chapter. Another notable special case with a very elegant theory is the situation where the random field is Gaussian with a smooth covariance structure. A more general tool may be applied in the Gaussian setting that involves a continuous parameter set T for cases where the covariance function does not have derivatives. There are far fewer tools available in order to deal with the even more general case where the random field is not Gaussian, and to handle in this non-Gaussian setting cases where the index set is discrete or the sample paths are not smooth.

The toolbox of the probabilist is not completely empty when faced with these more general problems. However, the optional methods are limited in number and none is very elegant. Admittedly, one may question the elegancy of the method that we will advocate. Still, the method seems to work in many different settings and thus may claim the title of generality.

The method is defined through a series of steps in which the large deviation part of the probability is accounted for first, followed by refinements that result from the identification of the contributions that are due to global and to more local fluctuations. These recommended steps may help to organize the investigation of the probabilistic characteristics of the problem at hand and assist in the evaluation of the relative contribution of the various sources of variability.

The demonstration of methods for approximating the extreme tail of the distribution of the maxima of a random field is initiated in the next chapter. The current introductory chapter is devoted to mental preparation. In Section 1.2 we provide a bird's view of the proposed method. In that section we outline the role and characteristics of the different steps. All the details are left out and are given in subsequent chapters.

Section 1.3 presents the type of random fields in which one may hope to apply the method as it is presented in the current text. Essentially, we are motivated by the analysis of Gaussian random fields, yet the the method is marketed as a tool that works for non-Gaussian fields as well. Nonetheless, the approximation is based to some extent on the application of the central limit theorem. Consequently, the type of fields that we target are those that obey the central limit theorem in an appropriate sense. We discuss such fields in Section 1.3.

In Section 1.4 we give a list of applications. Random fields associated with two relatively simple applications will serve throughout the first part of the book as a demonstration of classical methods for approximating the distribution of extremes, as well as the demonstration of our method. Other, more complicated examples will be discussed in the second part of the book, the part devoted to applications. For these examples we use the method described in Section 1.2.

1.2   Outline of the method

The method we propose is motivated by statistical considerations. We are inspired to think of the parameter t of the field as specifying a statistical model and consider X_t as a statistic that summarizes the information regarding the parameter. In many of the applications that we will consider this indeed is the context in which the field emerges. In other applications, when this is not the case, we may still consider that point of view as a motivation and a guiding principle.

More specifically, we propose to consider the problem of finding the tail distribution of a random field in the context of statistical hypothesis testing. In statistical hypothesis testing competing models for the distribution of the observations are grouped into two sub-collections. One sub-collection is called the null hypothesis. It reflects the absence of a scientifically significant signal in the data. The other sub-collection is composed of alternative distributions which are a reflection of the presence of such a signal. A statistical test is constructed with the role of determining which of the two hypotheses is more consistent with empirical observations.

Inspired by that point of view one may regard the random variable $\sup_{t \in T} X_t$ as the test statistic, with the test itself rejecting the null hypothesis in favor of the alternative if this test statistic is above a threshold x. The null hypothesis itself in this context is composed of a single distribution: the actual distribution of the data. Consequently, rejecting the null hypothesis is an error. The probability that we seek is the probability of making such an error, which in statistical vocabulary is called the significance level of the test.

The alternative collection of models is associated with the set T. Each $t \in T$ specifies a model P_t of the distribution of the data. This distribution may be the actual distribution that was considered, if the problem emerged in the considered statistical context, or it may involve some artificially constructed model that fits our needs. At the heart of the method is the proposal to translate the original problem of computing the probability of being above the threshold under the null hypothesis to a problem of computing expectations under alternative models. The vehicle that carries out this translation is the likelihood ratio identity.

The likelihood ratio identity employs likelihood ratios. A likelihood is the probability of the observed data under a given probabilistic model. If the distribution of the data is continuous then the likelihood refers to the probability's density. A likelihood ratio is the ratio between two likelihoods. Here we consider likelihood ratios in which the denominator is the likelihood of the data under the current distribution (the null distribution) and the numerator is the likelihood of the data under the alternative distribution P_t. We denote this likelihood ratio by $\exp\{\ell_t\}$, with ℓ_t being the log-likelihood ratio. We relate each X_t to ℓ_t, whether or not X_t emerged originally as a log-likelihood ratio, and rephrase the original problem of crossing the threshold by elements from the

collection $\{X_t : t \in T\}$ using instead elements from the collection $\{\ell_t : t \in T\}$, possibly with a different threshold.

In the book we present a recipe for the application of the method that involves the likelihood ratio identity. This recipe is executed in a series of steps, and it is concluded by producing an approximation for the tail probability that we analyze. Unlike the baking of real cakes, one need not follow the proposed steps meticulously in order to avoid disasters. On the contrary, these steps are only guidelines and are not necessarily optimal in all scenarios. Still, we find them useful.

The first step involves the identification of the large deviation rate. The method itself produces refinements to the first-order approximation that is produced by this rate. Frequently, one may find the large deviation rate by the maximization of the marginal probabilities $\{P(\sup_{t \in T} X_t \geq x) : t \in T\}$ over the collection T.

A large deviation rate is associated with a collection of values in the parameter set. Preparation towards the application of the likelihood ratio identity may involve the identification of a subset of parameter values that are most likely to contain the maximizing value and restricting the analysis only to that sub-region.

In the case where the parameter space is continuous one may consider another preparation step in which the maximization is restricted to a dense, but discrete, sub-collection. Although the method can, and has been, implemented directly to a continuous parameter set there are some technical advantages to its implementation in the context of a discrete set of parameters.

After preparations one may invoke the likelihood ratio identity. The outcome of this step is a presentation of the probability, under the given distribution, that the maximum of the field exceeds a threshold in terms of a sum of expectations. The sum extends over the different values of the parameters. Each element in the sum is computed in the context of the alternative probability model specified by that value. The expectation involves a deterministic term that is associated with the large deviation rate and a product of two random terms, one measuring the global behavior of the field in the context of maximization and the other measuring local fluctuations.

The localization theorem, the subsequent step, applies a local limit theorem to the global term in order to prove the asymptotic independence between the effect of the global term and the effect of local fluctuations. In the examples that are considered the local limit theorem emerges as a refinement of the central limit theorem. Consequently, the given approach is more natural in problems where the global term obeys a central limit theorem and converges to the normal distribution. The outcome of this step is an approximation of each parameter-specific expectation in the representation by the product of three factors: a factor associated with the large deviation rate, a factor associated with the density of the normal limit of the global term, and a factor that measures local fluctuations. The integrated approximation of the tail probability is obtained by the summation of these products over the collection of parameter values.

The method is employed in settings consistent with large deviation formulation. Accordingly, the probability $P(\sup_{t \in T} X_t \geq x)$ converges to zero when the

threshold x diverges to infinity. In other applications, when the parameter set T is increasing fast enough as a function of x, the probability may be converging to a positive constant. In such a case, Poisson-type approximations may be applied in order to extend the approximation obtained in large deviation settings to the setting of non-vanishing probabilities. Convergence in distribution that emerges from Poisson approximations, in conjunction with statement of uniform integrability, may be used in order to approximate functionals associated with extremes that involve expectation.

1.3   Gaussian and asymptotically Gaussian random fields

The random field is a collection of random variables $\{X_t : t \in T\}$ with a joint distribution. The joint distribution is uniquely specified in terms of the finite-dimensional joint distribution of the random vector $\{X_t : t = t_1, t_2 \ldots, t_k\}$, for any finite sub-collection of parameter values $t_1, \ldots, t_k \in T$. In the special case where these finite-dimensional joint distributions are all Gaussian we say that the field is a Gaussian random field.

The joint distribution of a Gaussian random vector, the multinormal distribution, is a function only of the vector of expectations and the matrix of variances and covariances. As a conclusion we get that the distribution of a Gaussian random field is fully specified in terms of the expectation function: the expectation $E(X_t)$, for each $t \in T$, and the variance-covariance function: the covariance $Cov(X_t, X_s)$, for any pair $(t, s) \in T \times T$. The distribution of the maximum of the Gaussian field is influenced both by the deterministic component of the field, namely the expectation, and by the variability, which is determined by the covariance function.

The theory that deals with the investigation of extreme values in Gaussian fields is highly developed. The role of the expectation, and the more delicate role of the covariance structure, in the determination of the distributions of such extremes is well understood. Extremely accurate asymptotic approximations of the distribution of these extremes exist for some subfamilies of Gaussian random fields. Good asymptotic approximations exist for other subfamilies.

In the next chapter we will present the two main tools for analyzing the distribution of extremes in Gaussian fields. One tool is based on the computation of the expectation of the Euler characteristic of the excursion set. This tool is applicable when the realizations of the random field are differentiable and is extremely accurate. The other, more general approach, is known as the double-sum method. It is not as accurate as the first tool but it may apply in situations where the realizations of the field are not smooth.

The smoothness, or lack thereof, of the realizations of a Gaussian random field is determined by the covariance function in general, and the smoothness of this function in the vicinity of the diagonal $\{(t, t) : t \in T\}$ in particular. Consequently, the answer to a question regarding which of the tools can be used is related to

the ability to take partial derivatives of the covariance function at the points of the diagonal. Basically, if second-order partial derivatives exist then the more accurate method of the Euler characteristic may be applied. Otherwise, one should refer to the more general double-sum method.

Our method is more like the double-sum method in the sense that it may be applied under more general conditions, paying for the generality of the application in terms of accuracy of the approximation. It has a further advantage that it can be applied in settings where the random field is not Gaussian, although it should obey a local limit theorem that emerges from the central limit theorem.

The central limit theorem deals with the convergence of sums of random elements to a Gaussian element. If, for example, the elements are independent fields then the resulting limit is a Gaussian field. A central limit theorem in the context of random field relies, typically, on the convergence of the finite-dimensional distribution of the field to a Gaussian limit and on a tightness property. The role of tightness is to ensure that distribution of the field, along the process of convergence, may be approximated uniformly well using a finite and fixed collection of parameter points.

A tempting approach for dealing with the distribution of the extremes in a non-Gaussian setting, in the case where the field in question belongs to a sequence that converges to a Gaussian limit, is to apply the approximation of the distribution of the maximum to the Gaussian limit distribution. In this approach one separates between the convergence of the field to the Gaussian limit, which is carried out first, and the convergence of the tail distribution of the field to zero, which is assessed after the first convergence took place. Tempting as it may be, this approach may produce misleading outcomes. The reason for this is that the central limit theorem, as the name suggests, deals with the central part of the distribution, not with the extreme tail of the distribution. There may very well exist a big difference between the tail behavior of the original field and the tail behavior of a Gaussian field with the same expectation and the same covariance function as the original field. However, this difference is washed away by the central limit theorem. A better approach is to deal directly with the distribution of extreme values of the original field itself and assess its asymptotic behavior.

In the approach outlined in the previous section the probability of the maximum of the field is presented as a sum of terms. These terms are composed of a determinist factor that relates to the large deviation rate, a factor that relates to the contribution of the global term, and a factor that measures the contribution of local perturbations. Separating out the effect of large deviation guarantees an honest assessment of the extreme tail distribution. The central limit theorem plays a part in the derivation of the contribution of the global term and, in some cases, in the contribution of the local fluctuations.

The part played by the central limit theorem in the assessment of the contribution of the global term is in the form of a local limit theorem. A local limit theorem deals with the probability that a statistic, typically a statistic produced by taking a sum, obtains values from an interval of fixed width. The statistic has a variance that goes to infinity. Consequently, the probability of belonging

to the interval goes to 0 at a rate proportional to the standard deviation of the statistic. An accurate assessment of the rate by which the probability converges to zero may be deduced from a central limit theorem that involves a higher order expansion of the approximation error. A famous theorem of that sort is the Berry–Esseen theorem. An important point to make is that the distribution of the global term is assessed in the context of an appropriate alternative distribution, not the original null distribution. Consequently, convergence may hold for the selected alternative distribution even if it does not exist for the original null distribution.

The method relies on a statement regarding the joint limit distribution of the global term and a local field that is derived by local deviations of the field. The requirement is a local limit for the global term and asymptotic independence between the global term and local deviations. The local deviations are not required to converge to a Gaussian limit. This requirement is much less than the requirement that the field converges to a Gaussian limit. In the particular important case where the field does converge to a Gaussian limit, for example when the field is a sum of independent fields, the factor in the approximation that is associated with the local fluctuation is the same as the factor that emerges for Gaussian fields (still, the factor that is associated with large deviation may be different). In other cases, the factor that is associated with local fluctuations may differ from the factors encountered in the Gaussian setting.

1.4   Applications

This book targets people with an interest in statistics and probability as branches of applied mathematics. As such, it will not do an honest job if it does not demonstrate the applicability of the theory to 'real life situations'. 'Application of mathematical theory' is, to some extent, an oxymoron. Typically, what is presented as applications are as abstract and removed from physical reality as the theory that it serves to demonstrate and justify. To justify the relation between an application and real life the mathematician tells a story that portrays the application as something that actual practitioners, people that do real science or put their money (or other people's money) at real risk, care about. These stories can fool outsiders but not people who actually know the details.

This description is true also for the applications given in this book. I am not an insider in either of the topics that appear in the second part of the book. In some cases, for example in genetics, I can say that I know people who know actual practitioners. In other cases I have only my imagination and what I read in Wikipedia to guide me.

Let us list the applications that will be presented in the book. In the first part of the book we will use two simple applications in order to demonstrate the solutions provided by the theory, both classical solutions that will be presented in Chapter 2, and the solution provided by our method. That solution will be mainly given in Chapter 3.

The first example is an example of sequential testing of hypotheses. In this example we will consider the case of two simple hypotheses and the application of the sequential probability ratio test for testing one hypothesis versus the other. In this example the parameter space is one-dimensional – the positive integers – and the random field is the process of a random walk. Tools that are applicable to random walks, in particular random walks stopped by a stopping time, can be used for the analysis of this example.

The second example is an example of a scanning statistic in the spirit of the example that was mentioned at the beginning of the chapter. Scanning statistics are used in order to detect rare signals in an environment contaminated by random noise. In the particular example we will consider a signal of the form of a region of elevated expectation in a linear environment with a Gaussian white noise. The elevated region is parameterized by the location of its center and by its width. The resulting field of scanning statistics is two-dimensional and Gaussian. We will consider cases where the field is smooth and a case where it is not. The classical tools for the analysis of Gaussian fields will be applied in Chapter 2 and our alternative approach will be used in Chapter 3. The Poisson approximation for the example of a scanning statistic and for a modified version of the sequential example is discussed in Chapter 4.

The second part of the book begins with a problem of intermediate difficulty. The task in that problem is to produce an approximation for the significance level of the Kolmogorov–Smirnov test and the Peacock test. Peacock's test is a generalization of the well known Kolmogorov–Smirnov nonparametric test of goodness-of-fit to higher dimensions. The Kolmogorov–Smirnov test compares the empirical distribution of a random variable to a theoretical distribution. Peacock's test, on the other hand, compares the empirical distribution of a random multivariate vector to its theoretical multivariate distribution. The analysis of the significance level of the Kolmogorov–Smirnov test is a classical exercise in many advanced statistics textbooks. Knowledge about the multivariate version of the test is not so wide spread. In Chapter 6 we give an alternative asymptotic derivation of the result for the Kolmogorov–Smirnov test on the basis of our approach. We then show how little effort is needed in order to extend the analysis to the multivariate setting.

More complex applications appear in the rest of the second part of the book. The applications presented are a reflection of the projects that I was involved with in recent years and in which the method was used.

The first application involves scanning for DNA copy number variations. Most of the genetic material in somatic cells comes in two copies, one originating from the mother and the other from the father. Occasionally, a segment of the DNA may be missing or may have multiple copies resulting in a copy number different than 2. A scanning statistic for detecting such intervals using genetic measurements generated from a sample can be constructed. The resulting scanning problem is not unlike the simple example that is used in the first part of the book. However, the produced two-dimensional field is not Gaussian, although it is asymptotically so. This problem is analyzed in Chapter 7.

The second example combines a scanning statistic with a sequential tool for change-point detection. This example involves a scenario in which an image is scanned for a signal of a specific structure. The specifications of the signal are not known nor is the time in which it will appear, if at all. The goal is to identify the emergence of the signal as fast as possible after it appeared, but not to do so before it did. The noise involved is Gaussian but the statistics that are used, and hence the associated random field, are not. The investigation of the properties of an appropriate stopping time that may be used to detect the emergence of a signal is carried out in Chapter 8.

The third application is discussed in Chapter 9. The issue in that application is the design of a buffer that is large enough to store packets waiting to be transmitted in an outgoing communication line. These packets arrive from a large number of independent sources and the outgoing line is of a fixed bandwidth. A simple model will characterize an incoming source by the distribution of the duration of time that is active, the 'on' period, and the distribution of time that it is idle, the 'off' period. The size of the required buffer can be associated with a level and the probability of a buffer overflow can be associated with the probability that a random field associated with the sum of the on−off processes exceeds the level. Apart from the fact that the field is not normally distributed, but only asymptotically so, it is also the case that the characteristic behavior of local fluctuations in this case differs from the characterizations in the other examples. This local behavior is specified by a parameter we denote by α that may take a value between 0 and 2 in general, and between 1 and 2 in the specific case. The local behavior is characterized by constants known as Pickands' constants that depend on this parameter α.

The Pickands' constants emerged as part of the development of the classical double-sum method for analyzing Gaussian fields and were defined a long time ago. However, their numerical evaluation remained an elusive problem and only crude upper and lower bounds to their value existed. In Chapter 10 we will explain how the representation that emerges from the method can be used in order to evaluate the constants efficiently. With this last application we conclude the book.