2

Random Fields

2.1 Stochastic Processes and Random Fields

As you read in the Preface, for us a *random field* is simply a stochastic process, taking values in a Euclidean space, and defined over a parameter space of dimensionality at least one. Actually, we shall be rather loose about exchanging the terms 'random field' and 'stochastic process'. In general, we shall use 'field' when the geometric structure of the parameter space is important to us, and shall use 'process' otherwise.

We shall usually denote parameter spaces by either T or M, generally using T when the parameter space is simple Euclidean domain, such as a cube, and M when refering to manifolds, or surfaces. Elements of both T and M will be denoted by t, in a notation harking back to the early days of stochastic processes when the parameter was always one-dimensional 'time'.

Of course, we have yet to define what a stochastic process is. To do this properly, we should really talk about notions such as probability spaces, measurability, separability, and so forth, as we did in *RFG*. However, we shall send the purist to *RFG* to read about such things, and here take a simpler route.

Definition 2.1.1. *Given a parameter space T, a stochastic process f over T is a collection of random variables*

$$\{f(t) : t \in T\}.$$

If T is a set of dimension N, and the random variables $f(t)$ are all vector valued of dimension d, then we call the vector valued random field f a (N, d) random field.

Note that while in what follows we shall adopt the standard convention of denoting random variables by upper case Latin characters, we shall use lower case to denote random processes. The reasons for this will become clear later, since we shall need objects such as $X(t)$ to denote geometric objects.

Examples of random fields abound, and we have already seen a few in Chapter 1, but perhaps the easiest to visualise is given by the ocean surface. In fact, this one example yields many, which together give a good introduction to many of the basic concepts of random field theory.

The height of an ocean surface above some nominal mean plane is, obviously, a function of both time and space, and so we acknowledge this by representing it as

$$f(t,x), \qquad t \in [0,\infty) \equiv \mathbb{R}_+, \quad x \in \mathbb{R}^2.$$

It is clear that the dependencies of f on the two parameters t and x are quite different, so, for the moment, let us fix a time t and consider only the ocean surface at this time, which we denote by $f(x)$.

The behavior of $f(x)$ as a function of x, however, depends very much on where the ocean is being observed. For example, at a sea shore we see mainly waves, and so f has the structure of a cosine wave (cf. Figure 2.4.1) with local amplitudes and wavelengths that have been perturbed in a random fashion. However, despite additional random perturbations, these waves generally have a very well defined direction.

The same would not be true if we were to observe an ocean at its center, on a day with little wind. While we would still see waves, their directions would not be so well defined, and in fact would change randomly with time. This lack of preferred direction is called *isotropy*, a concept that we shall meet more formally in Section 2.4.7.

While waves at the sea shore do not exhibit this behavior, it is true that if they are neither very close nor very far from the shoreline, waves at any two positions behave much the same, in that the random perturbations to local amplitude and wavelength are stochastically similar, and the same dominant direction is preserved. This property is known as *stationarity*, and we shall look at it in more detail in Section 2.4.

Adding time back into the picture, you may now want to think about whether or not ocean surfaces are ever going to be stationary or isotropic in a joint space-time parameter space. (Answers will be given in Section ?????.) It is clear, however, that the different axes in this three-dimensional parameter space no longer play interchangeable roles, as, for example, the two components of space do for an ocean near its center, for a fixed time on a windless day.

The ocean surface also provides some very natural examples of vector valued random fields. Over half a century ago, Michael Longuet-Higgins, in a series of works that gave birth to the study of the stochastic properties of sea surfaces (e.g. [59, 60]) studied the $(2,2)$ gradient random field

$$\nabla f (x_1, x_2) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right),$$

with the aim of understanding the 'specular points' of random waves, those points x at which the two components of $\nabla f(x)$ take given values. For exam-

ple, depending on where one is standing on the sea shore, on where the sun is, and what the particular values of ∇f are, these might be the points at which the sea surface reflects the sun into one's eyes.

We shall meet many additional examples as we progress through this book, but the sea surface is probably the easiest and canonical example on which to try out most new ideas.

Before leaving this introductory section, there is one rather important, theoretical point that needs to be made. While Definition 2.1.1 may effectively define stochastic processes/random fields, it does not really tell us how to differentiate between them. A celebrated theorem of Kolmogorov, known as his *Consistency* or *Existence Theorem*, says that the distributional properties of a (N, d) random field over T are determined by its *finite-dimensional distributions*[1]. These are the distributions

$$\mathbb{P}\left\{ f(t_1) \in B_1, \ldots, f(t_n) \in B_n \right\}, \tag{2.1.1}$$

for all $n \geq 1$ and all collections $\{t_j\}_{1 \leq j \leq n}$ and Borel $\{B_j\}_{1 \leq j \leq n}$ with $t_j \in T$ and $B_j \in \mathbb{R}^d$. If f is a process which possesses joint probability densities, then the probabilities in (2.1.1) can be expressed, in a self-explanatory notation, as

$$\int_{B_1} \ldots \int_{B_n} p_{t_1, \ldots, t_n}(x_1, \ldots, x_n) \, dx_1 \ldots dx_n, \tag{2.1.2}$$

and so Kolmogorov's theorem reduces to demanding that we know these densities.

2.2 Gaussian and Gaussian Related Random Fields

At the core of this book will be Gaussian and Gaussian-related random fields, and so it is appropriate that we define them before all others[2]. At the same time, we shall take the opportunity to collect a number of basic results about univariate and multivariate Gaussian random variables. We imagine that most readers will be familiar with these, and so can skip the next subsection on first reading, returning to it later for notational conventions.

2.2.1 Gaussian Variables

A real-valued random variable X is said to be *Gaussian* (or *normally distributed*) if it has the density function

---

[1] Of course, Kolmogorov's theorem only holds under certain regularity conditions. However, since these can be found in any mathematically rigorous textbook on stochastic processes, we shall let you search for them there.

[2] In fact, we already did this informally in Section 1.2, but now the time has come to do it properly.

$$\varphi(x) \overset{\Delta}{=} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/2\sigma^2}, \qquad x \in \mathbb{R},$$

for some $m \in \mathbb{R}$ and $\sigma > 0$. It is elementary calculus that the mean of X is m and the variance σ^2, and that the characteristic function is given by

$$\phi(\theta) = \mathbb{E}\left\{e^{i\theta X}\right\} = e^{i\theta m - \sigma^2\theta^2/2}.$$

We abbreviate this by writing $X \sim N(m, \sigma^2)$. The case $m = 0$, $\sigma^2 = 1$ is rather special and in this situation we say that X has a *standard* normal distribution. In general, if a random variable or process has zero mean we call it *centered*.

Since the indefinite integral of φ is not a simple function, we also need notation (Φ) for the distribution function and (Ψ) for the tail probability function of a standard normal variable, so that

$$\Phi(x) \overset{\Delta}{=} 1 - \Psi(x) \overset{\Delta}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du. \tag{2.2.1}$$

While Φ and Ψ may not be explicit, there are simple, and rather important, bounds which hold for every $x > 0$ and become sharp very quickly as x grows. In particular, in terms of Ψ, we have

$$\left(\frac{1}{x} - \frac{1}{x^3}\right)\varphi(x) < \Psi(x) < \frac{1}{x}\varphi(x). \tag{2.2.2}$$

(See Exercise 2.8.1.)

An $\mathbb{R}^d$-valued random variable $X = (X_1, \ldots, X_d)$ is said to be *multivariate Gaussian* if, for every $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d$, the real valued variable $\langle \alpha, X \rangle = \sum_{i=1}^{d} \alpha_i X_i$ is Gaussian. In this case there exists a mean vector $m \in \mathbb{R}^d$ with $m_j = \mathbb{E}\{X_j\}$ and a non-negative definite[3] $d \times d$ covariance matrix C, with elements $c_{ij} = \mathbb{E}\{(X_i - m_i)(X_j - m_j)\}$, such that the probability density of X is given by

$$\varphi(x) = \frac{1}{(2\pi)^{d/2}|C|^{1/2}} e^{-\frac{1}{2}(x-m)C^{-1}(x-m)'}, \tag{2.2.3}$$

where $|C| = \det C$ is the determinant[4] of C. Consistently with the one-dimensional case, we write this as $X \sim N(m, C)$, or $X \sim N_d(m, C)$ if we need to emphasise the dimension.

---

[3] A $d \times d$ matrix C is called non-negative definite, or positive semi-definite, (positive definite) if $xCx' \geq 0$ (> 0) for all $x \in \mathbb{R}^d$. A function $C : T \times T \to \mathbb{R}$ is called non-negative (positive) definite if the matrices $(C(t_i, t_j))_{i,j=1}^{n}$ are non-negative (positive) definite for all $1 \leq n < \infty$ and all $(t_1, \ldots, t_n) \in T^n$.

[4] Two important comments on notation: Throughout the book, vectors are taken to be row vectors and a prime indicates transposition. The inner product between x and y in $\mathbb{R}^d$ is usually denoted by $\langle x, y \rangle$, (x, y) or, occasionally, by $x \cdot y$ and even xy when there is no chance of confusion.

In view of (2.2.3) we have that Gaussian distributions are completely determined by their first and second order moments and that uncorrelated Gaussian variables are independent. Both of these facts will be of crucial importance later on.

While the definitions are fresh, note for later use that it is relatively straightforward to check from (2.2.3) that the characteristic function of a multivariate Gaussian X is given by

$$\phi(\theta) = \mathbb{E}\{e^{i\langle\theta,X\rangle}\} = e^{i\langle\theta,m\rangle - \frac{1}{2}\theta C\theta'}, \qquad (2.2.4)$$

where $\theta \in \mathbb{R}^d$.

An immediate consequence of either (2.2.3) or (2.2.4) is that if A is any $d \times d$ matrix and $X \sim N_d(m, C)$, then

$$XA \sim N(mA, A'CA). \qquad (2.2.5)$$

A judicious choice of A (see Exercise 2.8.2) then allows us to compute conditional distributions as well. If $n < d$ make the partitions

$$X = (X^1, X^2) = ((X_1, \dots, X_n), (X_{n+1}, \dots X_d)),$$
$$m = (m^1, m^2) = ((m_1, \dots, m_n), (m_{n+1}, \dots m_d)),$$
$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix},$$

where C_{11} is an $n \times n$ matrix. Then each X^i is $N(m^i, C_{ii})$ and the conditional distribution of X^i given X^j is also Gaussian, with mean vector

$$m_{i|j} = m^i + (X^j - m^j)C_{jj}^{-1}C_{ji} \qquad (2.2.6)$$

and covariance matrix

$$C_{i|j} = C_{ii} - C_{ij}C_{jj}^{-1}C_{ji}. \qquad (2.2.7)$$

2.2.2 Gaussian Fields

We can now define a real valued *Gaussian (random) field* or *Gaussian (random) process* to be a random field f on a parameter set T for which the (finite

---

Secondly, at various places we shall use the notation $|\cdot|$ to denote any of 'absolute value', 'Euclidean norm', 'determinant' or 'Lebesgue measure', depending on the argument, in a natural fashion. The notation $\|\cdot\|$ is used only for either the norm of complex numbers or for special norms, when it usually appears with a subscript. This is unless it is used, as in Section 2.6 in particular, as $\|f\|$ to denote the supremum of a function f, which is not a norm at all. Despite this multitude of uses of a simple symbol, its meaning should always be clear from the context.

dimensional) distributions (2.1.1) of $(f_{t_1}, \ldots, f_{t_n})$ are multivariate Gaussian for each $1 \leq n < \infty$ and each $(t_1, \ldots, t_n) \in T^n$.

Since multivariate Gaussian distributions are determined by means and covariances, it is immediate that Gaussian random fields are determined[5] by their mean and covariance functions[6], defined by

$$m(t) = \mathbb{E}\{f(t)\} \tag{2.2.8}$$

and

$$C(s,t) = \mathbb{E}\left\{(f(s) - m(s))\,(f(t) - m(t))\right\}. \tag{2.2.9}$$

In fact, this is one of the main reasons, beyond ubiquitous but not always justified appeals to the central limit theorem, that Gaussian processes are such popular and useful choices for models for random processes on general spaces.

Multivariate Gaussian fields taking values in $\mathbb{R}^d$ are fields for which $\langle \alpha, f_t \rangle$ is a real valued Gaussian field for every $\alpha \in \mathbb{R}^d$. In this case, $m(t)$ takes values in $\mathbb{R}^d$ and the covariance function of (2.2.9) is replaced by a function whose values are non-negative definite, $d \times d$, matrices. In particular,

$$C(s,t) = \mathbb{E}\left\{(f(s) - m(s))'\,(f(t) - m(t))\right\}, \tag{2.2.10}$$

so that the individual elements of C are given by

$$C_{ij}(s,t) = \mathbb{E}\left\{(f_i(s) - m_i(s))\,(f_j(t) - m_j(t))\right\}.$$

As in the real valued case, for Gaussian f the vector function m and the matrix function C determine all of its statistical properties.

2.2.3 Gaussian Related Fields

As convenient as Gaussian random fields may be for the mathematician, in that the form of the multivariate Gaussian distribution makes many things computable in the Gaussian case that are uncomputable otherwise, it would

---

[5] In fact, one can also go in the other direction as well. It is a consequence of this structure of Gaussian densities and the Kolmogorov consistency theorem that, given *any* set T, a function $m : T \to \mathbb{R}$, and a non-negative definite function $C : T \times T \to \mathbb{R}$ (see the following footnote) there exists a Gaussian process on T with mean function m and covariance function C.

[6] Covariance functions are always positive semi-definite; cf. Exercise 2.8.4. In the vector valued case there are two types of semi-definiteness to worry about. On the one hand, for each fixed pair s, t, the matrix $C(s,t)$ is positive semi-definite. On the other hand, for each pair i, j, the function C_{ij} is positive semi-definite, as a function, on $T \times T$.

be a poor modeller or statistician who would work in a Gaussian scenario only.

Leaving the Gaussian scenario is, however, not all that easy to do, and so we shall leave it in a fashion that, while somewhat limited, turns out to be broad enough to cover many, if not most, statistical applications. To be more precise, we shall call a random field $f : T \to \mathbb{R}^d$ a *Gaussian related* field if we can find a vector valued Gaussian random field,

$$g(t) = (g_1(t), \ldots, g_k(t)) : T \to \mathbb{R}^k,$$

and a function

$$F : \mathbb{R}^k \to \mathbb{R}^d,$$

such that f has the same finite dimensional distributions as $F(g)$. That is,

$$f(t) \overset{\mathcal{L}}{=} F(g(t)) = F\left(g_1(t), \ldots, g_k(t)\right), \tag{2.2.11}$$

where $\overset{\mathcal{L}}{=}$ indicates equivalence in distribution.

When $k = 1$, or, in general $k = d$ and F is invertible, then the corresponding Gaussian related process is not much harder to study than the original Gaussian one. After all, what happens at the level u for f is precisely what happens at the uniquely defined level $F^{-1}(u)$ for g. However, in other cases, f can provide a process that is qualitatively different to g.

For example, suppose that the g_j are centred and of unit variance, and consider the following three choices for F, where in the third we set $k = n+m$.

$$\sum_1^k x_i^2, \qquad \frac{x_1 \sqrt{k-1}}{(\sum_2^k x_i^2)^{1/2}}, \qquad \frac{m \sum_1^n x_i^2}{n \sum_{n+1}^{n+m} x_i^2}. \tag{2.2.12}$$

The corresponding random fields are known as χ^2 fields with k degrees of freedom, the T field with $k - 1$ degrees of freedom, and the F field with n and m degrees of freedom. All of these will appear as important models in the applications we shall be looking at in Part IV of the book, and so we shall not attempt to motivate their usefulness now beyond stating that they are as fundamental to the statistical applications of random field theory as are their univariate distributions to standard statistical theory.

Before we leave Gaussian related processes, we want to make one comment about some interesting geometric problems associated with them.

Recall that one of the central themes of this book will be the study and application of the geometric properties of the excursion sets[7] of f over T and above the level u, viz.

---

[7] One final reminder, particularly for the physicists among our readers, that in much of the mathematical literature, and in most of physics, excursion sets are known as *nodal domains*.

$$A_u(f,T) = \{t \in T : f(t) \geq u\}.$$

For a Gaussian related field of the form (2.2.11), these can be rewritten as

$$\begin{aligned} A_u(f,T) \;=\; A_u(F(g),M) &= \{t \in T : (F \circ g)(t) \geq u\} \qquad (2.2.13)\\ &= \{t \in T : g(t) \in F^{-1}[u,\infty)\}\\ &= T \cap g^{-1}(F^{-1}[u,+\infty)). \end{aligned}$$

Thus, the excursion set of a *real valued* non-Gaussian $f = F \circ g$ above a level u is equivalent to the excursion set for a *vector valued* Gaussian g in $F^{-1}[u,\infty)$. This set will generally have an interesting (albeit deterministic) structure of its own. For example, in the case of a χ_k^2 field it will be the complement, in $\mathbb{R}^k$, of a ball of radius $\sqrt{u}$ and centered at the origin. Thus, regardless of the simplicity of the underlying parameter space T, the moment we leave real valued Gaussian processes and turn to their Gaussian related relatives, geometry is going to be important to us.

   We shall return to this point later in the book, once we know a little more about random fields and about geometry.

2.3 Examples of Random Fields

Before looking at specific examples, we need to make one small digression, to introduce the notions of stationarity and isotropy.

2.3.1 Stationary and Isotropic Fields

To formally define stationarity, which we actually already met in Section 1.2 and will learn a lot more about in Section 2.4 below, recall the definitions (2.2.8)–(2.2.10) of the mean and covariance functions of a stochastic process. These lead to two important definitions.

Definition 2.3.1. *Suppose that f is an (N,d) random field defined over all of $\mathbb{R}^N$. Suppose furthermore that the mean function $m(t)$ is constant, and that the covariance function $C(s,t)$ is a function of the difference $t - s$ only. Then we say that f is homogeneous or stationary.*

   If it is also true that $C(s,t)$ is a function of the Euclidean distance $|t - s|$ only, then we say that f is also isotropic[8].

   Two comments are called for following on from this definition. The first is that, in fact, we have only really defined what is usually called *weak* or

---

[8] Isotropy can also be defined in the non-stationary case, the defining property then being that the distribution of f is invariant under rotation. Under stationarity, this is equivalent to the current definition. We shall treat isotropy *only* in the scenario of stationarity.

second order stationarity , which depends only on first and second moments. A stronger form of stationarity is that of the finite dimensional distributions, that requires that joint distributions of the form

$$\{f(t_1 + \tau), \ldots, f(t_n + \tau)\} \tag{2.3.1}$$

be independent of τ, for all $n \geq 1$ and all $t_j \in \mathbb{R}^N$. It is obvious that the two notions coincide when f is Gaussian or Gaussian related.

The second is more an issue of notation than a comment. When f is stationary, we shall generally abuse notation and write

$$C(s, t) = C(s - t). \tag{2.3.2}$$

While the first C here is defined on $T \times T$, the second is defined on T only. In the isotropic case, we shall go even further, and write

$$C(s, t) = C(|s - t|),$$

where the second C is defined on $\mathbb{R}_+$.

2.3.2 Random Polynomials

What appear at first sight to be the simplest of stochastic processes are the *random polynomials* on $\mathbb{R}$, functions of the form

$$f(t) = \sum_{k=0}^{n} \xi_k t^k,$$

for some finite n, where the random coefficients ξ_k are generally taken to be either uncorrelated or independent. The random field version, on $\mathbb{R}^N$, is given by

$$f(t) = \sum_{k=0}^{n} \sum_{j_1 + \cdots + j_N = k} \xi_{j_1 \ldots j_N} t_1^{j_1} \cdots t_N^{j_N},$$

again with uncorrelated or independent coefficients.

Perhaps surprisingly (since polynomials really are the simplest of all non-random functions) random polynomials are actually very hard to study. For a start, they are clearly non-stationary, not even having constant variance. Nevertheless, despite this, or perhaps because of it, they are at the centre of considerable research activity at the moment, interest concentrating on the number of points in the zero set $\{t : f(t) = 0\}$ in the one-dimensional case, and the structure of this set in higher dimensions.

We shall take a look at this zero set using the tools we have developed, much later, in Section 12.1.

2.3.3 The Cosine Process

If random polynomials is where the theory of random processes might have begun, the true development of the theory began with what is perhaps the grandfather of all stationary processes, the so-called *cosine random process* on $\mathbb{R}$. It is defined as

$$f(t) \stackrel{\Delta}{=} \xi \cos \lambda t + \xi' \sin \lambda t, \tag{2.3.3}$$

where ξ and ξ' are uncorrelated, equidistributed, random variables and λ is a positive constant.

The cosine process provides the simplest version of the the wave example that we met in Section 2.1, and we shall soon see that it also provides the elementary building block for general stationary processes.

It is elementary trigonometry to see that the cosine process can also be written as

$$f(t) = R \cos(\lambda(t - \theta)), \tag{2.3.4}$$

where $R^2 = \xi^2 + (\xi')^2 \geq 0$ and $\theta = \arctan(\xi/\xi') \in (-\pi, \pi]$, from whence the name 'cosine process'. It is now obvious that the cosine process is no more than a cosine function with fixed wavelength $2\pi/\lambda$, but random amplitude and phase. Furthermore, supposing for convenience that $\mathbb{E}\{\xi\} = 0$, we have that its covariance function is given by

$$\begin{aligned}
C(s,t) &= \mathbb{E}\{f(s)f(t)\} \\
&= \mathbb{E}\{(\xi \cos \lambda s + \xi' \sin \lambda s)(\xi \cos \lambda t + \xi' \sin \lambda t)\} \\
&= \mathbb{E}\{\xi^2\} \cos(\lambda(t - s)),
\end{aligned}$$

on using the fact that ξ and ξ' are uncorrelated and equidistributed. Consequently, regardless of the distribution of ξ, the cosine process is stationary. In fact, as we shall soon see, it is the archetype of all stationary processes, all of which can be represented as a sum of cosine processes.

One of the nice aspects of the cosine process is that many things that are either difficult or impossible to compute for more general processes can be computed exactly, and from first principles, once some assumptions are made on the distribution of ξ. We shall therefore now assume that ξ and ξ' are independent, Gaussian variables, with zero mean and common variance σ^2.

Now consider, for $u > 0$, the exceedence probability

$$\mathbb{P}\left\{ \sup_{0 \leq t \leq T} f(t) \geq u \right\}, \tag{2.3.5}$$

which you know from reading the Preface is one of the important quantities of this book.

Under the Gaussian assumption, R^2 has an exponential distribution with mean $1/(2\sigma^2)$, θ has a uniform distribution on $(-\pi, \pi]$, and R and θ are

independent. We can use this information to compute some exceedence probabilities directly, and start by defining

$$N_u = N_u(f, T) = \#\{t \in [0, T] : f(t) = u \text{ and } df(t)/d(t) > 0\}.$$

Then N_u is known as the *number of upcrossings* by f of the level u in time $[0, T]$, and the points being counted as *upcrossings*. Upcrossings and their generalizations will play a major rôle throughout this book.

It is trivial to see that the exceedence probability that we are after can now be written as

$$\mathbb{P}\left\{\sup_{0 \leq t \leq T} f(t) \geq u\right\} = \mathbb{P}\left\{f(0) \geq u\right\} + \mathbb{P}\left\{f(0) < u, N_u \geq 1\right\} \quad (2.3.6)$$

$$= \Psi\left(\frac{u}{\sigma}\right) + \mathbb{P}\left\{f(0) < u, N_u \geq 1\right\}.$$

We now restrict attention to the case $T \leq \pi/\lambda$, in which case the event $\{f(0) \geq u, N_u \geq 1\}$ is empty, implying that

$$\mathbb{P}\left\{f(0) < u, N_u \geq 1\right\} = \mathbb{P}\left\{N_u \geq 1\right\}.$$

Again using the fact that $T < \pi/\lambda$, note that N_u is either 0 or 1. In order that it be 1, two independent events must occur. Firstly, we must have $R > u$, with probability $e^{-u^2/2\sigma^2}$. Secondly (draw a picture) θ must fall in an interval of length λT, so that the final result is

$$\mathbb{P}\left\{\sup_{0 \leq t \leq T} f(t) \geq u\right\} = \Psi\left(\frac{u}{\sigma}\right) + \frac{\lambda T}{2\pi\sigma} e^{-u^2/2\sigma^2}, \quad (2.3.7)$$

and the probability density of the supremum is given by

$$\frac{1}{\sigma}\varphi\left(\frac{u}{\sigma}\right) + \frac{\lambda T u}{2\pi\sigma^2} e^{-u^2/2\sigma^2}. \quad (2.3.8)$$

This computation was so simple, that one is tempted to believe that it must be easy to extend to many other processes. In fact, this is not the case, and the cosine process and field (see below) are the *only* differentiable, stationary, Gaussian processes for which the exceedence probabilities are explicitly known.

However, before we leave it, we can use it to nevertheless motivate a more general approach. Note first that since, as noted above, N_u is either 0 or 1 when $T < \pi/\lambda$, we can rewrite (2.3.6) as

$$\mathbb{P}\left\{\sup_{0 \leq t \leq T} f(t) \geq u\right\} = \Psi\left(\frac{u}{\sigma}\right) + \mathbb{E}\{N_u\}. \quad (2.3.9)$$

Thus, rather than arguing as above, we could concentrate on finding an expression for the mean number of upcrossings.

More importantly, note that for any T, and, indeed, for *any differentiable random process*, the above argument always gives

$$\mathbb{P}\left\{\sup_{0 \leq t \leq T} f(t) \geq u\right\} \leq \mathbb{P}\left\{f(0) \geq u\right\} + \mathbb{E}\{N_u\}. \qquad (2.3.10)$$

Thus there would seem to be a close relationship between exceedence probabilities and level crossing rates, that actually becomes exact for the cosine process over certain intervals. In fact, since, for a one dimensional set, its Euler characteristic is given by the number of its connected components, the right hand sides of both (2.3.9) and (2.3.10) could be written as $\mathbb{E}\{\varphi(A_u(f, T))\}$, where φ is the Euler characteristic and $A_u(f, T)$ the excursion set $\{t \in [0, T] : f(t) \geq u\}$.

We shall investigate and exploit this, in detail, in Chapter 5.

2.3.4 The Cosine Field

The cosine field is a straightforward extension to $\mathbb{R}^N$ of the cosine process, and has the respresentation

$$f(t) = f(t_1, \ldots, t_N) \triangleq \frac{1}{\sqrt{N}} \sum_{k=1}^{N} f_k(\lambda_k t_k), \qquad (2.3.11)$$

where each f_k is the process on $\mathbb{R}$ given by

$$f_k(t) = \xi_k \cos t + \xi_k' \sin t.$$

The λ_k are fixed, and the ξ_k and ξ_k' are taken to be identically distributed and uncorrelated.

Again, it is a simple excercise to check that the cosine field is both stationary and isotropic (Exercise 2.8.5) but is a little harder to compute its exceedence probabilities. To see what can be done, we restrict attention to the cosine process on a rectangle of the form $T = \prod_{k=1}^{N}[0, T_k]$. Then, given the structure of the cosine field as a sum, it is immediate that

$$\sup_{t \in T} f(t) = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \sup_{0 \leq t_k \leq T_k} f_k(t).$$

If we assume that the ξ_k and ξ_k' are all independent $N(0, \sigma^2)$, then the suprema of the individual f_k are also independent. Further assuming that each $T_k \in (0, \pi/2\lambda_k]$, (2.3.7) and (2.3.8) give their individual distributions. The distribution of the supremum of the cosine field is then the convolution of these. In Exercise 2.8.6 you are asked to actually do the computation, to find that, if $p_N(u)$ is the density function of the supremum, φ the usual standard Gaussian density and $\varphi^{(k)}$ its k-derivative, then there are simple constants, C_{nk}, depending only on n and k, such that

$$(2.3.12)$$

$$p_N(u) = \varphi\left(\frac{u}{\sigma}\right) + \sum_{k=1}^{N} (-1)^k C_{nk} \varphi^{(k)}\left(\frac{u}{\sigma}\right) \sum_{j_1 \ldots j_k} \prod_{i=1}^{k} \frac{\lambda_{j_i} T_{j_i}}{\sigma}.$$

The inner sum here is over the $\binom{N}{k}$ subsets of size k of $\{1, \ldots, N\}$.

This is actually the archetype of a far more general result of a geometrical nature. To see it in its easiest formulation, firstly take $\sigma^2 = 1$ and $\lambda_k = \lambda$ for all k, so that the product over the λ_{j_i} becomes merely λ^k, and consider the sums

$$\sum_{j_1 \ldots j_k} T_{j_1} \cdots T_{j_k}.$$

When $k = N$, this gives the volume of the parameter space. When $k = N - 1$, it gives its surface area. In fact, if you recall the $\mathcal{L}_k$ of Chapter 1, then this is precisely what these sums are, and we now define[9] the $\mathcal{L}_k$ of N-dimensional rectangles to be

$$\mathcal{L}_j\left(\prod_1^N [0, T_j]\right) = \sum_{j_1 \ldots j_k} T_{j_1} \cdots T_{j_k}, \qquad (2.3.13)$$

where the sum is over the $\binom{N}{k}$ *distinct* choices of k indices between 1 and N. Thus (2.3.12) becomes[10]

$$p_N(u) = \sum_{k=0}^{N} (-1)^k C'_{nk} \varphi^{(k)}(u) \lambda^k \mathcal{L}_j(T). \qquad (2.3.15)$$

Integrating over u, and applying some asymptotics[11], one finds that

---

[9] Precisely why this is a useful definition will be made clear in Chapter 3. See also Exercise 3.6.4.

[10] Actually, (2.3.15) still makes sense if the λ_j are not all identical and $\sigma^2 \neq 1$ but we define the $\mathcal{L}_k$ by

$$\mathcal{L}_k\left(\prod_1^N [0, T_j]\right) = \sum_{j_1 \ldots j_k} \prod_{i=1}^{k} \frac{\lambda_{j_i} T_{j_i}}{\sigma}. \qquad (2.3.14)$$

This is an example of what we shall call the 'inducing of a Riemannian metric on $\mathbb{R}^N$ via the random field f', something that we shall meet up with in some detail in Chapter 4 and that will be an important tool in later parts of the book.

[11] The asymptotics, while not being intrinsically difficult, are also not easy, and this is left to you as an exercise. You can find the details in Section 2.5 of Piterbarg's important monograph [75], which is where, to the best of our knowledge, (2.3.16) appeared for the first time.

$$\mathbb{P}\left\{\sup_{t \in T} f(t) \geq u\right\} \tag{2.3.16}$$

$$= \Psi(u) + e^{-u^2/2} \sum_{k=1}^{N} C''_{nk} H_{k-1}(u) \lambda^k \mathcal{L}_j(T) + o\left(e^{-(1+\eta)u^2/2}\right),$$

for some $\eta > 0$, where the H_k are Hermite polynomials, which we shall have cause to meet many times in the future. (cf. (3.3.20) for a definition.)

It turns out that the cosine process and field are the *only* smooth (non-degenerate) Gaussian processes for which it is possible to actually compute the distribution of the supremum[12]. This means that in general we shall have to adopt other approaches, as we have mentioned before. However, the fact that there is at least one class of Gaussian fields for which the distribution of the supremum is explicitly known is extremely useful. We shall soon see that Slepian's Inequality (Theorem 2.6.3) will enable us to use the known exceedence probabilities of the cosine process and field to bound those of other Gaussian processes and fields as well.

2.3.5 Orthogonal Expansions

The idea of summing simple random fields to get more complicated ones, intrinsic to the definition of the cosine field, can be carried much further.

For example, suppose we have a collection $\{\varphi_n\}$ of real valued[13] functions on our parameter space T, and a collection of uncorellated random variables $\{\xi_n\}$ with, for convenience, mean zero and variances σ_n^2. From these, we can form the sum

$$f(t) = \sum_{n=1}^{\infty} \xi_n \, \varphi_n(t), \tag{2.3.17}$$

which will have covariance function

$$\mathbb{E}\{f(s)f(t)\} = \sum_{n=1}^{\infty} \sigma_n^2 \, \varphi_n(s)\varphi_n(t), \tag{2.3.18}$$

and variance function

---

[12] There are five non-smooth, essentially Markov, random processes on $\mathbb{R}$ for which this is possible, but no other random fields. See Footnote 2 in Chapter 4 of *RFG* for a list.

[13] We are treating only the real valued case for convenience. The extension to vector valued random fields in all that follows is straightforward. One needs only require that the φ_n map T to $\mathbb{R}^d$ and that the random variables ξ_n are replaced by random matrices ξ_n which are either independent or uncorellated in the sense that $\mathbb{E}\{\xi'_n\xi_m\}$ is the zero matrix if $n \neq m$. The covariance function (2.3.18) is then replaced by the covariance matrix, $\sum_{n=1}^{\infty} \varphi'_n(s)\mathbb{E}\{\xi'_n\xi_n\}\varphi_n(t)$, etc.

$$\mathbb{E}\{f^2(t)\} = \sum_{n=1}^{\infty} \sigma_n^2 \, \varphi_n^2(t), \qquad\qquad (2.3.19)$$

as long as everything that needs to converge does in fact do so.

Random fields with the above structure are attractive to theoreticians and practitioners alike. For the theoretician, the mathematical structure of a summation of uncorellated or independent variables is easy to handle. For the practitioner, this structure is even more useful. For a start, it generally matters little to a practitioner if the summations are finite or infinite. After all, even if they are infinite, the fact that they converge must mean that all but a finite number of terms are very small, or at least too small to see on a computer screen, so that they can be taken to be finite. Doing so means that one needs only a finite number of *deterministic* functions, and a finite number of random variables, to understand an entire random field. This, of course, has significant implications for simulation, which we shall look at in some detail in Chapter 7.

For the moment, however, we shall concentrate on the more theoretical aspects of fields defined by such sums. A particularly interesting question is how general they are.

Clearly, if the random coefficients in (2.3.17) are Gaussian, then so is the sum. However, there is also a converse, which states that *every* centered Gaussian process with a continuous covariance function has an expansion[14] of the form (2.3.17), in which the the ξ_n are i.i.d. $N(0,1)$, and the φ_n are certain functions on T determined by the covariance function C of f, and possess certain orthogonality properties[15].

---

[14] In general, the convergence in the expansion (2.3.17) is in $L^2(\mathbb{P})$ for each $t \in T$, but, if f is Gaussian and continuous with probability one, then the convergence is uniform over T, and with probability one. For more details – and there are many – see Chapter 3 of *RFG* for an introduction to a rich and beautiful theory.

[15] Orthogonality is in the so-called *reproducing kernel Hilbert space* (RKHS) of f. In essence, the RKHS is made up of functions that have about the same smoothness properties that $C(s,t)$ has, as a function in t for fixed s, or vice versa. To construct it, start with

$$S = \Big\{ u : T \to \mathbb{R} : u(\cdot) = \sum_{i=1}^{n} a_i C(s_i, \cdot), \ a_i \text{ real}, \ s_i \in T, \ n \geq 1 \Big\},$$

and define an inner product on S by setting

$$(u, v)_H = \Big(\sum_{i=1}^{n} a_i C(s_i, \cdot), \sum_{j=1}^{m} b_j C(t_j, \cdot) \Big)_H = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j C(s_i, t_j).$$

Note that this inner product has the unusual property that

$$(u, C(t, \cdot))_H = \Big(\sum_{i=1}^{n} a_i C(s_i, \cdot), C(t, \cdot) \Big)_H = \sum_{i=1}^{n} a_i C(s_i, t) = u(t).$$

To see how this works, we consider the case in which the parameter set T is a compact domain in $\mathbb{R}^N$. The corresponding expansion of f, given by (2.3.22) below, is then known as the *Karhunen-Loève expansion*. To set up the expansion, suppose that f has covariance function C and define an operator $\mathcal{C}$, taking the space of square integrable functions on T to itself, by

$$(\mathcal{C}\psi)(t) = \int_T C(s,t)\psi(s)\,ds.$$

Suppose that $\lambda_1 \geq \lambda_2 \geq \ldots$, and $\psi_1, \psi_2, \ldots$, are, respectively, the (ordered) eigenvalues and normalised eigenfunctions of the operator. That is, the λ_n and ψ_n solve the integral equation

$$\int_T C(s,t)\psi(s)\,ds \;=\; \lambda\psi(t), \tag{2.3.20}$$

with the normalisation

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

These eigenfunctions lead to a natural expansion of C, known as *Mercer's Theorem*, which states that

$$C(s,t) = \sum_{n=1}^{\infty} \lambda_n \psi_n(s)\psi_n(t), \tag{2.3.21}$$

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

The Karhunen-Loève expansion of f is then obtained by setting $\varphi_n = \lambda_n^{\frac{1}{2}}\psi_n$ in the expansion (2.3.17), so that

$$f_t = \sum_{n=1}^{\infty} \lambda_n^{\frac{1}{2}}\, \xi_n\, \psi_n(t), \tag{2.3.22}$$

where the ξ_n are i.i.d. $N(0,1)$. For an example of how this approach works for Brownian motion on the line, see Exercise 2.8.10.

As simple as this approach might seem, it is generally limited by the fact that it is usually not easy to analytically solve the integral equation (2.3.20), even for parameter sets as simple as the unit interval. Moreover, the handful of examples for which it can be solved involve non-differentiable processes, well outside our range of interests. Nevertheless, it is nice to know that such expansions exist, and, as we shall soon see, they have a wealth of applications.

This is called the *reproducing kernel* property, and the closure of S under the corresponding norm is caled the RKHS of f or of C.

In particular, the same approach will work if T is replaced by a finite set of points[16]. Then the integral in (2.3.20) becomes a finite sum, as does the expansion (2.3.22), and eigenfunctions become eigenvectors. Thus, even if no analytic solution is possible, we now have an eigenvector problem that will be easy to solve numerically, as long as the number of points is not too large.

We shall return to this in Chapter 7, when we discuss simulations. However, the important thing to note, already at this stage, is that, whether we are in the finite or infinite scenario, in order to simulate f via the Karhunen-Loève expansion one needs *only once* to solve a deterministic eigenfunction or eigenvector problem, and then simulate nothing more difficult than a set of i.i.d. standard normals.

2.4 Stationarity and Isotropy

Although we have already met both stationarity and isotropy more than once, the time has now come to look at both of them in some detail. Understanding the structure of random fields with these properties is imperative for working in the subject at all, and understanding the analytic properties of stationary and isotropic covariance functions will be crucial for carrying out many of the calculations we shall need to make later.

We imagine that most readers of this book will be familiar with most of the material of this section, and so will skip it and return only when specific details are required later.

The most important classic results of this chapter are the spectral distribution and spectral representation theorems for $\mathbb{R}^N$ which we shall meet soon in Section 2.4.2. However, the most important results for us will be some of the consequences of these theorems for relationships between spectral moments, and these are concentrated in Section 2.4.3. This is the one part of this section that you will almost definitely need to come back to, even if you have decided that you are familiar enough with stationarity to skip the remainder for the moment.

2.4.1 Basic Definitions

Although our primary interest lies in the study of real valued random fields it is mathematically more convenient to discuss stationarity in the framework of complex valued processes[17]. Hence, unless otherwise stated, we shall assume

---

[16] Think of a continuous parameter space being replaced by a grid of points, and the smooth random field being replaced by the values of the original field on this grid. This kind of discretization is, of course, what one always has on a computer.

[17] The main reason for this is that it is so much more convenient to multiply complex exponentials than it is to multiply sines and cosines. No damage is done, since in dealing with real valued processes one just restricts to complex processes whose imaginary parts have mean and variance both zero.

throughout this section that $f(t) = (f_R(t) + if_I(t))$ takes values in the complex plane $\mathbb{C}$ and that

$$\mathbb{E}\left\{\|f(t)\|^2\right\} = \mathbb{E}\left\{f_R^2(t) + f_I^2(t)\right\} < \infty.$$

(Both f_R and f_I are, obviously, to be real valued.) The parameter space of f will be $\mathbb{R}^N$ although it could be any space with a group structure without much change in notation. As for a definition of normality in the complex scenario, we first define a complex random variable to be Gaussian if the vector of its two components is bivariate Gaussian[18]. A complex process f is Gaussian if $\sum_i \alpha_i f_{t_i}$ is a complex Gaussian variable for all sequences $\{t_i\}$ in T and complex $\{\alpha_i\}$.

Since $f_t \in \mathbb{C}$, it follows that the mean function $m(t) = \mathbb{E}\{f(t)\}$ is also complex valued, as is the covariance function, which we redefine for the complex case as

$$C(s,t) \overset{\Delta}{=} \mathbb{E}\left\{[f(s) - m(s)]\left[\overline{f(t)} - \overline{m(t)}\right]\right\}, \tag{2.4.1}$$

with the bar denoting complex conjugation.

Two basic properties of covariance functions follow immediately from (2.4.1).

(i) $C(s,t) = \overline{C(t,s)}$, which becomes the simple symmetry $C(s,t) = C(t,s)$ if f (and so C) is real valued.

(ii) For any $k \geq 1$, $t_1, \ldots, t_k \in T$ and $z_1, \ldots, z_k \in \mathbb{C}$, the Hermitian form $\sum_{i=1}^k \sum_{j=1}^k C(t_i, t_j) z_i \overline{z}_j$ is always real and non-negative. We summarise this, as before, by saying that C is non-negative definite.

Stationarity and homogeneity of f are defined as in the real valued case, whether they be strict or second order, the latter requiring that the mean is constant and $C(s,t)$ is a function of the difference $s - t$ only.

The Gaussian situation is, however, a little different in the complex case. Whereas weakly stationary, real valued, Gaussian processes are also strongly stationary, in the complex case we also require that[19]

$$C'(s,t) = \mathbb{E}\{[f(s) - m(s)]\,[f(t) - m(t)]\}$$

is also a function only of $s - t$. If f is real valued, $C \equiv C'$ and so we are back to the usual situation that strong and weak stationarity are equivalent for Gaussian processes.

Isotropy has the same definition in the complex case as in the real one.

---

[18] It therefore follows that a complex Gaussian variable $X = X_R + iX_I$ is defined by five parameters: $\mathbb{E}\{X_I\}$, $\mathbb{E}\{X_R\}$, $\mathbb{E}\{X_I^2\}$, $\mathbb{E}\{X_R^2\}$ and $\mathbb{E}\{X_I X_R\}$.

[19] The reason for the additional condition in the complex case lies in the structure of the multivariate complex normal distribution. Following on from Footnote 18, in order to move from weak stationarity to strong stationarity we need that *all* covariances, including those between real and imaginary parts and among themselves. Knowing both C and C' is enough to do this, but knowing only one is not enough.

2.4.2 Spectral Distribution Theorem

We shall now investigate some of the basic properties of covariance functions, and what information can be easily gleaned from them, starting with an important characterization result.

The result, which dates back to Bochner [21], in the setting of (non-stochastic) Fourier analysis, is classical and a proof can be found in almost any text on Fourier analysis.

Theorem 2.4.1 (spectral distribution theorem). *A continuous function* $C : \mathbb{R}^N \to \mathbb{C}$ *is non-negative definite (i.e. a covariance function) if and only if there exists a finite measure ν on the Borel σ-field $\mathbb{B}^N$ of $\mathbb{R}^N$ such that*

$$C(t) = \int_{\mathbb{R}^N} e^{i\langle t, \lambda \rangle} \, \nu(d\lambda), \qquad (2.4.2)$$

for all $t \in \mathbb{R}^N$.

With randomness in mind, we write $\sigma^2 = C(0) = \nu(\mathbb{R}^N)$. The measure ν is called the *spectral measure* (for C) and the function $F : \mathbb{R}^N \to [0, \sigma^2]$ given by

$$F(\lambda) \triangleq \nu \left(\prod_{i=1}^{N} (-\infty, \lambda_i] \right), \qquad \lambda = (\lambda_1, \ldots, \lambda_N) \in \mathbb{R}^N, \qquad (2.4.3)$$

is called the *spectral distribution function*[20]. When F is absolutely continuous the corresponding density is called the *spectral density*. .

Note that if C is the covariance of a real valued random field, then C must also be real valued. In this case, it follows that ν must be a symmetric measure, in the sense that $\nu(A) = \nu(-A)$ for all $A \in \mathbb{B}^N$. In terms of spectral densities p, this implies that $p(x) = p(-x)$ for all $x \in \mathbb{R}^N$.

Similarly, if C is isotropic then ν must be spherically symmetric, and the spectral density must satisfy $p(x) = p(xM)$ for all $x \in \mathbb{R}^N$ and any rotation matrix M.

It is important to note that the spectral distribution theorem is a purely analytic result and would have nothing to do with random fields were it not for the fact that covariance functions are non-negative definite. Understanding of the result comes from the spectral representation theorem (Theorem 2.4.2) for which we offer two approaches, in Section 2.4.6

However, before we turn to this more powerful result, we shall collect some easy but important facts that can be gleaned from covariance functions per se.

---

[20] Of course, unless ν is a probability measure, so that $\sigma^2 = 1$, F is not a distribution function in the usual usage of the term.

2.4.3 Spectral Moments and Derivatives of Random Fields

We start by taking a closer look at spectral measures and, in particular, their moments. Given the the spectral representation (2.4.2) we define the *spectral moments*

$$\lambda_{i_1 \ldots i_N} \triangleq \int_{\mathbb{R}^N} \lambda_1^{i_1} \cdots \lambda_N^{i_N} \, \nu(d\lambda), \qquad (2.4.4)$$

for all $(i_1, \ldots, i_N)$ with $i_j \geq 0$. Assuming that the underlying random field, and so the covariance function, are real valued, so that, as described above, stationarity implies that $C(t) = \overline{C(-t)}$ and $\nu(A) = \nu(-A)$, it follows that the odd ordered spectral moments, when they exist, are zero; i.e.

$$\lambda_{i_1 \ldots i_N} = 0, \qquad \text{if } \sum_{j=1}^{N} i_j \text{ is odd.} \qquad (2.4.5)$$

There are two ways to understand the meaning of the spectral moments of even order. One of these has to do with the 'high frequency components' of f, and relies on understanding the spectral representation theorem of Section 2.4.6. The other, which is already in reach, is related to the mean square, or L^2, derivatives of random fields.

To define L^2 derivatives, choose a point $t \in \mathbb{R}^N$ and a sequence of k 'directions' $t_1', \ldots, t_k'$ in $\mathbb{R}^N$, and write these as $t' = (t_1', \ldots, t_k')$. We say that f has a *k-th order L^2 partial derivative* at t, in the direction t', if the limit

$$D_{L^2}^k f(t, t') \triangleq \lim_{h_1, \ldots, h_k \to 0} \frac{1}{\prod_{i=1}^{k} h_i} \Delta^k f(t, t', h) \qquad (2.4.6)$$

exists in mean square, where $h = (h_1, \ldots, h_k)$. Here $\Delta^k f(t, t', h)$ is the symmetrized difference

$$\Delta^k f(t, t', h) = \sum_{s \in \{0,1\}^k} (-1)^{k - \sum_{i=1}^{k} s_i} f\left(t + \sum_{i=1}^{k} s_i h_i t_i'\right)$$

and the limit in (2.4.6) is interpreted sequentially, i.e. first send h_1 to 0, then h_2, etc. A simple sufficient condition for L^2 partial differentiability of order k in all directions and throughout a region $T \subset \mathbb{R}^N$ is that

$$(2.4.7)$$

$$\lim_{h_1, \ldots, h_k, \widehat{h}_i, \ldots, \widehat{h}_k \to 0} \frac{1}{\prod_{i=1}^{k} h_i \widehat{h}_i} \mathbb{E}\left\{\Delta^k f(t, t', h) \, \Delta^k f\left(s, s', \widehat{h}\right)\right\}$$

exists[21] for all $s, t \in T$, all directions s', t', and all $h = (h_1, \ldots, h_k)$, $\widehat{h} = (\widehat{h}_i, \ldots, \widehat{h}_k)$, where the limits are again to be interpreted sequentially. Note that if f is Gaussian then so are its L^2 derivatives, when they exist.

---

[21] This is an immediate consequence of the fact that a sequence X_n of random variables converges in L^2 if, and only if, $\mathbb{E}\{X_n X_m\}$ converges to a constant as $n, m \to \infty$.

By choosing $t' = (e_{i_1}, \ldots, e_{i_k})$, where e_i is the vector with i-th element 1 and all others zero, we can talk of the mean square partial derivatives

$$\frac{\partial^k}{\partial t_{i_1} \ldots \partial t_{i_k}} f(t) \;\triangleq\; D_{L^2}^k f\left(t, (e_{i_1}, \ldots, e_{i_k})\right)$$

of f of various orders.

It is then straightforward (Exercise 2.8.8) to see that the the covariance function of such partial derivatives must be given by

$$\mathbb{E}\left\{ \frac{\partial^k f(s)}{\partial s_{i_1} \partial s_{i_1} \ldots \partial s_{i_k}} \; \frac{\partial^k f(t)}{\partial t_{i_1} \partial t_{i_1} \ldots \partial t_{i_k}} \right\} = \frac{\partial^{2k} C(s,t)}{\partial s_{i_1} \partial t_{i_1} \ldots \partial s_{i_k} \partial t_{i_k}}. \quad (2.4.8)$$

The corresponding variances have a nice interpretation in terms of spectral moments when f is stationary. For example, if f has mean square partial derivatives of orders $\alpha + \beta$ and $\gamma + \delta$ for $\alpha, \beta, \gamma, \delta \in \{0, 1, 2, \ldots\}$, then (still Exercise 2.8.8)

$$\mathbb{E}\left\{ \frac{\partial^{\alpha+\beta} f(t)}{\partial^\alpha t_i \partial^\beta t_j} \; \frac{\partial^{\gamma+\delta} f(t)}{\partial^\gamma t_k \partial^\delta t_l} \right\} = (-1)^{\alpha+\beta} \left. \frac{\partial^{\alpha+\beta+\gamma+\delta}}{\partial^\alpha t_i \partial^\beta t_j \partial^\gamma t_k \partial^\delta t_l} C(t) \right|_{t=0} \quad (2.4.9)$$

$$= (-1)^{\alpha+\beta} \, i^{\alpha+\beta+\gamma+\delta} \int_{\mathbb{R}^N} \lambda_i^\alpha \lambda_j^\beta \lambda_k^\gamma \lambda_l^\delta \, \nu(d\lambda).$$

Note that although this equation seems to have some asymmetries in the powers, these disappear due to the fact that all odd ordered spectral moments, like all odd ordered derivatives of C, are identically zero.

Here are some important special cases of the above, for which we adopt the shorthand $f_j = \partial f / \partial t_j$ and $f_{ij} = \partial^2 f / \partial t_i \partial t_j$ along with a corresponding shorthand for the partial derivatives of C.

(i) f_j has covariance function $-C_{jj}$ and thus variance $\lambda_{2e_j} = -C_{jj}(0)$, where e_j, as usual, is the vector with a 1 in the j-th position and zero elsewhere.

(ii) In view of (2.4.5), and taking $\beta = \gamma = \delta = 0$, $\alpha = 1$ in (2.4.9)

$$f(t) \text{ and } f_j(t) \text{ are uncorrelated}, \quad (2.4.10)$$

for all j and all t. If f is Gaussian, this is equivalent to independence. Note that (2.4.10) does *not* imply that f and f_j are uncorrelated *as processes*. In general, for $s \neq t$, we will have that $\mathbb{E}\{f(s)f_j(t)\} = -C_j(s-t) \neq 0$.

(iii) Taking $\alpha = \gamma = \delta = 1$, $\beta = 0$ in (2.4.9) gives that

$$f_i(t) \text{ and } f_{jk}(t) \text{ are uncorrelated} \quad (2.4.11)$$

for all i, j, k and all t. Again, if f is Gaussian, this is equivalent to independence.

Under the additional condition of isotropy, with its implication of spherical symmetry for the spectral measure, the structure of the spectral moments simplifies significantly, as do the correlations between various derivatives of f. In particular, it follows immediately from (2.4.9) that

$$\mathbb{E}\{f_i(t)f_j(t)\} = -\mathbb{E}\{f(t)f_{ij}(t)\} = \lambda_2\delta_{ij} \qquad (2.4.12)$$

where δ_{ij} is the Kronecker delta and λ_2 is the *second spectral moment*

$$\lambda_2 \triangleq \int_{\mathbb{R}^N} \lambda_i^2\, \nu(d\lambda),$$

which, because of isotropy, is independent of the value of i. Consequently, if f is Gaussian, then the first order derivatives of f are independent of one another, in addition to being independent of f itself.

2.4.4 Constant Variance

It will be important for us in later chapters that some of the relationships of the previous section continue to hold under a condition much weaker than stationarity. Of particular interest is knowing when (2.4.10) holds; i.e. when $f(t)$ and $f_j(t)$ are uncorrelated.

Suppose that f has constant variance, $\sigma^2 = C(t,t)$, throughout its domain of definition, and that its L^2 first-order derivatives all exist. In this case, analagously to (2.4.9), we have that

$$\mathbb{E}\{f(t)\,f_j(t)\} = \left.\frac{\partial}{\partial t_j}C(t,s)\right|_{s=t} = \left.\frac{\partial}{\partial s_j}C(t,s)\right|_{s=t}. \qquad (2.4.13)$$

Since constant variance implies that $\partial/\partial t_j C(t,t) \equiv 0$, this and the equivalence of the above two partial derivatives implies that these must also be identically zero. Consequently, f and its first order derivatives are uncorrelated.

One can, of course, continue in this fashion. If first derivatives have constant variance, then they, in turn, will be uncorrelated with second derivatives, in the sense that f_i will be uncorrelated with f_{ij} for all i,j. It will not necessarily be true, however, that f_i and f_{jk} will be uncorrelated if $i \neq j$ and $i \neq k$.

2.4.5 White Noise and Integration

With some easy, but extremely important, consequences of the spectral distribution theorem established, we now turn to understanding the structure of stationary random fields in a little more depth. This structure relies on some rather simple stochastic calculus. However, although stochastic calculus is a

subject with an enormous and difficult literature, we shall need only a very small part of it, and even then we shall do it without detailed proofs.

The principal aim of the current section is to set up some of the basic machinery of stochastic integration. When this is done, we can turn to the spectral representation theorem, which will provide the deeper understanding of stationary that we are looking for. *En passant*, in this section we shall meet for the first time an extremely important class of random fields, known as moving averages.

We start with a collection of independent Gaussian random variables. While such a collections are easy to construct if they are finite or countable, serious technical difficulties obstruct the construction for uncountable collections. This is why white noise, which is essentially a collection of i.i.d. random variables indexed by the points $\mathbb{R}^N$, is such a delicate object. The way around this is to avoid giving the value of the noise at specific points, and to treat is as a sort of signed measure over subsets of $\mathbb{R}^N$.

More formally, suppose ν is a measure on $\mathbb{R}^N$. The classical example is Lebesgue measure, but any measure that can be written in the form

$$\nu(A) = \int_A p(x)\,dx, \tag{2.4.14}$$

will do. Here $A \subset \mathcal{B}^N$ and p is a non-negative, but not necessarily integrable function on $\mathbb{R}^N$. For obvious reasons φ is called the (Radon-Nikodym) density of ν, although, since it generally will not integrate to one (or, indeed, anything finite) it is not a probability density. If (2.4.14) holds we say that ν is continuous.

However, completely discrete ν, that put all their mass on a finite or countable set of points, will also be important to us, and we shall even meet measures that distribute their mass smoothly over lower dimensional subsets of $\mathbb{R}^N$, such as spheres.

We can now define a *Gaussian noise*[22] W based on ν, or 'Gaussian ν-noise' as a random process defined on the Borel subsets of $\mathbb{R}^N$ such that, for all $A, B \in \mathcal{B}^N$ with $\nu A()$ and $\nu(B)$ finite,

$$W(A) \sim N(0, \nu(A)). \tag{2.4.15}$$

$$A \cap B = \emptyset \Rightarrow W(A \cup B) = W(A) + W(B) \text{ a.s.} \tag{2.4.16}$$

$$A \cap B = \emptyset \Rightarrow W(A) \text{ and } W(B) \text{ are independent.} \tag{2.4.17}$$

Property (2.4.17) is described by saying that W has *independent increments*[23]. When ν is Lebesgue measure, W is called *white noise* and it is closely

---

[22] While the notation 'W' is inconsistent with our decision to use lower case Latin characters for random functions, we retain it as a tribute to Norbert Wiener, who is the mathematical father of these processes.

[23] Much of what follows can be done without the assumption of Gaussianity, in which case (2.4.15) is replaced by requiring that $\mathbb{E}\{W(A)\} = 0$ and $\mathbb{E}\{[W(A)]^2\} = \nu(A)$.

connected to Brownian motion ($N = 1$) and the Brownian sheet. See Exercise 2.8.11.

Having defined Gaussian noises, our next step is to make sense out of the integral

$$\int_{\mathbb{R}^N} \varphi(t)\, W(dt), \tag{2.4.18}$$

for *deterministic* φ with $\int \varphi^2(x)\, \nu(dx) < \infty$.

In principle, this is not hard to do. Roughly, the argument goes as follows: Start with simple functions

$$\varphi(t) = \sum_1^n a_i \mathbb{1}_{A_i}(t), \tag{2.4.19}$$

where $A_1, \ldots, A_n \subset \mathbb{R}^N$ are disjoint, and the a_i are real, and define

$$W(\varphi) \equiv \int_T \varphi(t)\, W(dt) = \sum_1^n a_i\, W(A_i). \tag{2.4.20}$$

It follows immediately from (2.4.15) and (2.4.17) that in this case $W(\varphi)$ has zero mean and variance given by $\sum a_i^2\, \nu(A_i)$. Now think of W as a mapping from simple functions to random variables, and extend it to all functions square integrable with respect to ν. The extension is standard, and we send you to either *RFG* or any standard text for details.

An important point to note, however, is that if φ and ψ are the simple functions

$$\varphi(t) = \sum_1^n a_i \mathbb{1}_{A_i}(t), \quad \psi(t) = \sum_1^n b_i \mathbb{1}_{A_i}(t),$$

then, again by (2.4.15) and (2.4.17),

$$\mathbb{E}\{W(\varphi)W(\psi)\} = \mathbb{E}\left\{ \sum_1^n a_i W(A_i) \cdot \sum_1^n b_i W(A_i) \right\} \tag{2.4.21}$$

$$= \sum_1^n a_i b_i \mathbb{E}\left\{ [W(A_i)]^2 \right\}$$

$$= \int_{\mathbb{R}^N} \varphi(t)\psi(t)\, \nu(dt),$$

a result that extends, in general, to

$$\mathbb{E}\left\{ \int_{\mathbb{R}^N} \varphi(t)\, W(dt) \int_{\mathbb{R}^N} \psi(t)\, W(dt) \right\} = \int_{\mathbb{R}^N} \varphi(t)\psi(t)\, \nu(dt). \tag{2.4.22}$$

Note also that since linear combinations and limits of Gaussian random variables remain Gaussian, it follows that $W(\varphi)$ is also Gaussian.

With our integral defined, we can now start looking at some examples of what can be done with it. One very important use is the definition of *moving average* random fields. These are now easily constructed by taking white noise with Lebesgue measure as control measure, choosing a square integrable function $\varphi : \mathbb{R}^N \to \mathbb{R}$, and defining

$$f(t) \triangleq \int_{\mathbb{R}^N} \varphi(t - s) \, W(ds). \qquad (2.4.23)$$

Since

$$\mathbb{E}\{f(t)f(s)\} = \int_{\mathbb{R}^N} \varphi(t - u)\varphi(s - u) \, du$$

$$= \int_{\mathbb{R}^N} \varphi(t - s + v)\varphi(v) \, dv$$

$$\triangleq C(t - s),$$

moving averages are clearly stationary.

Furthermore, if φ is spherically symmetric, so that it depends only on $|t|$, f is isotropic. This is easiest to see as follows: let M be a rotation matrix, then

$$C(sM, tM) = \mathbb{E}\{f(tM)f(sM)\}$$

$$= \int_{\mathbb{R}^N} \varphi(tM - u)\varphi(sM - u) \, du$$

$$= \int_{\mathbb{R}^N} \varphi(tM - uM)\varphi(sM - uM) \, du$$

$$= \int_{\mathbb{R}^N} \varphi(t - u)\varphi(s - u) \, du$$

$$= C(s, t)$$

2.4.6 Spectral Representation Theorem

Moving averages gave us examples of stationary fields that are rather easy to generate from white noise. Now, however, we want to use stochastic integrals to find a very general way of representing *all* stationary fields on $\mathbb{R}^N$, via the so-called *spectral representation*.

We require a minor extension of the integrals of the previous section, however, since we want to work in a complex valued scenario. Thus, given a measure ν on $\mathbb{R}^N$, define a *complex ν-noise* W to be a $\mathbb{C}$-valued process satisfying

$$\mathbb{E}\{W(A)\} = 0, \quad \mathbb{E}\{W(A)\overline{W(A)}\} = \nu(A). \qquad (2.4.24)$$

$$A \cap B = \emptyset \quad \Rightarrow \quad W(A \cup B) = W(A) + W(B) \text{ a.s.} \qquad (2.4.25)$$

$$A \cap B = \emptyset \quad \Rightarrow \quad \mathbb{E}\{W(A)\overline{W(B)}\} = 0. \qquad (2.4.26)$$

We can, in addition, assume that W is Gaussian, in which case you should note that (2.4.24) does not specify all its parameters[24].

It is then a straightforward exercise to extend the construction of the previous section to an L^2 stochastic integral

$$W(\varphi) = \int_{\mathbb{R}^N} \varphi(\lambda)\, W(d\lambda)$$

for $\varphi : \mathbb{R}^N \to \mathbb{C}$ with $\int_{\mathbb{R}^N} \|\varphi\|^2\, d\nu < \infty$. It is immediate from the construction that

$$\mathbb{E}\{W(\varphi)\overline{W(\psi)}\} = \int_{\mathbb{R}^N} \varphi(\lambda)\overline{\psi(\lambda)}\, \nu(d\lambda). \qquad (2.4.27)$$

This construction allows us to state the following important result.

Theorem 2.4.2 (spectral representation theorem). *Let ν be a finite measure on $\mathbb{R}^N$ and W a complex ν-noise. Then the complex valued random field*

$$f(t) = \int_{\mathbb{R}^N} e^{i\langle t,\lambda \rangle}\, W(d\lambda) \qquad (2.4.28)$$

has covariance

$$C(s,t) = \int_{\mathbb{R}^N} e^{i\langle (s-t),\lambda \rangle}\, \nu(d\lambda), \qquad (2.4.29)$$

and so is (weakly) stationary. If W is Gaussian, then so is f.

Furthermore, to every mean square continuous, centered, (Gaussian) stationary random field f on $\mathbb{R}^N$ with covariance function C and spectral measure ν there corresponds a complex (Gaussian) ν-noise W on $\mathbb{R}^N$ such that (2.4.28) holds in mean square for each $t \in \mathbb{R}^N$.

In both cases, W is called the spectral process corresponding to f.

In one direction, this theorem does need a proof. It is a consequence of the construction of the stochastic integral that the process f defined by (2.4.28) has covariance function (2.4.29). The other direction is not so easy, although not hard. You can find the details in almost any book on time series – our

---

[24] See Footnote 18 above. In general, if we split a complex ν-noise into its real and imaginary parts, W_R and W_I say, it does not follow from (2.4.26) that $A \cap B = \emptyset$ implies any of $\mathbb{E}\{W(A)W(B)\} = 0$, $\mathbb{E}\{W_R(A)W_R(B)\} = 0$, $\mathbb{E}\{W_I(A)W_I(B)\} = 0$, or $\mathbb{E}\{W_I(A)W_R(B)\} = 0$. Indeed, this is most definitely not the case for the complex W of Theorem 2.4.2 when, for example, the stationary process f there is real valued. See the discussion following (2.4.30) on 'real' spectral representations, in which the above problem is addressed by taking W to be defined on a half space rather than all of $\mathbb{R}^N$.

favourite is [23] – for processes on either $\mathbb{Z}$ or $\mathbb{R}$, and the extension to $\mathbb{R}^N$ is trivial[25].

When the basic field f is real, it is natural to expect a 'real' spectral representation, and this is in fact the case, although notationally, *and computationally*, it is still generally more convenient to use the complex formulation. Nevertheless, the real representation is also useful. To describe it, recall firstly that if f is real, the symmetry of the spectral measure ν allows us to introduce three[26] new measures, on $\mathbb{R}_+ \times \mathbb{R}^{N-1}$, by setting

$$\nu_1(A) = \nu(A \cap \{\lambda \in \mathbb{R}^N : \lambda_1 > 0\}),$$
$$\nu_2(A) = \nu(A \cap \{\lambda \in \mathbb{R}^N : \lambda_1 = 0\}),$$
$$\mu(A) = 2\nu_1(A) + \nu_2(A).$$

We can now rewrite[27] (2.4.2) in real form, as

$$C(t) = \int_{\mathbb{R}_+ \times \mathbb{R}^{N-1}} \cos(\langle \lambda, t \rangle) \, \mu(d\lambda). \qquad (2.4.30)$$

There is also a corresponding real form of the spectral representation (2.4.28). The fact that the spectral representation yields a real valued process also implies certain symmetries on the spectral process W. In particular, it turns out that there are two independent *real* valued μ-noises, W_1 and W_2, such that[28]

$$f_t = \int_{\mathbb{R}_+ \times \mathbb{R}^{N-1}} \cos(\langle \lambda, t \rangle) \, W_1(d\lambda) + \int_{\mathbb{R}_+ \times \mathbb{R}^{N-1}} \sin(\langle \lambda, t \rangle) \, W_2(d\lambda). \quad (2.4.31)$$

It is easy to check that f so defined has the right covariance function.

The real representation goes a long way to helping one develop a good understanding of what the spectral representation theorem says, and so we devote a few paragraphs to this. While it is not necesary for the rest of the theory, it does help develop intuition and very much comes into its own when we turn to issues of simulation in Chapter 7.

One way to think of the integral in (2.4.31) is via the approximating sum

$$\sum_i \left\{ \cos(\langle \lambda_i t \rangle) W_1(\Lambda_i) + \sin(\langle \lambda_i, t \rangle) W_2(\Lambda_i) \right\} \qquad (2.4.32)$$

---

[25] There is also an inverse to (2.4.28), expressing W as an integral involving f, but we shall have no need of it.

[26] Note that if ν is continuous (so that (2.4.14) holds) the second of these measures, ν_2, will be identically zero.

[27] There is nothing special about the half-space $\lambda_1 \geq 0$ taken in this representation. Any half space will do.

[28] In one dimension, it is customary to take W_1 as a μ-noise and W_2 as a $(2\nu_1)$-noise, which at first glance is different to what we have. However, noting that, when $N = 1$, $\sin(\lambda t) W_2(d\lambda) = 0$ when $\lambda = 0$, it is clear that the two definitions in fact coincide in this case.

where the $\{\Lambda_i\}$ give a partition of $\mathbb{R}_+ \times \mathbb{R}^{N-1}$ and $\lambda_i \in \Lambda_i$. Indeed, this sum will be exact if the spectral measure is discrete with atoms λ_i. In either case, what (2.4.32) does is to express the random field as the sum of a large number of sinusoidal components.

In the one-dimensional situation the basic components in (2.4.32) are simple sine and cosine waves of (random) amplitudes $|W_2(\Lambda_i)|$ and $|W_1(\Lambda_i)|$, respectively, and wavelengths equal to $2\pi/\lambda_i$. In two dimensions let λ be a generic λ_i in (2.4.32) (i.e. drop the subscript on λ_i) with components λ_1 and λ_2, we have that an elementary cosine wave is of the form $\cos(\lambda_1 t_1 + \lambda_2 t_2)$. The points λ_1 and λ_2 are fixed and the point (t_1, t_2) ranges over $\mathbb{R}^2$. This gives a sequence of waves travelling in a direction which makes an angle $\arctan(\lambda_2/\lambda_1)$ with the t_1 axis and having wavelength $2\pi/\sqrt{\lambda_1^2 + \lambda_2^2}$, the distance between troughs or crests, as measured along the line perpendicular to the crests. An example is given in Figure 2.4.1.
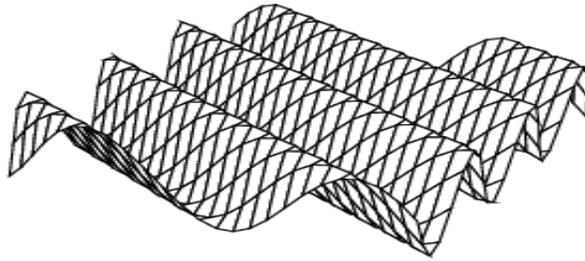
Fig. 2.4.1. The elementary wave form $\cos(\lambda_1 t_1 + \lambda_2 t_2)$ in $\mathbb{R}^2$.

The corresponding sine function is exactly the same, except for the obvious a phase shift of half a wavelength. As in the one-dimensional case, the amplitudes of the components $\cos(\langle \lambda_i, t \rangle)$ and $\sin(\langle \lambda_i, t \rangle)$ are given by the random variables $|W_1(\Lambda_i)|$ and $|W_2(\Lambda_i)|$. Figure 2.4.2 shows what a sum of 10 such components looks like, when the λ_i are chosen randomly in $(-\pi, \pi]^2$ and the $W_j(\lambda_i)$ are independent $N(0,1)$. Higher dimensional situations behave analogously.

Another way to think of the spectral representation theorem is as a particular (rather non-rigorous) application of the Karhunen-Loève expansion of Section 2.3.5. Let's start by assuming that all that we wrote back there also works for complex valued processes and on the definitely unbounded region $\mathbb{R}^N$. Since we are in the stationary scenario, it is actually quite easy to find eigenfunctions for the integral equation (2.3.20), via complex exponentials. In fact, for any $\lambda \in \mathbb{R}^N$, the function $e^{i\langle t, \lambda \rangle}$ (as a function of $t \in \mathbb{R}^N$) satisfies
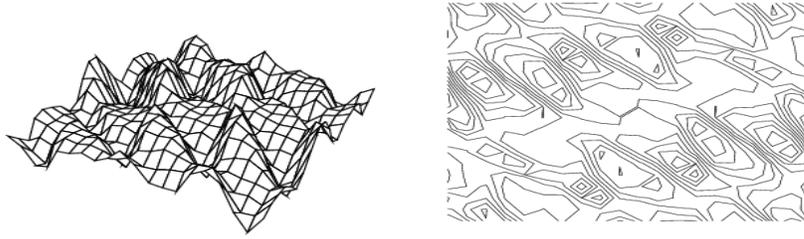
Fig. 2.4.2. A more realistic surface based on (2.4.32), along with contour lines at the zero level.

$$\int_{\mathbb{R}^N} C(s,t)\, e^{i\langle s,\lambda\rangle}\, ds = \int_{\mathbb{R}^N} C(t-s)\, e^{i\langle s,\lambda\rangle}\, ds$$

$$= e^{i\langle t,\lambda\rangle} \int_{\mathbb{R}^N} C(u)\, e^{-i\langle u,\lambda\rangle}\, du$$

$$= K_\lambda\, e^{i\langle t,\lambda\rangle},$$

for some, possibly zero, K_λ.

Suppose that $K_\lambda \neq 0$ for only a countable number of $\lambda \in \mathbb{R}^N$. Then the stationary, complex version of the Mercer expansion (2.3.21) can be written as

$$C(t) = \sum_\lambda K_\lambda\, e^{i\langle t,\lambda\rangle}, \tag{2.4.33}$$

while the Karhunen-Loève expansion becomes

$$f(t) = \sum_\lambda K_\lambda^{1/2} \xi_\lambda\, e^{i\langle t,\lambda\rangle}. \tag{2.4.34}$$

These are, respectively, special cases of the spectral distribution theorem (cf. (2.4.2)) and the spectral representation theorem (cf. (2.4.28)) when the spectral measure is discrete.

Despite the minor irregularity of assuming that f is complex valued, the above argument is completely rigorous. On the other hand, what follows for the case in which $K_\lambda \neq 0$ on an uncountable set, is not. Nevertheless, it is still somewhat enlightening to look at. In this case, one could imagine replacing the summations in (2.4.33) and (2.4.34) by integrals, to obtain

$$C(t) = \int_{\mathbb{R}^N} K_\lambda\, e^{i\langle t,\lambda\rangle}\, d\lambda$$

and

$$f(t) = \int_{\mathbb{R}^N} K_\lambda^{1/2} \xi_\lambda\, e^{i\langle t,\lambda\rangle}\, d\lambda. \tag{2.4.35}$$

Everything is well defined in the first of these integrals, but in the second we have the problem that the ξ_λ should be independent for each λ. The way to

make mathematics of this is, of course, via stochastic integration, which is precisely what we did above.

Nevertheless, despite the lack of rigor of the above ideas, they do give a useful way of thinking about the spectral representation theorem.

2.4.7 Isotropy

We now turn to the special case of stationary, isotropic random fields on $\mathbb{R}^N$, for which the covariance function $C(t)$ is a function only of $|t|$. Not surprisingly, isotropy implies significant simplifying consequences for the spectral distribution and respresentation theorems.

As we already noted in Section 2.4.2, spectral measures of isotropic random fields are sperically symmetric. Consequently, they cannot have all their mass concentrated in one small region in $\mathbb{R}^N$ away from the origin. In particular, it is not possible to have a spectral measure degenerate at one point, unless that point is the origin. The closest the spectral measure of an isotropic field can come to this sort of behavior is to have all its probability concentrated in an annulus of the form

$$\{\lambda \in \mathbb{R}^N : a \leq |\lambda| \leq b\}, \qquad a \leq b.$$

In such a case it is not hard to see that that the field itself is then composed of a 'sum' of waves travelling in all directions but with wavelengths between $2\pi/b$ and $2\pi/a$ only.

Other, often unexpected, restrictions on covariance functions also arise from isotropy. For example, isotropic covariance functions cannot be very negative, a result due originally to the Matérn [62], who, in the 1950's, was one of the first researchers to actually employ random fields as a statistical modelling tool. In particular, he showed that the covariance functions of centered, isotropic random fields must satisfy (Exercise 2.8.13)

$$C(t) \geq -C(0)/N, \qquad \text{for all } t \in \mathbb{R}^N. \tag{2.4.36}$$

More important than the above, however, are the consequences of isotropy for the spectral distribution and representation theorems. We shall state these formally, and even show how they are derived, since the first response of a modern mathematician on seeing Bessel functions is to imagine that they arise from something quite mysterious. In fact, here quite the opposite is the case, and although the original statement of the spectral distribution theorem, with its complex exponentials, may look simpler than the version that follows, there are other consequences, at the level of the spectral representation theorem for isotropic processes (Theorem 2.4.4) that have quite significant practical importance. In particular, this will help set the scene for what happens when we look at random fields on spheres.

The following result, due originally to Schoenberg [82] (in a somewhat different setting) and Yaglom [101] describes what happens.

Theorem 2.4.3. *For C to be the covariance function of a mean square continuous, isotropic, random field on $\mathbb{R}^N$ it is necessary and sufficient that*

$$C(t) \; = \; \int_0^\infty \frac{J_{(N-2)/2}(\lambda|t|)}{(\lambda|t|)^{(N-2)/2}}\, \mu(d\lambda), \qquad (2.4.37)$$

where μ is a finite measure on $\mathbb{R}_+$ and J_m is the Bessel function of the first kind of order m; viz.

$$J_m(x) \; = \; \sum_{k=0}^\infty (-1)^k \frac{(x/2)^{2k+m}}{k!\,\Gamma(k+m+1)}.$$

Proof. The proof consists in simplifying the basic spectral representation by using the symmetry properties of ν.

We commence by converting to polar coordinates, $(\lambda, \theta_1, \ldots, \theta_{N-1})$, $\lambda \geq 0$, $(\theta_1, \ldots, \theta_{N-1}) \in S^{N-1}$, where

$$S^{N-1} = \left\{ t \in \mathbb{R}^N : |t| = 1 \right\}$$

is the unit sphere in $\mathbb{R}^N$. Define a measure μ on $\mathbb{R}_+$ by setting $\mu([0, \lambda]) = \nu(B_\lambda^N)$, and extending as usual, where

$$B_\lambda^N = \left\{ t \in \mathbb{R}^N : |t| \leq \lambda \right\}$$

is the N-ball of radius λ and ν is the spectral measure of (2.4.2).

Then, on substituting into (2.4.2) with $t = (|t|, 0, \ldots, 0)$ and performing the coordinate transformation, we obtain

$$C(|t|) = \int_0^\infty \int_{S^{N-1}} \exp(i|t|\lambda \cos\theta_{N-1})\, \sigma(d\theta)\mu(d\lambda)$$

where σ is surface area measure on S^{N-1}. Integrating out $\theta_1, \ldots, \theta_{N-2}$ it follows that

$$C(|t|) = s_{N-2} \int_0^\infty \int_0^\pi e^{i\lambda|t|\cos\theta_{N-1}} (\sin\theta_{N-1})^{N-2}\, d\theta_{N-1}\mu(d\lambda)$$

where

$$s_N \triangleq \frac{2\pi^{N/2}}{\Gamma(N/2)}, \qquad N \geq 0, \qquad (2.4.38)$$

is the surface area[29] of S^{N-1}.

The inside integral can be evaluated in terms of Bessel functions to yield

$$\int_0^\pi e^{i\lambda|t|\cos\theta} \sin^{N-2}\theta\, d\theta \; = \; \frac{J_{(N-2)/2}(\lambda|t|)}{(\lambda|t|)^{(N-2)/2}}$$

which, on absorbing s_{N-2} into μ, completes the proof.   □

---

[29] Including the case $N = 0$ here is not a mistake. We shall need it later. In any case, it makes perfect sense if we think of $\mathbb{R}^0$ as the integers with counting measure.

For small values of the dimension N, (2.4.37) can be simplified even further. For example, substituting $N = 2$ into (2.4.37) yields that in this case

$$C(t) = \int_0^\infty J_0(\lambda|t|)\,\mu(d\lambda),$$

while substituting $N = 3$ and evaluating the inner integral easily yields that in this case

$$C(t) = \int_0^\infty \frac{\sin(\lambda|t|)}{\lambda|t|}\,\mu(d\lambda).$$

Given the fact that the covariance function of an isotropic field takes such a special form, it is natural to seek a corresponding form for the spectral representation of the field itself. Such a representation does in fact exist and we shall now describe it, albeit without giving any proofs. These can be found, for example, in the book by Wong [97], or as special cases in the review by Yaglom [102]. Another way to verify it would be to check that the representation given in Theorem 2.4.4 below yields the covariance structure of (2.4.37). This is essentially an exercise in the manipulation of special functions.

The spectral representation of isotropic fields on $\mathbb{R}^N$ is based on the so-called *spherical harmonics* on the $(N-1)$-sphere, which form an orthonormal basis for the space of complex, square integrable functions on S^{N-1} equipped with the usual surface measure. We shall denote them by $\{h_{ml}^{(N-1)},\ l = 1, \ldots, d_m,\ m = 0, 1, \ldots\}$ where $d_m = \binom{N+m-2}{N-2}(N+2m-2)/(N+m-2)$. Figure 2.4.3 is a representation of the first few spherical harmonics, with $N = 2$ and $0 \le m \le 3$ in descending order in m. In each representation the distance from the underlying sphere indicates the size of $|h_{ml}|$, while the color/shading represents the ray in the complex plane on which h_{ml} sits. The color/shading key is shown in the disc at top right. Thus h_{01} takes a constant, positive, real value (1), while h_{11} takes positive real values in the upper hemisphere and negative real values in the lower hemisphere, etc.

Given the spherical harmonics, we can now use the spectral decomposition

$$f(t) = \int_{\mathbb{R}^N} e^{i\langle t, \lambda \rangle}\,W(d\lambda)$$

to define a family of noises on $\mathbb{R}_+$ by setting

$$W_{ml}(A) = \int_A \int_{S^{N-1}} h_{ml}^{(N-1)}(\theta)\,W(d\lambda, d\theta)$$

where, once again, we work in polar coordinates. Note that since W is a ν-noise, where ν is the spectral measure, information about the covariance of f has been coded into the W_{ml}. From this family, define a family of mutually uncorrelated, stationary, one-dimensional processes $\{f_{ml}\}$ by

$$f_{ml}(r) = \int_0^\infty \frac{J_{m+(N-2)/2}(\lambda r)}{(\lambda r)^{(N-2)/2}}\,W_{ml}(d\lambda),$$
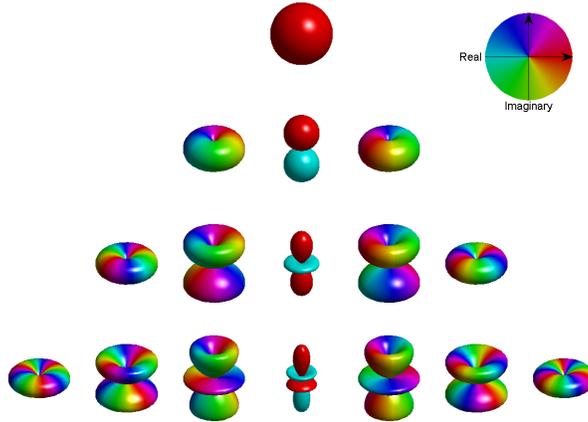
Fig. 2.4.3. Spherical harmonics, $N = 2$, $0 \le m \le 3$.

where, as in the spectral representation (2.4.28), one has to justify the existence of this L^2 stochastic integral. These are all the components we need in order to state the following.

Theorem 2.4.4. *A centered, mean square continuous, isotropic random field on $\mathbb{R}^N$ can be represented by*

$$f(t) \; = \; f(r, \theta) \; = \; \sum_{m=0}^{\infty} \sum_{l=1}^{d_m} f_{ml}(r) h_{ml}^{(N-1)}(\theta). \tag{2.4.39}$$

In other words, isotropic random fields can be decomposed into a countable number of mutually uncorrelated stationary processes with a one-dimensional parameter, a result which one would not intuitively expect. As noted above, there is still a hidden spectral process in (2.4.39), entering via the W_{ml} and f_{ml}. This makes for an important difference between (2.4.39) and the similar looking Karhunen-Loève expansion we met in Section 2.3.5. Another difference lies in the fact that while it is possible to truncate the expansion (2.4.39) to a finite number of terms and retain isotropy, this is not true of the standard Karhunen-Loève expansion. In particular, isotropic fields can never have finite Karhunen-Loève expansions. For a heuristic argument as to why this is the case, recall from that under isotropy the spectral measure must be invariant under rotations, and so cannot be supported on a finite, or even countable, number of points. Consequently, one also needs an uncountable number of independent variables in the spectral noise process to generate the process via (2.4.28). However a process with a finite Karhunen-Loève expansion provides only a finite number of such variables, which can never be enough.

We close this section on isotropy with a brief discussion of fields in 'space-time'. Taking the lead from the moving ocean waves with which we motivated this chapter, it is not at all uncommon in applications to find random fields that are functions of 'space' x and 'time' t, so that the parameter set is most conveniently written as $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^N$. Such processes are often stationary in t and isotropic in x, in the sense that

$$\mathbb{E}\{f(s, u)\overline{f(s + t, u + x)}\} \;=\; C(t, |x|),$$

where C is now a function from $\mathbb{R}^2$ to $\mathbb{C}$. In such a situation the methods that we used to study the purely isotropic case also suffice to show that C can be written in the form

$$C(t, x) \;=\; \int_{-\infty}^{\infty} \int_0^{\infty} e^{it\nu} G_N(\lambda x) \, \mu(d\nu, d\lambda),$$

where

$$G_N(x) = \left(\frac{2}{x}\right)^{(N-2)/2} \Gamma\left(\frac{N}{2}\right) J_{(N-2)/2}(x)$$

and μ is a measure on the half-plane $\mathbb{R}_+ \times \mathbb{R}^N$.

It is not hard to also develop a corresponding representation for the field itself.

2.4.8 Isotropic Fields on the Sphere and Other Groups

While the stationary and isotropic random fields that we have studied so far in this section have a very nice theory, they all had the common property that were defined over the Euclidean spaces $\mathbb{R}^N$. However, when we turn to applications we shall see that there are many random fields defined on more delicate parameter spaces, such as surfaces.

One classic example is electroencephalogram (EEG) data, neurophysiologic measurements of the electrical activity of the brain taking by recording from electrodes placed on the scalp. Unlike the fMRI data of Chapter 1, which is data taken over a solid domain in $\mathbb{R}^3$, EEG data is taken over a surface which, for the moment, we can think of as a sphere. Another example is provided by the Nobel Prize winning COBE (COsmic microwave Background Explorer) astrophysical data. This data is on the sphere of all directions away from the Earth, and measures an integral of microwave radiation throughout the universe, in each direction. We shall look at these two examples in some detail in Chapters 9 and 10.

However, since at the moment we are still doing theory rather than application, there is no reason to stay with the two dimensional sphere, and we introduce

$$S_\lambda^{N-1} = \{t \in \mathbb{R}^N : |t| = \lambda\}, \tag{2.4.40}$$

the sphere in $\mathbb{R}^N$ of radius λ. We shall write S^{N-1} for the unit sphere S_1^{N-1}.

To define stationarity and isotropy for random fields defined on S_λ^{N-1} we need a notion of translation, which is usually done by identifying each point $t \in S_\lambda^{N-1}$ with a rotation, and thinking of $t_1 + t_2$ as a composition of two rotations.

Then an interesting corollary of Theorem 2.4.4 is obtained by fixing r in (2.4.39). We then have, for a homogeneous and isotropic random field on S_λ^{N-1}, a simple representation in terms of uncorrelated random coefficients $f_{ml}(r)$ and spherical harmonics. If the random field is Gaussian, then the coefficients are actually independent, and we will, essentially, have generated a Karhunen-Loève expansion.

Furthermore, the covariance function can be written as

$$C(t_1, t_2) = C(\alpha_{12}) \;=\; \sum_{m=0}^{\infty} \sigma_m^2 C_m^{(N-1)/2}(\cos \alpha_{12}), \qquad (2.4.41)$$

where α_{12} is the angular distance between t_1 and t_2, the C_m^N are Gegenbauer polynomials, and σ_m^2 is the common variance of the $f_{ml}^2(r)$ in (2.4.39)[30].

By now it should be starting to become evident that all of these representations must be special cases of some general theory, that might also be able to cover quite non-Euclidean parameter spaces, as long as they have some group structure that mimics the ideas of translation and perhaps rotation. This is indeed the case, and what happens, in essence, is that the complex exponentials that have been at the core of all the representations of this section are replaced by the so-called 'characters' of the group.

You can find many more results of this kind, with much more detail, in the classic paper of Yaglom [102] or the book by Leonenko [54]. Other useful treatments, with various emphases, can be found in [14, 40, 55] and [103].

2.4.9 Transforming to stationarity and isotropy

The notions of stationarity and isotropy are clearly important, and particularly pleasant from a mathematical viewpoint, in that, as we have just seen, they allow for quite simple representations of random fields.

Nevertheless, Nature is not always as kind as to provide stationarity or isotropy, and it is natural to ask what can be done in that case. In particular, is would be natural to ask if there was a simple way to transform general random fields to stationarity and/or isotropy.

---

[30] An important special case of (2.4.41) occurs when $\sigma_m^2 = 0$ for all m other than $m = 1$, and $\sigma_1^2 = 1/(N-1)$. In this case the covariance function is given by $C(t_1, t_2) = \langle t_1, t_2 \rangle$ and the resulting random field is known as the canonical (isotropic) Gaussian field on the sphere. This canonical field plays a central rôle in the theory of smooth Gaussian fields, and is the point at which the Gaussian kinematic formulae of Chapter 4 meet the kinematic fundamental formula of Chapter 3. The details of the mathematics can be found in Chapter 15 of *RFG*.

Despite the fact that the applied literature abounds with papers having titles which seem to indicate that this is indeed possible, the fact, sadly, is that in general it is not. We shall nevertheless take a moment to indicate what *is* known.

In a literature that seems to have begun with a statistical paper [79] by Sampson and Guttorp in 1992, there has been considerable interest in models of the form

$$f(t) = g(\varphi(t)),$$

where g is a stationary, isotropic random field and φ a bijective transformation of some kind. The now generally non-stationary and non-isotropic f can therefore be transformed to stationarity and isotropy by studying $f(\varphi^{-1}(t))$.

In [79] a statistical procedure was developed, based on the so-called *variogram*[31]

$$V(s,t) \stackrel{\Delta}{=} \mathbb{E}\left\{ [f(t) - f(s)]^2 \right\},$$

for estimating the transformation φ in a non-parametric fashion. The procedure, while obviously useful, applies only for two-dimensional, real valued random fields, and does not seem to have an easy extension to higher dimensions.

There are some more recent papers [7, 8] which approach this problem a different way, by assuming the transformation φ to be quasi-conformal. Once again, the approach is limited to two dimensions, and, in this case, the extension to higher dimensions seems completely blocked.

However, neither these procedures nor related ones (see [7, 8] for a recent bibliography) answer the general question we originally asked: Can general non-stationary or non-isotropic fields be easily treated within the stationary/isotropic framework; i.e. fields which we do not know a priori to be simple transforms?

Unfortunately, the answer to this question is negative. Nevertheless, there are some techniques that can help. For example, there is a notion of *local isotropy* that is often useful. To define this, recall from Section 2.4.3 that if a random field is stationary and isotropic, then, among many other properties, its first order partial derivatives are uncorrelated and have common variance. It turns out that many of the properties of Gaussian and related random fields rely only on these consequences, and so we shall define a *locally isotropic random field* to be one with constant variance and

$$\mathbb{E}\left\{ f_i(t) f_j(t) \right\} = \lambda_2 \delta_{ij}, \tag{2.4.42}$$

---

[31] While the name is different, the variogram is no more than the square of the canonical metric that we shall meet and heavily use in the following section. While there, as the canonical metric, it will play an important rôle for developing theory, it plays no less an important rôle in the applied literature as the variogram.

for some common second spectral moment λ_2, where $f_i(t) = \partial f(t)/\partial t_i$.

A natural question to now ask is how easy is it to transform general random fields to locally isotropic ones. For stationary random fields f this is straightforward. If Λ is the $N \times N$ matrix of second spectral moments $\lambda_{ij} = \mathbb{E}\{f_i(t)f_j(t)\}$, then it is trivial to check that the field defined by $f(\Lambda^{-1/2}t)$ is locally isotropic.

In general, however, there is no such simple transformation available. However, there is a trick, based on Riemannian geometry, that allows one to compute many things about constant variance Gaussian random fields as if they were, in fact, locally isotropic. It was this trick that, in many ways, was one of the most important themes of *RFG*. It will be explained in Section 4.3 and heavily used thereafter. However, we shall have to develop a few more concepts and notation before we can introduce it, so we leave it for the moment.

Finally, we note that there is a technique [34, 68, 69, 70] for making stationary fields from non-stationary ones, at the price of enlarging the parameter space. Suppose f is a non-stationary random field f on $T \subset \mathbb{R}^N$, and $\varphi : \mathbb{R}^N \to \mathbb{R}^{2N}$ a bijective mapping. Consider the parameter space

$$\widehat{T} \triangleq \{t, \varphi(t)) : t \in T\} \subset \mathbb{R}^{2N},$$

and the random field $\widehat{f}$ on the manifold $\widehat{T}$ defined by $\widehat{f}(x, \varphi(x)) = f(\varphi^{-1}(x))$. Then it is easy to check that $\widehat{f}$ is 'stationary' on $\widehat{T}$ in the sense

$$\mathbb{E}\left\{\widehat{f}(x, \varphi(x))\widehat{f}(y, \varphi(y))\right\}$$

is a function of $(x, \varphi(x)) - (y, \varphi(y))$ only[32]. However, since it is an open (and hard) question as to whether $\widehat{f}$ can be extended to a stationary process on all of $\mathbb{R}^{2N}$, it is hard to exploit this fact in any practical fashion.

2.5 Smoothness of Random Fields

Now that we have discussed the basic construction and structure of random fields, we turn to discussing basic sample path properties. For example, we would like to know when a random field f is continuous, or continuously differentiable. These two issues are precisely what we shall look at in this section, mainly for Gaussian fields.

It is actually quite easy to describe what kinds of conditions are needed to ensure that these smoothness properties hold, at least in the Gaussian case. For simplicity, suppose that f is Gaussian, centered, and stationary, with covariance function C. Since C determines the finite dimensional distributions of f, and these determine its properties, ultimately we are looking for conditions on C.

---

[32] In essence, this result is a consequence of the famous Nash embedding theorem.

Now, if f is to be smooth, it must be true that for s and t close, the difference $f(t) - f(s)$ must be small. However, these differences have a known distribution, viz.

$$f(t) - f(s) \sim N\left(0, \mathbb{E}\left\{\left|f(s) - f(t)|^2\right|\right\}\right)$$
$$= N\left(0, 2[C(0) - C(t - s)]\right),$$

the second line a consequence of stationarity. Thus, what we require is that C itself is smooth in the neighborhood of the origin[33]. The only question is "exactly how smooth does it need to be?", and this is what we plan to answer.

Still in the stationary Gaussian scenario, one can also consider smoothness properties from the point of view of the spectral representation theorem. As we described in Section 2.4.6, one can think of stationary processes as a sum of infinitely many sinusoids. In particular, think of the (non-rigorous) Karhunen-Loève type representation (2.4.34), in terms of complex exponentials, in which we wrote

$$f(t) = \sum_{\lambda} K_{\lambda}^{1/2} \xi_{\lambda}\, e^{i\langle t, \lambda \rangle}.$$

Obviously, since each function $e^{i\langle t, \lambda \rangle}$ is infinitely differentiable, the first few terms in the sum have no affect on the level of smoothness of f. In fact, no finite number of terms can affect it. Thus, the issue of smoothness is determined by the behavior of the tail of the sum, or, equivalently, by the behavior of $\mathbb{E}\{|K_{\lambda}|\}$ as $|\lambda| \to \infty$. Equivalently, it is determined by the variance of the high frequency components, or, in terms of the spectrum, by its decay rate at infinity.

These two considerations – the behavior of the covariance function at the origin and the behavior of the spectral measure at infinity – are actually the same mathematically, since they are linked by the Tauberian theorems of Fourier analysis. However, from a stochastic point of view, they give us two seemingly quite different ways to think about the smoothness of random fields.

To turn these ideas into mathematics is not trivial, and we shall decribe how to do it in the remainder of this section. The details (but not the results) can, however, be skipped by the reader who cares only about the 'what' rather than the 'why' of smoothness, or by the reader who feels that rigor leads only to *rigor mortis.* However, the reader who is prepared to accept that all his random fields satisfy regularity requirements, without checking, will do well to remember the principle of *caveat emptor.*

We start with the issue of continuity. Rather interestingly, and importantly, this turns out to be a problem which can be studied, at least for Gaussian processes, just as easily over very general parameter spaces as over $\mathbb{R}^N$, as long

---

[33] As a consequence, it will be smooth everywhere. See Exercise 2.8.15.

as they have metrics defined on them[34]. First of all, however, it is important to recall that there are several notions of continuity for random processes over a parameter space T, among them

Continuity in probability:

$$\lim_{s \to t} \mathbb{P}\left\{|f(t) - f(s)| \geq \varepsilon\right\} = 0, \quad \text{for each } t \in T \text{ and each } \varepsilon > 0.$$

Continuity in mean square, or L^2 continuity:

$$\lim_{s \to t} \mathbb{E}\left\{|f(t) - f(s)|^2\right\} = 0, \quad \text{for each } t \in T.$$

Continuity with probability one, sample path, or almost sure (a.s.), continuity:

$$\mathbb{P}\{\lim_{s \to t} |f(t) - f(s)| = 0, \text{ for all } t \in T\} = 1.$$

That these can be quite different is shown in Exercise 2.8.14. We are interested in the strongest of these, continuity with probability one.

2.5.1 The General Gaussian Theory

The aim of this section is to describe useful *sufficient* conditions for a centered Gaussian random field on a parameter space T to be almost surely bounded and/or continuous; i.e. to determine conditions for which

$$\mathbb{P}\{\sup_{t \in T} |f(t)| < \infty\} = 1 \quad \text{or} \quad \mathbb{P}\{\lim_{s \to t} |f(t) - f(s)| = 0, \text{ for all } t \in T\} = 1.$$

To start, define a metric d on T, known as the *canonical metric* for T and/or f, by setting

$$d(s,t) \stackrel{\Delta}{=} \left\{ \mathbb{E}\left[(f(s) - f(t))^2 \right] \right\}^{\frac{1}{2}}, \tag{2.5.1}$$

in a notation that will henceforth remain fixed[35].

The next thing we need is notation for the ball, of radius ε, in the canonical metric, centered at a point $t \in T$, which we denote by

$$B_d(t, \varepsilon) \stackrel{\Delta}{=} \left\{ s \in T : d(s,t) \leq \varepsilon \right\}. \tag{2.5.2}$$

---

[34] In fact, not even a metric is required. It suffices that T is metrizable, and one can then study everything in terms of the canonical metric (2.5.1). That the main questions of continuity are actually independent of the choice of metric is proven, for example, in Section 1.3 of *RFG*.

[35] Actually, d is only a pseudo-metric, since although it satisfies all the other demands of a metric, $d(s,t) = 0$ does not necessarily $s = t$. To see why, think of a periodic process on $\mathbb{R}$, with period p. Then $d(s,t) = 0$ implies no more than $s - t = kp$ for some integer k.

It is important to realise that even if the underlying parameter space is as simple as $\mathbb{R}^2$ these 'balls' need not be 'round'. For example, consider the four balls in Figure 2.5.1, which are all assumed to have the same radius in the canonical metric. Balls a and b are physically (in a Euclidean sense) round, which indicates that in the regions in which they are placed the canonical distance between two points is proportional to the Euclidean distance. However, the constant of proportionality is different in the two regions. In both cases it makes sense to talk about some sort of local isotropy, at least as seen through the canonical metric. Of course, since this metric only takes into account second moments, such local isotropy is in the limited sense of Section 2.4.9.
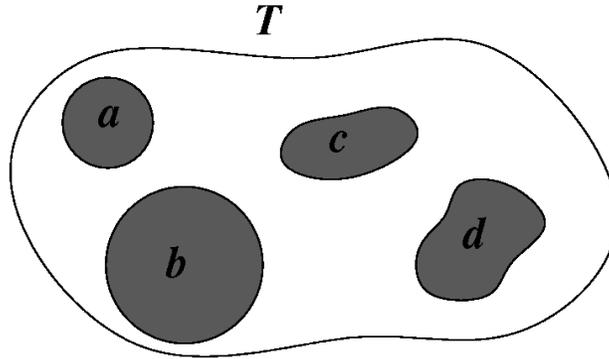

Fig. 2.5.1. Balls in the canonical metric

To understand what this means about the random process, note from (2.5.1) that for, $s, t \in T$, the d-distance between any two points is the standard deviation of the difference $f(t) - f(s)$. Thus, in regions where this standard deviation is large, we expect the process to move around comparatively rapidly, and, as a consequence, the 'physical' balls of fixed d-radius will be smaller. Thus while both a and b are regions of local isotropy, we would expect f to behave far more erratically in the former.

In regions c and d we have also lost local isotropy, and so the balls, albeit of the same d-radius, look quite different.

Once this is understood, it is not a big jump to realise that the number of balls needed to cover T must, in some way, measure the behavior of f over T, and this brings us to an important definition.

Definition 2.5.1. *Let f be a centered Gaussian field on T, and d the canonical metric (2.5.1). Assume that T is d-compact, in that*

$$\operatorname{diam}(T) \overset{\Delta}{=} \sup_{s,t \in T} d(s,t) < \infty. \tag{2.5.3}$$

Fix $\varepsilon > 0$ and let $N(T, d, \varepsilon) \equiv N(\varepsilon)$ denote the smallest number of d-balls of radius ε whose union covers T. Set

$$H(T, d, \varepsilon) \equiv H(\varepsilon) = \ln\left(N(\varepsilon)\right). \qquad (2.5.4)$$

Then N and H are called the (metric) entropy and log-entropy functions for T (or f). We shall refer to any condition or result based on N or H as an entropy condition/result.

Note that since we are assuming that $\mathrm{diam}(T)$ is finite, it follows that $H(\varepsilon) < \infty$ for all $\varepsilon > 0$. On the other hand, the same need not be (nor generally is) true for $\lim_{\varepsilon \to 0} H(\varepsilon)$, since as $\varepsilon \to 0$ the d-balls get smaller and smaller, and we need more and more of them to cover T. It is precisely the growth rate of H at zero that captures the 'smoothness of the covariance at the origin' that we described in the setting of stationary fields at the beginning of this section.

Here then is the main result about Gaussian continuity and boundedness, due originally, more or less in the form given below, to Richard Dudley [30, 31]. However this result has a long and rich history, and, as we shall describe briefly below, is far from being the last word on the subject.

Theorem 2.5.2. *Let f be a centered Gaussian field on a d-compact T, d the canonical metric, and H the corresponding entropy. Then there exists a universal constant K such that*

$$\mathbb{E}\left\{\sup_{t \in T} f_t\right\} \leq K \int_0^{\mathrm{diam}(T)/2} H^{1/2}(\varepsilon)\, d\varepsilon, \qquad (2.5.5)$$

and

$$\mathbb{E}\left\{\omega_{f,d}(\delta)\right\} \leq K \int_0^{\delta} H^{1/2}(\varepsilon)\, d\varepsilon, \qquad (2.5.6)$$

where

$$\omega_f(\delta) \equiv \omega_{f,d}(\delta) \stackrel{\Delta}{=} \sup_{d(s,t) \leq \delta} |f(t) - f(s)|, \quad \delta > 0, \qquad (2.5.7)$$

is the modulus of continuity of f on a T with respect to the canonical metric d. Furthermore, there exists a random $\eta \in (0, \infty)$ and a universal constant K such that

$$\omega_{f,d}(\delta) \leq K \int_0^{\delta} H^{1/2}(\varepsilon)\, d\varepsilon, \qquad (2.5.8)$$

for all $\delta < \eta$.

We are not going to attempt to prove this general result here, since you can find a full proof in Section 1.3 of *RFG*. Furthermore, the proof, while

not difficult, is a little to technical for our current purposes. Nevertheless, we encourage you to go to *RFG* for the details.

Reading Theorem 2.5.2 carefully, you will note that there is no claim that the sufficient conditions given there are also necessary and, indeed, they generally are not[36]. However, when f is a stationary, centered, Gaussian process there is such a result, and then it can be shown that

$$f \text{ is a.s. continuous on } T \iff f \text{ is a.s. bounded on } T \qquad (2.5.9)$$
$$\iff \int_0^\infty H^{1/2}(\varepsilon)\, d\varepsilon \ < \ \infty.$$

We shall not go into further details for the general case, but now do want to see how to apply these results to Gaussian random fields on $\mathbb{R}^N$.

2.5.2 Gaussian Fields on $\mathbb{R}^N$

Returning to Euclidean space after the abstraction of entropy on general metric spaces, it is natural to expect that conditions for continuity and boundedness will become so simple to both state and prove that there was really no need to introduce such abstruse general concepts.

This expectation is both true and false. Although we did not actually give detailed proofs here, it turns out that avoiding the notion of entropy does not make it any easier to establish continuity theorems, and, indeed, reliance on the specific geometry of the parameter space often confounds the basic issues.

On the other hand, the basic results for Gaussian processes on $\mathbb{R}^N$ are easy to state without specifically referring to any abstract notions. We shall do this first for continuity, and then for differentiability. In both, we assume that f is a centered Gaussian process with continuous covariance function C defined on a compact $T \subset \mathbb{R}^N$.

Continuity

We start by defining a function which captures the size of increments, by setting

$$p^2(u) \overset{\Delta}{=} \sup_{|s-t|\leq u} \mathbb{E}\left\{|f_s - f_t|^2\right\} \qquad (2.5.10)$$
$$= \sup_{|s-t|\leq u} \left[C(t,t) + C(s,s) - 2C(s,t) \right].$$

If f is stationary, then

---

[36] There is a theory which gives necessary and sufficient conditions for boundedness and continuity of general Gaussian processes, based on the ideas of *majorising measures*. For details, see, for example, [53, 89].

$$p^2(u) \;=\; 2 \sup_{|t| \leq u} [C(0) - C(t)], \qquad (2.5.11)$$

and so one sees that the rate of convergence on p to zero as $u \to 0$ is closely related to the smoothness of C at the origin discussed above.

Here is the main result on continuity of Gaussian random fields on $\mathbb{R}^N$.

Theorem 2.5.3. *If, for some $\delta > 0$, either*

$$\int_0^\delta (-\ln u)^{\frac{1}{2}} \, dp(u) < \infty \quad \text{or} \quad \int_\delta^\infty p\left(e^{-u^2}\right) du < \infty, \qquad (2.5.12)$$

then f is continuous and bounded on T with probability one. A sufficient condition for either integral in (2.5.12) to be finite is that, for some $0 < K < \infty$ and $\alpha, \eta > 0$,

$$\mathbb{E}\left\{|f_s - f_t|^2\right\} = C(t,t) + C(s,s) - 2C(s,t) \qquad (2.5.13)$$

$$\leq \frac{K}{|\log|s-t||^{1+\alpha}},$$

for all s, t with $|s - t| < \eta$. Furthermore, there exists a constant K', dependent only on the dimension N, and a random $\delta_o > 0$, such that, for all $\delta < \delta_o$,

$$\omega_f(\delta) \leq K' \int_0^{p(\delta)} (-\ln u)^{\frac{1}{2}} \, dp(u), \qquad (2.5.14)$$

where the modulus of continuity ω_f is as in (2.5.7), but taken with respect to the usual Euclidean metric rather than the canonical one. A similar bound, in the spirit of (2.5.6), holds for $\mathbb{E}\{\omega_f(\delta)\}$.

Proof. For once, we shall actually prove something in detail, the motivation being that it is worthwhile seeing how to convert the abstract entropy conditions to something more concrete.

Note first that since $p(u)$ is obviously non-decreasing in u, the Riemann-Stieljes integral (2.5.12) is well defined. The proof that both integrals in (2.5.12) converge and diverge together and that the convergence of both is assured by (2.5.13) is simple calculus and is Exercise 2.8.17. Of more significance is relating these integrals to the entropy integrals of Theorem 2.5.2. Indeed, all the claims of the theorem will follow from these results if we show that

$$\int_0^\delta H^{1/2}(\varepsilon) \, d\varepsilon \;\leq\; K \int_0^{p(\delta)} (-\ln u)^{\frac{1}{2}} \, dp(u) \qquad (2.5.15)$$

for small enough δ.

Since T is compact, we can enclose it in a N-cube C_L of side length $L = \max_{i=1,\dots,N} \sup_{s,t \in T} |t_i - s_i|$. Since p is non-decreasing, there is no problem in defining

$$p^{-1}(\varepsilon) \overset{\Delta}{=} \sup\{u : p(u) \le \varepsilon\}.$$

Now note that, for each $\varepsilon > 0$, the cube C_L, and so T, can be covered by $[1 + L\sqrt{N}/(2p^{-1}(\varepsilon))]^N$ (Euclidean) N-balls, each of which has radius no more than ε in the canonical metric d. Thus,

$$\int_0^\delta H^{1/2}(\varepsilon)\,d\varepsilon \le \sqrt{N} \int_0^\delta \left(\ln(1 + L\sqrt{N}/(2p^{-1}(\varepsilon)))\right)^{\frac{1}{2}}\,d\varepsilon$$

$$= \sqrt{N} \int_0^{p(\delta)} \left(\ln(1 + L\sqrt{N}/2u)\right)^{\frac{1}{2}}\,dp(u)$$

$$\le 2\sqrt{N} \int_0^{p(\delta)} (-\ln u)^{\frac{1}{2}}\,dp(u)$$

for small enough δ. This completes the proof. $\square$

The various sufficient conditions for continuity of Theorem 2.5.3 are quite sharp, but not necessary. There are two stages at which necessity is lost. One is simply that, as we mentioned earlier, entropy conditions, in general, need not be necessary in the non-stationary case. The second is that something is lost in the passage from entropy to the conditions on p. For an example of the latter, see Exercise 2.8.16.

Despite these drawbacks, the results of Theorem 2.5.3 are, from a practical point of view, reasonably definitive. For example, if f is stationary, then, following on from (2.5.9), it is possible to check that if

$$\frac{K_1}{(-\ln|t|)^{1+\alpha_1}} \le C(0) - C(t) \le \frac{K_2}{(-\ln|t|)^{1+\alpha_2}}, \qquad (2.5.16)$$

for $|t|$ small enough, then f will continuous if $\alpha_2 > 0$ and discontinuous if $\alpha_1 < 0$.

In practial situations, it is rare indeed that one even gets close to the logarithmic behavior of (2.5.13) or (2.5.16). The more common situation is that the covariance function has a power series representation of the form

$$C(s,t) = C(t,t) + (t-s)\Lambda_t(t-s)' + o\left(|t-s|^{2+\delta}\right), \qquad (2.5.17)$$

for $|t-s|$ small and some $\delta > 0$, or, in the stationary case

$$C(t) = C(0) + t\Lambda t' + o\left(|t|^{2+\delta}\right), \qquad (2.5.18)$$

for t in the neighborhood of the origin. The matrices Λ_t and Λ are $N \times N$ and positive definite. In the stationary case, the elements of Λ are the second order spectral moments (cf. (2.4.4)). In each of these cases the upper bound of (2.5.13) or (2.5.16) holds, with room to spare.

The condition (2.5.18) on the covariance function can easily be translated to spectral terms. It is standard Tauberian theory, which translates the behaviour of C at the origin to that of the spectral measure ν at infinity, to see that if the integral

$$\int_{\mathbb{R}^N} \left(\log(1 + |\lambda|) \right)^{1+\alpha} \nu(d\lambda) \tag{2.5.19}$$

converges for some $\alpha > 0$ then f is continuous, while if it diverges for some $\alpha < 0$ then f is discontinuous. (cf. Exercise 2.8.18.)

In other words, it is the behavior of the 'high frequency oscillations' in the spectral representation that are controlling the continuity/discontinuity dichotomy. This, of course, is what we suggested, on a purely heuristic basis, at the beginning of this section.

Differentiability

The step beyond continuity for a function on $\mathbb{R}^N$ is differentiability, which we investigate now, again in the Gaussian setting.

As for continuity, there are various notions of differentiability. We have already seen one, in the mean square setting of (2.4.6). Now, however, we would like to know when the limits in (2.4.6) not only exist, for all points $t \in T$ and for all directions t', with probability one, but when they are the same as what one gets by differentiating the sample paths of f in the usual sense. Furthermore, we would like to know when these derivatives are continuous.

It turns out that, at least in the Gaussian scenario, this question can be handled within the framework of basic continuity since derivatives, if they exist, must still be Gaussian. To see how to do this, first endow the space $\mathbb{R}^N \times \otimes^k \mathbb{R}^N$ with the norm

$$\|(s, s')\|_{N,k} \overset{\Delta}{=} |s| + \|s'\|_{\otimes^k \mathbb{R}^N} = |s| + \left(\sum_{i=1}^k |s'_i|^2 \right)^{1/2},$$

and write $B_{N,k}(y, h)$ for the ball centered at $y = (t, t')$ and of radius h in the metric induced by $\| \cdot \|_{N,k}$. Furthermore, write

$$T_{k,\rho} \overset{\Delta}{=} T \times \{t' : \|t'\|_{\otimes^k \mathbb{R}^N} \in (1 - \rho, 1 + \rho)\}$$

for the product of T with the ρ-tube around the unit sphere in $\otimes^k \mathbb{R}^N$. This is enough to allow us to formulate

Theorem 2.5.4. *Suppose f is a centered Gaussian random field on an open $T \in \mathbb{R}^N$, possessing k-th order partial derivatives in the L^2 sense in all directions everywhere inside T. Suppose, furthermore, that there exists $0 < K < \infty$, and $\rho, \delta, h_0 > 0$ such that for $0 < \eta_1, \eta_2, h < h_0$,*

$$\mathbb{E}\left\{ \left[\Delta^k f(t, \eta_1 t') - \Delta^k f(s, \eta_2 s') \right]^2 \right\} \tag{2.5.20}$$

$$< K \left| \ln \left(\|(t, t') - (s, s')\|_{N,k} + |\eta_1 - \eta_2| \right) \right|^{-(1+\delta)},$$

for all

$$((t, t'), (s, s')) \in T_{k,\rho} \times T_{k,\rho} : (s, s') \in B_{N,k}((t, t'), h),$$

where $\Delta^k f(t, t')$ is the symmetrized difference

$$\Delta^k f(t, t') = \sum_{s \in \{0,1\}^k} (-1)^{k - \sum_{i=1}^k s_i} f\left(t + \sum_{i=1}^k s_i t_i'\right)$$

Then, with probability one, f is k times continuously differentiable.

Proof. Recalling that we have assumed the existence of L^2 derivatives, we can define the Gaussian field

$$\widehat{f}(t, t', \eta) = \begin{cases} \Delta^k f(t, \eta t') & \eta \neq 0, \\ D_{L^2}^k f(t, t') & \eta = 0, \end{cases}$$

where $D_{L^2}^k f$ is the mean square derivative (2.4.6). This process is defined on the parameter space $\widehat{T} \overset{\Delta}{=} T_{k,\rho} \times (-h, h)$, an open subset of the finite dimensional vector space $\mathbb{R}^N \times \otimes^k \mathbb{R}^N \times \mathbb{R}$, with norm

$$\|(t, t', \eta)\|_{N,k,1} = \|(t, t')\|_{N,k} + |\eta|.$$

Whether or not f is k times differentiable on T is clearly the same issue as whether or not $\widehat{f}$ is continuous in $\widehat{T}$, with the issue of the continuity of $\widehat{f}$ really being only on the hyperplane where $\eta = 0$. But this puts us back into the setting of Theorem 2.5.3, and it is easy to check that condition (2.5.13) there translates to (2.5.20) in the current scenario. $\square$

The left hand side of (2.5.20) can, obviously, be written in terms of the covariance function C, although it becomes rather messy rather quickly. Consider what is perhaps the simplest of cases, in which $k = 1$, $t = s$, and $t' = s'$, so that we are looking at the simple, first order derivative of a random process f on the real line. Then (2.5.20) becomes

$$C(t + h, t + h) - C(t + h, t) - C(t, t + h) + C(t, t) \leq \frac{K}{|\log(|h|)|^{1+\delta}},$$

for all $h \in \mathbb{R}^1$ with $|h|$ sufficently small.

If we want to ensure that all first order partial derivatives of a random field on $\mathbb{R}^N$ exist and are continuous, then precisely the same condition suffices, with the understanding that now t and h are both in $\mathbb{R}^N$.

Second order partial derivatives of f of general order k require similar bounds, but the 'difference' of C on the left hand side will always be a difference of order $2k$.

As for continuity, it is rare in practice to get close to the upper bound in (2.5.20), and this condition will easily be satisfied if, in analogy to (2.5.17) and (2.5.18), the covariance function has a Taylor series expansion of up to order $2k$ with a remainder of $o(|h|^{2k+\eta})$ for some $\eta > 0$. There is also a corresponding spectral result, which you are asked to prove in Exercise 2.8.18.

2.5.3 Non-Gaussian Processes

A natural question to ask is whether or not the results and methods relating to continuity and differentiabilty that we have seen so far only for Gaussian random fields extend naturally to the non-Gaussian scenario.

In fact, for most of the processes that will concern us, this will not be terribly relevant, since we plan to concentrate on Gaussian related fields which can be written in the form

$$f(t) \; = \; F\left(g^1(t), \ldots, g^d(t)\right),$$

where the g^i are i.i.d. Gaussian and $F : \mathbb{R}^d \to \mathbb{R}$ is smooth. In this setting, continuity and boundedness of the non-Gaussian f follow deterministically from similar properties on F and the g^i, and so no additional theory is needed.

Nevertheless, there are many processes that are not attainable in this way, With these in mind, and for completeness, we state the following result, in which f_t is a random field on a parameter space T with a metric τ. (This is usually something corresponding to the canonical metric d of the Gaussian case.) Recall that a function $\varphi : \mathbb{R} \to \mathbb{R}$ is called a *Young function* if it is even, continuous, convex, and satisfies

$$\lim_{x \to 0} \frac{\varphi(x)}{x} = 0 \quad \text{and} \quad \lim_{x \to \infty} \frac{\varphi(x)}{x} = \infty.$$

Theorem 2.5.5. *Take f as above and let N_τ be the metric entropy function for T with respect to the metric τ. If there exist an $\alpha \in (0,1]$ and a Young function φ such that*

$$\mathbb{E}\left\{\varphi\left(\frac{\|f(t) - f(s)\|_B^\alpha}{\tau(s,t)}\right)\right\} \leq 1 \quad \text{and} \quad \int_{N_\tau(u) > 1} \varphi^{-1}\left(N_\tau(u)\right) du < \infty,$$

then f is continuous with probability one.

The best place to read about this is in Ledoux and Talagrand [53].

Note that, in the Gaussian case, once we had solved the issue of continuity, continuous differentiability was in essence a corollary. This will not be true in general, since then derivatives need not belong to the same class of processes as the original one.

2.6 Gaussian Exceedence Probabilities

If you have read this far, you know that one of the things that we care about most are the exceedence probabilities

$$\mathbb{P}\left\{\sup_{t \in T} f_t > u\right\} \tag{2.6.1}$$

for Gaussian random fields.

There are at least four ways to approach these probabilities.

(i) Choose a particular random field, and use its special properties to calculate the exceedence probabilities from first principles. This is what we did for the cosine process and field in Section 2.3.4. However, we already stated there, that, as far as differentiable processes are concerned, these are the only cases for which this direct approach works.

(ii) Search for general inequalities, that will always work. After all, basic statistics has its Chebychev inequality, martingale theory has its maximal inequalities, Markov processes have large deviations, and so on, so surely there should be something for Gaussian processes. In fact, there is, and we shall meet it in Theorem 2.6.1 below. It was discovered independently, and established with very different proofs, by Borell [22] and Csirelson, Ibragimov and Sudakov (CIS) [93]. For brevity, we shall call it the Borell-TIS inequality. However, like all blanket inequalities, while it provides an excellent tool for probabilists who wish to prove theorems, it is a poor tool for statisticians who need to see numbers.

(iii) Use comparison techniques. We do have exact results for the cosine processes. Perhaps these could be used as a basis for comparison. It seems reasonable, for example, that, other things (like means and variances) being equal, a random field with a weak correlation structure should have higher exceedence probabilities than a tightly correlated one. To a certain extent this is true, and the basic results quantifying this are Slepian's inequality, Theorem 2.6.3, and its extensions, below.

(iv) Finally, we could compromise. Rather than assuming everything about the process, as in (i), or nothing, as in (ii), we could assume perhaps a little more than we would like to, and obtain approximations to exceedence probabilities which, while not perfect, yield numbers that can be used by practicing statisticians.

In fact, (iv) is the path that we shall concentrate on later. However, since (ii) and (iii) are important, and often useful, we shall first invest a little time saying something about them, along with some of their extensions. We shall not give proofs. They can all be found in *RFG*.

2.6.1 Borell-TIS Inequality

In a notation that will remain fixed for the rest of the book, for any random field f on any parameter space T, set

$$\sigma_T^2 \triangleq \sup_{t \in T} \mathbb{E}\{f_t^2\}.$$

Then the Borell-TIS inequality[37] is

---

[37] Actually, Theorem 2.6.1 is not in the same form as Borell's original inequality, in which $\mathbb{E}\{\|f\|\}$ was replaced by the median of $\|f\|$. However, the two forms are equivalent.

Theorem 2.6.1 (Borell-TIS inequality). *Let f_t be a centered Gaussian process, a.s. bounded on T[38]. Write*

$$\|f\| = \|f\|_T \overset{\Delta}{=} \sup_{t \in T} f_t.$$

Then

$$\mathbb{E}\{\|f\|\} \; < \; \infty,$$

and, for all $u > 0$,

$$\mathbb{P}\left\{\|f\| - \mathbb{E}\{\|f\|\} \; > \; u\right\} \le e^{-u^2/2\sigma_T^2}. \tag{2.6.2}$$

Note that, despite the misleading notation, $\| \; \| \equiv \sup$ is not a norm, and that very often one needs bounds on the tail of $\sup_t |f_t|$, which does give a norm. However, symmetry immediately gives

$$\mathbb{P}\left\{\sup_t |f_t| \; > \; u\right\} \; \le \; 2\,\mathbb{P}\left\{\sup_t f_t \; > \; u\right\} \le 2e^{-u^2/2\sigma_T^2}. \tag{2.6.3}$$

To obtain an appreciation of how tight this very general inequality is, note that it immediately implies that

$$\mathbb{P}\{\|f\| \ge u\} \le e^{\mu_u - u^2/2\sigma_T^2}, \tag{2.6.4}$$

where $\mu_u = (2u\mathbb{E}\{\|f\|\} - [\mathbb{E}\{\|f\|\}]^2)/\sigma_T^2$, which tells us that, for high levels u, the dominant behavior of *all* Gaussian exceedence probabilities is determined by $e^{-u^2/2\sigma_T^2}$.

This is somewhat surprising, for if X is a single Gaussian variable with distribution $N(0, \sigma_T^2)$ then we already know (cf. (2.2.2)) that, for all $u > 0$,

$$\left(\frac{\sigma_T}{\sqrt{2\pi}u} - \frac{\sigma_T^3}{\sqrt{2\pi u^3}}\right) e^{-\frac{1}{2}u^2/\sigma_T^2} \; \le \; \mathbb{P}\{X > u\} \; \le \; \frac{\sigma_T}{\sqrt{2\pi}u} \, e^{-\frac{1}{2}u^2/\sigma_T^2}.$$

In other words, at high levels, the exceedence probability of a Gaussian random field is not that different from the exceedence probability of the field at the point of maximal variance.

For a stationary field, where the maximal variance is achieved at every point in T, this strengthens to the claim that, still at high levels, the exceedence probability of a Gaussian random field is not that different from the exceedence probability of the field *at any given point*.

Does this mean that we need go no further in studying Gaussian extrema? Not at all, for a number of reasons:

---

[38] The parameter space T in the Borell-TIS inequality is completely general, the only requirement being that it be compact with respect to the canonical metric d, in the sense of (2.5.3) .

(i)   First of all, while the exponent μ_u in (2.6.4) grows only linearly in u, e^{μ_u} grows much faster. It can therefore hardly be ignored, other than perhaps by a probabilist in the throes of a proof where "large u" is so large that linear terms are irrelevent when compared to quadratic, regardless of whatever constants may be around. Statisticians, however, will rarely feel this way.

(ii) If, in view of (i), we decide to keep track of the term e^{μ_u}, we find it involves the expectation of the supremum, and this is hard to obtain. While it is true that under entropy conditions we did manage to find an upper bound for it (cf. Theorem 2.5.2) this bound involved an admittedly universal, but essentially unknown[39], constant.

(iii) As we shall see later, for most smooth random fields the factor of e^{μ_u}, which is generally $O(e^{cu})$, can be replaced for a factor of the form Cu^α, where both C and α are explicitly computable. Furthermore, it can be shown that this is the correct order of growth. This is where expected Euler characteristics will come into their own, and justify the claims that we made back in Chapter 1 (cf. (1.5.12)).

As an example of how easy it is to improve on the Borell-TIS bound in the spirit of (ii), when a little more is assumed, consider the following, definitely sub-optimal, result.

Theorem 2.6.2. *Let be f a centered, a.s. continuous Gaussian field over T with entropy function N. If $N(\varepsilon) \leq K\varepsilon^{-\alpha}$, then, for all sufficiently large u,*

$$\mathbb{P}\left\{\sup_{t\in T} f(t) \geq u\right\} \leq C_\alpha\, u^{\alpha+\eta}\, e^{-u^2/2\sigma_T^2}, \qquad (2.6.5)$$

for every $\eta > 0$, where $C_\alpha = C(K, \alpha, \sigma_T^2)$ is a finite constant.

Proof. Take $\varepsilon > 0$ and define

$$\mu(t, \varepsilon) = \mathbb{E}\left\{\sup_{s\in B_d(t,\varepsilon)} f_s\right\},$$

and

$$\mu(\varepsilon) = \sup_{t\in T} \mu(t, \varepsilon),$$

where $B_d(t, \varepsilon)$ is a ball of radius ε around t in the canonical metric d of (2.5.1). Since $N(\varepsilon)$ balls of radius ε cover T, it is an immediate consequence of the Borell-TIS inequality that, for $u > \mu(\varepsilon)$,

---

[39] Actually, there are known values for this constant. However, they are so large as to yield bounds that are effectively valueless from the point of view of generating useful numerical bounds to exceedence probabilities.

$$\mathbb{P}\left\{\sup_{t\in T} f(t) \geq u\right\} \leq N(\varepsilon)\, e^{-\frac{1}{2}(u-\mu(\varepsilon))^2/\sigma_T^2}. \qquad (2.6.6)$$

Allowing $C = C(\alpha)$ to denote a constant, dependent only on α, that may change from line to line, we have from Theorem 2.5.2 that

$$(2.6.7)$$

$$\mu(t,\varepsilon) \leq C \int_0^\varepsilon \left(\log N(\varepsilon)\right)^{\frac{1}{2}} d\varepsilon \leq C \int_0^\varepsilon \left(\log(1/\varepsilon)\right)^{\frac{1}{2}} d\varepsilon \leq C\, \varepsilon\, \sqrt{\log(1/\varepsilon)}.$$

Set $\varepsilon = \varepsilon(u) = u^{-1}$, choose u large enough so that $u > Cu^{-1}\sqrt{\log u}$ and substitute into $(2.6.6)$ to obtain

$$\mathbb{P}\left\{\sup_{t\in T} f(t) \geq u\right\} \leq C_1\, u^\alpha\, e^{-\frac{1}{2}\left(u - C_2 u^{-1}\sqrt{\log u}\,\right)^2/\sigma_T^2}$$

$$\leq C_3\, u^\alpha\, e^{C_4\sqrt{\log u}} e^{-u^2/2\sigma_T^2}.$$

Since for $\eta > 0$ and u large enough $e^{C\sqrt{\log u}} < u^\eta$, this gives us $(2.6.5)$ and so completes the proof. $\square$

One can do much better than Theorem 2.6.2, by assuming a little more on the entropy function, or by working a little harder. However, the basic idea should be clear from this one simple result and its proof. In particular, in the situations in which we shall generally be working, of smooth, twice differentiable functions, power law behavior for the entropy function is always assured (cf. the proof of Theorem 2.5.3). Thus, in those cases, we can expect bounds on exceedence probabilities which are much better than those given by the Borell-TIS inequality. On the other hand, it is going to be difficult to get bounds with useful constants using only entropy arguments.

2.6.2 Comparison Inequalities

Having now seen that it is not going to be easy to get good, quantitative estimates for exceedence probabilities, we can now investigate the second path, that of using comparison with known cases.

The theory of Gaussian processes is rich in comparison inequalities, where by this term we mean results of the form "if f is a 'rougher' process than g, and both are defined over the same parameter space, then $\|f\|$ will generally be 'larger' than $\|g\|$", where, as in the previous section, we write $\|\cdot\|$ for supremum. The most basic of these results is Slepian's inequality, which, like the Borell-TIS inequality, holds for all parameter spaces T.

Theorem 2.6.3 (Slepian's inequality). *If f and g are a.s. bounded, centered Gaussian processes on T such that $\mathbb{E}\{f_t^2\} = \mathbb{E}\{g_t^2\}$ for all $t \in T$ and*

$$\mathbb{E}\left\{(f_t - f_s)^2\right\} \leq \mathbb{E}\left\{(g_t - g_s)^2\right\}, \qquad (2.6.8)$$

for all $s, t \in T$, then for all real u

$$\mathbb{P}\{\|f\| > u\} \leq \mathbb{P}\{\|g\| > u\}. \qquad (2.6.9)$$

Furthermore,

$$\mathbb{E}\{\|f\|\} \leq \mathbb{E}\{\|g\|\}. \qquad (2.6.10)$$

Slepian's inequality is so natural that it hardly seems to require a proof, and hardly the rather analytic, non-probabilistic one that will follow. To see that there is more to the story than meets the eye, one need only note that (2.6.9) does *not* follow from (2.6.8) if we replace $\sup_T f_t$ by $\sup_T |f_t|$ and $\sup_T g_t$ by $\sup_T |g_t|$ (cf. Exercise 2.8.19).

Furthermore, the proof of Slepian's inequality does not use the above 'obvious' heuristic argument at all, but uses little more than basic calculus and some approximation arguments. Since the proof is not long, and the result is important, we shall give a full proof. The proof is based on the following lemma, the proof of which, in all its important details, goes back to Slepian's original paper [84].

Lemma 2.6.4. *Let $f_1, ..., f_k$ be centered Gaussian variables with covariance matrix $C = (c_{ij})_{i,j=1}^k$, $c_{ij} = \mathbb{E}\{f_i f_j\}$. Let $h : \mathbb{R}^k \to \mathbb{R}$ be C^2, and assume that, together with its derivatives, it satisfies a $o(|x|^d)$ growth condition at infinity for some finite d. Let*

$$H(C) = \mathbb{E}\{h(f_1, \ldots, f_k)\}, \qquad (2.6.11)$$

and assume that for a pair (i, j), $1 \leq i < j \leq k$

$$\frac{\partial^2 h(x)}{\partial x_i \partial x_j} \geq 0 \qquad (2.6.12)$$

for all $x \in \mathbb{R}^k$. Then $H(C)$ is an increasing function of c_{ij}.

Proof. We have to show that

$$\frac{\partial H(C)}{\partial c_{ij}} \geq 0$$

whenever $\partial^2 h / \partial x_i \partial x_j \geq 0$.

To make our lives a little easier we assume that C is non-singular, so that it makes sense to write $\varphi(x) = \varphi_C(x)$ for the centered Gaussian density on $\mathbb{R}^k$ with covariance matrix C. Straightforward algebra shows that[40]

$$\frac{\partial \varphi}{\partial c_{ii}} = \frac{1}{2} \frac{\partial^2 \varphi}{\partial x_i^2}, \qquad \frac{\partial \varphi}{\partial c_{ij}} = \frac{\partial^2 \varphi}{\partial x_i \partial x_j}, \qquad i \neq j. \qquad (2.6.13)$$

---

[40] This is, of course, little more that the heat equation of PDE theory.

Applying this and our assumptions on h to justify two integrations by parts, we obtain, for $i \neq j$,

$$\frac{\partial H(C)}{\partial c_{ij}} = \int_{\mathbb{R}^k} h(x)\, \frac{\partial \varphi(x)}{\partial c_{ij}}\, dx = \int_{\mathbb{R}^k} \frac{\partial^2 h(x)}{\partial x_i \partial x_j}\, \varphi(x)\, dx \geq 0.$$

This completes the proof for the case of non-singular C. The general case can be handled by approximating a singular C via a sequence of non-singular covariance matrices. $\qquad\square$

Proof of Theorem 2.6.3 Actually, we are only going to prove the main inequality, (2.6.9), for T finite. The extension to general T can be found in *RFG*, and while slightly technical, is not hard.

Note that since $\mathbb{E}\{f_t^2\} = \mathbb{E}\{g_t^2\}$ for all $t \in T$, (2.6.8) implies that $\mathbb{E}\{f_s f_t\} \geq \mathbb{E}\{g_s g_t\}$ for all $s, t \in T$. Let $h(x) = \prod_{i=1}^k h_i(x_i)$, where each h_i is a positive non-increasing, C^2 function satisfying the growth conditions placed on h in the statement of Lemma 2.6.4, and k is the number of points in T. Note that, for $i \neq j$

$$\frac{\partial^2 h(x)}{\partial x_i \partial x_j} \;=\; h_i'(x_i) h_j'(x_j) \prod_{\substack{n \neq i \\ n \neq j}} h_n(x_n) \;\geq\; 0,$$

since both h_i' and h_j' are non-positive. It therefore follows from Lemma 2.6.4 that

$$\mathbb{E}\left\{ \prod_{i=1}^k h_i(f_i) \right\} \geq \mathbb{E}\left\{ \prod_{i=1}^k h_i(g_i) \right\}. \tag{2.6.14}$$

Now take $\{h_i^{(n)}\}_{n=1}^\infty$ to be a sequence of positive, non-increasing, C^2 approximations to the indicator function of the interval $(-\infty, \lambda]$, to derive that

$$\mathbb{P}\{\|f\| < u\} \geq \mathbb{P}\{\|g\| < u\},$$

which implies (2.6.9).

To complete the proof, all that remains is to show that (2.6.9) implies (2.6.10). But this is a simple consequence of integration by parts, since

$$
\begin{aligned}
\mathbb{E}\{\|f\|\} &= \int_0^\infty \mathbb{P}\{\|f\| > u\}\, du \;-\; \int_{-\infty}^0 \mathbb{P}\{\|f\| < u\}\, du \\
&\leq \int_0^\infty \mathbb{P}\{\|g\| > u\}\, du \;-\; \int_{-\infty}^0 \mathbb{P}\{\|g\| < u\}\, du \\
&= \mathbb{E}\{\|g\|\}.
\end{aligned}
$$

This completes the proof. $\qquad\square$

There are many extensions of Slepian's inequality, the most important of which is probably the following, which we shall not attempt to prove.

Theorem 2.6.5 (Sudakov-Fernique inequality).   *Let f and g be a.s. bounded Gaussian processes on T. If*

$$\mathbb{E}\{f_t\} = \mathbb{E}\{g_t\}$$

and

$$\mathbb{E}\left\{(f_t - f_s)^2\right\} \;\leq\; \mathbb{E}\left\{(g_t - g_s)^2\right\}$$

for all $s, t \in T$, then

$$\mathbb{E}\{\|f\|\} \leq \mathbb{E}\{\|g\|\}. \tag{2.6.15}$$

In other words, a Slepian-like inequality holds without a need to assume either zero mean or identical variance for the compared processes. However, in this case we have only the weaker ordering of expectations of (2.6.10) and not the stochastic domination of (2.6.9).

2.6.3 Exceedence Probabilities for Smooth Processes

Now that you have had a little taste of the general theory, the time has come so say something about the cases that will be of central interest to us, those in which the random field has smooth sample paths.

Perhaps the first result of this kind, in the setting of random fields, goes back to [18, 19] and treats stationary, zero mean, Gaussian random fields on $\mathbb{R}^N$ with covariance functions which, near the origin, can be written as

$$C(t) = 1 - t\Lambda t' + o(|t|^2), \tag{2.6.16}$$

where Λ is the matrix of second order spectral moments (cf. (2.4.4)) and, for convenience, we have also assumed unit variance. We already know that these random fields are both continuous with probability one, and differentiable in mean square.

In this case, one can show[41] that, for N-dimensional rectangles $T = \prod_{i=1}^{N}[0, T_i]$,

$$\lim_{u \to \infty} \frac{\mathbb{P}\left\{\sup_{t \in T} f_t \geq u\right\}}{u^N \Psi(u)} = \frac{|T|\,|\Lambda|^{1/2}}{(2\pi)^{N/2}}, \tag{2.6.17}$$

where $|T| = \prod T_i$ is the volume of T, but $|\Lambda|$ is the determinant of Λ.

---

[41] You can learn a lot more about results of this kind, and how to prove them, in a variety of places. The book by Leadbetter, Lindgren and Rootzén [51] treats mainly processes on the real line, but is very readable. Piterbarg's monograph [75] is harder going, but treats random fields on $\mathbb{R}^N$ as well as processes on the line. There you will also find a detailed treatment of the so-called 'double-sum method' which can be used to compute results like (2.6.17) for a far wider collection of random fields.

In view of what we know about the tail probabilities $\Psi(u)$ (cf. (2.2.2)) another way to write this result would be

$$\mathbb{P}\left\{\sup_{t\in T} f_t \geq u\right\} = u^{N-1}e^{-u^2/2}\left[\frac{|T||\Lambda|^{1/2}}{(2\pi)^{(N+1)/2}} + R(u)\right].$$

The remainder term $R(u)$ tends to 0 as $u \to \infty$, but otherwise (2.6.17) tells us nothing about it.

On the other hand, if one assumes just a little more, that f is *twice* continuously differentiable, it is possible to show that there exist explicitly computable constants α, n and C_j such that

$$\mathbb{P}\left\{\sup_{t\in T} f(t) \geq u\right\} = \Psi(u) + u^\alpha\, e^{-u^2/2}\left[\sum_{j=0}^{n} C_j u^{-j} + \text{error}\right], \qquad (2.6.18)$$

where the error is small for large u. Furthermore, one can get a good handle on the error term, which turns out to be much smaller than expected[42].

Indeed, while, for convenience, we have been assuming stationarity above, results like (2.6.18) hold in quite wide generality and also without the assumption that T be a simple rectangle in $\mathbb{R}^N$.

To explain all of this properly, however, needs an excursion into geometry, which is the content of the next chapter, followed by some hard computation, which is the content of Chapter 4. We shall return to extremal problems, and results like (2.6.18), only in Chapter 5.

2.7 An Expectation Meta-Theorem: The Rice-Kac Formula

In this section, we quote a basic tool, the Rice-Kac formula, which we shall use later, in Chapter 4, to compute the expectations of certain functionals of the excursion sets of smooth random fields. Consider two vector-valued random fields $f = (f^1, \ldots, f^N)$ and $g = (g^1, \ldots, g^K)$ defined on some compact set $T \subset \mathbb{R}^N$ with non-empty interior. For $B \subset \mathbb{R}^K$, we need formulae for the expectations

$$\mathbb{E}\big\{\#\,\{t \in T : f(t) = u,\ g(t) \in B\}\big\}. \qquad (2.7.1)$$

Perhaps the most basic application of (2.7.1) is to prove the famous Rice, or Rice-Kac formula[43]. The Rice formula gives an expression for the number

---

[42] Since (2.6.18) looks like the beginning of a power series expansion, one would 'expect' the error term to be $o(u^{-(n+1)})$. In fact, it turns out to be $o(e^{-\delta u^2})$, for an identifiable $\delta > 0$. This, of course, is much smaller, and is reminiscent of what we saw for the cosine field at (2.3.16).

[43] In fact, the name Rice-Kac formula comes from the fact that the first version of the above formula was used in exactly this way in [45] and [77].

of upcrossings of the level u of a real-valued process h on the line where an upcrossing is defined as a point t where $h(t) = u$ and $h(t)$ is increasing. In this example, we set $f = h$, $g = \dot{h}$ and $B = [0, +\infty)$. See Example 2.8.21.

However, we shall require these expectations for a number of reasons. For example, in Chapter 3 we shall see how, using Morse theory, critical points of random fields above a level u, which are local features, can be used to compute certain global properties of excursion sets. In turn, these can be related to the exceedence probabilities (2.6.1) of f above the level u.

To see how (2.7.1) helps, note that that the critical points of a random field Z, say, are solutions to the equation $\nabla Z = 0$. If we are to count critical points above the level Z suggests setting $f = \nabla Z$, $g = Z$ and $B = [u, +\infty)$. The expression (2.7.1) therefore provides the expected number of critical points of Z above the level u. See Example 2.8.23 on applying this formula to maxima in one dimension.

We are now ready to quote the theorem. Below, ∇f denotes the derivative of f. Since f takes values in $\mathbb{R}^N$ this is now a $N \times N$ matrix of first-order partial derivatives of f; i.e.

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

Theorem 2.7.1. *Let f, g, T and B be as above, with the additional assumption that the boundaries of T and B have finite $N-1$ and $K-1$ dimensional measures, respectively. Furthermore, assume that the following conditions are satisfied for some $u \in \mathbb{R}^N$:*

(a) *All components of f, ∇f, and g are a.s. continuous and have finite variances (over T).*

(b) *For all $t \in T$, the marginal densities $p_t(x)$ of $f(t)$ (implicitly assumed to exist) are continuous at $x = u$.*

(c) *The conditional densities $p_t(x|\nabla f(t), g(t))$ of $f(t)$ given $g(t)$ and $\nabla f(t)$ (implicitly assumed to exist) are bounded above and continuous at $x = u$, uniformly in $t \in T$.*

(d) *The conditional densities $p_t(z|f(t) = x)$ of $\det \nabla f(t)$ given $f(t) = x$, are continuous for z and x in neighbourhoods of 0 and u, respectively, uniformly in $t \in T$.*

(e) *The conditional densities $p_t(z|f(t) = x)$ of $g(t)$ given $f(t) = x$, are continuous for all z and for x in a neighbourhood u, uniformly in $t \in T$.*

(f) *The following moment condition holds:*

$$\sup_{t \in T} \max_{1 \leq i,j \leq N} \mathbb{E}\left\{ \left| f_j^i(t) \right|^N \right\} < \infty. \tag{2.7.2}$$

(g) *The moduli of continuity with respect to the usual Euclidean norm (cf. (2.5.7)) of each of the components of f, ∇f, and g satisfy*

$$\mathbb{P}\left\{\omega(\eta) > \varepsilon\right\} = o\left(\eta^N\right), \qquad as\ \eta \downarrow 0, \tag{2.7.3}$$

for any $\varepsilon > 0$.

Then, if

$$N_u \equiv N_u(T) \equiv N_u(f, g : T, B)$$

denotes the number of points in T for which

$$f(t) = u \in \mathbb{R}^N \quad and \quad g(t) \in B \subset \mathbb{R}^K,$$

and $p_t(x, \nabla y, v)$ denotes the joint density of $(f_t, \nabla f_t, g_t)$, we have, with $D = N(N+1)/2 + K$,

$$\mathbb{E}\{N_u\} = \int_T \int_{\mathbb{R}^D} |\det \nabla y|\ \mathbb{1}_B(v)\, p_t(u, \nabla y, v)\, d(\nabla y)\, dv\, dt. \tag{2.7.4}$$

It is sometimes more convenient to write this as

$$\mathbb{E}\{N_u\} = \int_T \mathbb{E}\left\{ |\det \nabla f(t)|\, \mathbb{1}_B(g(t)) \,\Big|\, f(t) = u \right\} p_t(u)\, dt, \tag{2.7.5}$$

where p_t here is the density of $f(t)$.

In our applications of Theorem 2.7.1, the expression (2.7.5) will be the principal form used.

For a full proof, see *RFG*, where this result appears as Theorem 11.2.1. An outline of the beginning of this proof, that at least shows from where the result comes, will be given in a moment.

Conditions (a)–(g) are often tedious to check, but almost disappear when both f and g are either Gaussian, or simple functions of Gaussian (vector-valued) Gaussian random fields. In these situations, the primary consideration becomes one of sample path continuity and differentiability, which we already looked at in some detail in Section 2.5.1. Here is the Gaussian result.

Corollary 2.7.2. *Let f and g be centered Gaussian fields with T and B satisfying the conditions of Theorem 2.7.1. Assume that, for each $t \in T$, the joint distribution of $(f(t), \nabla f(t), g(t))$ is non-degenerate.*

Write $C_f^i = C_f^i(s,t)$ for the covariance function of f^i, $C_{f_j}^i = \partial^2 C_f^i / \partial s_j \partial t_j$ for the covariance function of $f_j^i = \partial f^i / \partial t_j$, and C_g^i for the covariance function of g^i. If

$$\max_{i,j} \left| C_{f_j}^i(t,t) + C_{f_j}^i(s,s) - 2C_{f_j}^i(s,t) \right| \leq K\, |\ln|t-s||^{-(1+\alpha)},$$
$$\max_i \left| C_g^i(t,t) + C_g^i(s,s) - 2C_g^i(s,t) \right| \leq K\, |\ln|t-s||^{-(1+\alpha)}, \tag{2.7.6}$$

for some finite $K > 0$, some $\alpha > 0$ and all $|t-s|$ small enough, then the conclusions of Theorem 2.7.1 hold.

To close this Chapter, we look at a partial proof of Theorem 2.7.1. For a start, however, you should try Exercise 2.8.24, which shows that, if $T = [0,1]^N$, then, under conditions (a) and (b) of Theorem 2.7.1, there are no points $t \in \partial T$ satisfying $f(t) = u$. We shall use this and other regularity properties of f and g in what follows. This is the easiest one to prove.

To start, let $\delta_\varepsilon : \mathbb{R}^N \to \mathbb{R}$ be an approximate delta function, constant on the N-ball $B(\varepsilon) = \{t \in \mathbb{R}^N : |t| < \varepsilon\}$, zero elsewhere, and normalized so that

$$\int_{B(\varepsilon)} \delta_\varepsilon(t) \, dt \;=\; 1. \tag{2.7.7}$$

We then claim that

$$N_u(f, g; T, B) \;=\; \lim_{\varepsilon \to 0} \int_T \delta_\varepsilon(f(t) - u) \, \mathbb{1}_B(g(t)) \, |\mathrm{det} \nabla f(t)| \, dt. \tag{2.7.8}$$

If this is true, then, with no further pretense to rigor, take expectations on both sides and freely change the orders of limit and expectation to find that

$$\mathbb{E}\{N_u\} = \lim_{\varepsilon \to 0} \mathbb{E} \int_T \delta_\varepsilon(f(t) - u) \, \mathbb{1}_B(g(t)) \, |\mathrm{det}\, \nabla f(t)| \, dt$$

$$= \int_T \int_{\mathbb{R}^{N(N+1)/2}} \int_{\mathbb{R}^K} \mathbb{1}_B(v) \, |\mathrm{det} \nabla y|$$

$$\times \left\{ \lim_{\varepsilon \to 0} \int_{\mathbb{R}^N} \delta_\varepsilon(x - u) \, p_t(x, \nabla y, v) \, dx \right\} d\nabla y \, dv \, dt,$$

where the p_t are the obvious densities. Taking the limit in the innermost integral yields

$$\mathbb{E}\{N_u\} = \int_T \int \int \mathbb{1}_B(v) \, |\mathrm{det} \nabla y| \, p_t(u, \nabla y, v) \, d\nabla y \, dv \, dt$$

$$= \int_T \mathbb{E}\{|\mathrm{det} \nabla f(t)| \, \mathbb{1}_B(g(t)) \, \big| \, f(t) = u\} \, p_t(u) \, dt,$$

which is what we wanted to show.

Of course, interchanging the order of integration and the limiting procedure requires justification, and not only is it far from trivial it is in fact so hard to do that a fully rigorous proof requires a rather different approach. See Theorem 11.2.1 of RFG for details.

Now, however, we return to the proof of (2.7.8).

To save on notation, and without any loss of generality, we take $u = 0$. Consider those $t \in T$ for which $f(t) = 0$, of which we claim (without proof) that there is only a finite number. Furthermore, by Exercise 2.8.24, none of them lie in ∂T. Consequently, each one can be surrounded by an open ball, of radius η, say, in such a way that the balls neither overlap nor intersect ∂T. Furthermore (again as an unproven consequence of the assumptions) we can

take η small enough so that within each ball $g(t)$ always lies in either B or the interior of its complement, but never both.

Let $\sigma(\varepsilon)$ be the ball $|f| < \varepsilon$ in the image space of f. From what we have just claimed follows the fact that we can also choose ε small enough for the inverse image of $\sigma(\varepsilon)$ in T to be contained within the union of the η spheres.

Furthermore, by the inverse mapping theorem we can choose ε, η so small that, for each η sphere in T, $\sigma(\varepsilon)$ is contained in the f image of the η sphere, so that the restriction of f to such a sphere will be one-one. Since the Jacobian of the mapping of each η sphere by f is $|\det \nabla f(t)|$ it follows that we can choose ε small enough so that

$$N_0 \ = \ \int_T \delta_\varepsilon(f(t)) \, \mathbb{1}_B(g(t)) \, |\det \nabla f(t)| \, dt.$$

This follows since each η sphere in T over which $g(t) \in B$ will contribute exactly one unit to the integral, while all points outside the η spheres will not be mapped onto $\sigma(\varepsilon)$. Since the left-hand side of this expression is independent of ε we can take the limit as $\varepsilon \to 0$ to obtain (2.7.8), as required.

2.8 Exercises

Exercise 2.8.1. Prove the basic, but very important, Gaussian inequality (2.2.2), that

$$\left(\frac{1}{x} - \frac{1}{x^3} \right) \varphi(x) < \Psi(x) < \frac{1}{x} \varphi(x),$$

for all $x > 0$.

Exercise 2.8.2. Show that the conditional distributions of multivariate Gaussian variables, are also Gaussian, with mean vectors given by (2.2.6) and covariance matrices given by (2.2.7).
Hint: Take

$$A = \begin{pmatrix} I_n & -C_{11}^{-1} C_{12} \\ 0 & I_{d-n} \end{pmatrix}$$

in (2.2.5) and define $Y = (Y^1, Y^2) = XA$, where Y^1 has length n. Check that $Y^1 \equiv X^1$ and Y^2 are independent and use this to obtain (2.2.6) and (2.2.7) for $i = 2$, $j = 1$.

Exercise 2.8.3. Prove the Wick product formula for Gaussian random variables, which says the following:

Let $X_1, X_2, \ldots, X_n$ be a set of real-valued random variables having a joint Gaussian distribution and zero means. Then, for any integer m,

$$\mathbb{E}\{X_1 X_2 \cdots X_{2m+1}\} = 0,$$
$$\mathbb{E}\{X_1 X_2 \cdots X_{2m}\} = \sum \mathbb{E}\{X_{i_1} X_{i_2}\} \cdots \mathbb{E}\{X_{i_{2m-1}} X_{i_{2m}}\},$$

where the sum is taken over the $(2m)!/m!\,2^m$ different ways of grouping $X_1, \ldots, X_{2m}$ into m pairs.
Hint: Use characteristic functions.

Exercise 2.8.4. Using nothing but the definition (2.2.9) of the covariance function,

(i) Show that covariance functions of random processes must always be non-negative definite. (cf. Footnote 3 for a definition of non-negative definiteness.)
(ii) Find a simple condition on a random process for its covariance function to be positive definite.

Exercise 2.8.5. Compute the covariance function of the cosine random field (2.3.11), and show that it is both stationary and isotropic.

Exercise 2.8.6. Let f be the cosine field of (2.3.11) with the ξ_k and ξ_k' all independent $N(0, \sigma^2)$.

(i) Following the argument described there, derive the density (2.3.12) for the supremum, identifying the C_{nk}.
Hint: Use characteristic functions to handle the convolution.
(ii) Using (2.3.12) (and quite a lot of calculus) establish (2.3.16).

Exercise 2.8.7. In the statement of the spectral distribution theorem, Theorem 2.4.1, there is an implicit assumption that every non-negative definite function is also a covariance function. Show that this is true.

Exercise 2.8.8. Prove that (2.4.8) and (2.4.9) hold, under condition (2.4.7).
Hint: It is easiest, but not necessary, to use both the spectral distribution theorem *and* the spectral respresentation theorem.
Warning: Don't forget that $i^2 = -1$ if you want to get the power of -1 correct!

Exercise 2.8.9. Let f be a stationary, C^2, random field, ∇f the vector of its first order partial derivatives, and $\nabla^2 f(t)$ the matrix of its second order derivatives.

(i) Show that the covariance of of $(f(t), \nabla f(t), \nabla^2 f(t))$ can be written as

$$\mathrm{Var}\begin{pmatrix} f(t) \\ \nabla f(t) \\ \nabla^2 f(t) \end{pmatrix} = \begin{pmatrix} \sigma^2 & 0 & -\Lambda \\ 0 & \Lambda & 0 \\ -\Lambda & 0 & \mathcal{E} \end{pmatrix} \qquad (2.8.1)$$

where

$$\Lambda_{ij} = \mathbb{E}\{f_i(t) f_j(t)\}$$

and

$$\mathcal{E}_{ij,kl} \triangleq \mathbb{E}\{f_{ij}(t) f_{kl}(t)\}.$$

(ii) Express Λ and $\mathcal{E}$ in terms of the spectral measure (2.4.2) of f.

(iii) Show that $\mathcal{E}$ is symmetric in i, j, k, l.

(iv) How do these covariances simplify if f is isotropic?

Exercise 2.8.10. Taking W to be the Brownian motion on $[0, 1]$ – i.e. the zero mean Gaussian process with covariance function $C(s, t) = s \wedge \overset{\Delta}{=} \min(s, t)$ – show that integral equation defining the eigenvalues λ_n and eigenfunctions ψ_n needed for finding a Karhunen-Loève expansion for W (i.e. (2.3.20)) is

$$\lambda \psi(t) = \int_0^1 \min(s, t)\, \psi(s)\, ds = \int_0^t s\, \psi(s)\, ds \; + \; t \int_t^1 \psi(s)\, ds.$$

Differentiate both sides twice with respect to t to find a second order ordinary differential equation whose solution, together with the appropriate conditions, gives

$$\psi_n(t) \;=\; \sqrt{2} \sin\left(\tfrac{1}{2}(2n+1)\pi t\right), \quad \lambda_n \;=\; \left(\frac{2}{(2n+1)\pi}\right)^2.$$

Exercise 2.8.11. Let W be a Gaussian white noise based on Lebesgue measure, and use it to define a random field on $\mathbb{R}_+^N = \{(t_1, ..., t_N) : t_i \geq 0\}$ by setting

$$W(t) \;=\; W\left([0, t]\right), \tag{2.8.2}$$

where $[0, t]$ is the rectangle $\prod_{i=1}^N [0, t_i]$. W_t is called the Brownian sheet on $\mathbb{R}_+^N$, or *multiparameter Brownian motion*. If $N = 1$ it is standard Brownian motion.

(i) Show that W is a centered Gaussian field on $\mathbb{R}_+^N$ with covariance

$$\mathbb{E}\{W_s W_t\} \;=\; (s_1 \wedge t_1) \times \cdots \times (s_N \wedge t_N), \tag{2.8.3}$$

where $s \wedge t \overset{\Delta}{=} \min(s, t)$.

(ii) Suppose $N > 1$, and fix $N - k$ of the indices. Show that W is a scaled k-parameter Brownian sheet in the remaining variables.

(iii) Using the result of Exercise 2.8.10, find a Karhunen-Loève expansion for W on $[0, 1]^N$.

Exercise 2.8.12. Show that for an isotropic random field on $\mathbb{R}^N$ the orthogonal expansion (2.3.17) can never be finite.

Exercise 2.8.13. By using the fact that covariance functions must be non-negative definite, show that (2.4.36) is true.

Exercise 2.8.14. Let $N(t)$, $t \geq 0$, be a standard, unit rate, Poisson process.

(i) Show that N is continuous in probability and in mean square, but not with probability one.

(ii) Can you find an example of a stochastic process that is continuous in probability but not in mean square. What about the other way around?

(iii) Show that $M(t) \triangleq \int_0^t N(s)\, ds$ is differentiable in mean square, but not with probability one.

Exercise 2.8.15. Let f be any stochastic process on $\mathbb{R}^N$.

(i) Show that f is mean square continuous if and only if its covariance function C is continuous on $T \times T$.

(ii) Show that if C is continuous at diagonal points (t, t), then it is continuous everywhere.

Exercise 2.8.16. Suppose f is a continuous Gaussian process on $[0, 1]$, φ a homeomorphism of $[0, 1]$, and g a new process defined, also on $[0, 1]$, by $g_t = f(\varphi(t))$. Show that f and g have identical entropy functions. Show, by example, that it is possible for the covariance function of f to satisfy (2.5.12) while that of g does not.

Hint: Use Brownian motion as your process.

Exercise 2.8.17. Prove the equivalence of the finiteness of the two integrals in (2.5.12). and that the convergence of both is assured by (2.5.13).

Exercise 2.8.18. Given a stationary Gaussian field f on $\mathbb{R}^N$ with spectral measure ν,

(i) Show that the finiteness of the spectral integral (2.5.19) ensures that f is sample path continuous.

(ii) Show that f will be k times continuously differentiable if

$$\int_{\mathbb{R}^N} |\lambda|^{2k} \big(\log(1 + |\lambda|) \big)^{1+\alpha} \nu(d\lambda) < \infty,$$

for some $\alpha > 0$.

Exercise 2.8.19. Find a counterexample to "Slepian's inequality for absolute values" which, if it were correct, would claim that the inequality holds with absolute values.

Hint: It suffices to find a counterexample based on a parameter space with only two points.

Exercise 2.8.20. Suppose that f is a stationary, zero mean, Gaussian random field on $\mathbb{R}^N$, with a covariance function $C(t)$ that can be expanded as

$$C(t) = C(0) + \tfrac{1}{2} t \Lambda t' + o(|t|^2),$$

for some matrix Λ of second spectral moments, and for $|t|$ in some neighborhood of the origin.

(i) Using Slepian's inequality, show that there is (perhaps another) neigh-
     borhood of the origin throughout which the exceedence probabilities of f
     can be bounded, above and below, by those of a cosine process.
(ii) Identify the parameters of the cosine process.

Exercise 2.8.21 (Rice formula). Let f be a C^1 process defined on the real
line. Let $N_u(T)$ denote the number of upcrossings by f of the level u in $[0, T]$,
viz.

$$N_u(T) = \left\{ \# \left\{ t \in [0, T] : f(t) = u, \ \dot{f}(t) > 0 \right\} \right\}$$

(i) Use Theorem 2.7.1 to derive Rice's original formula:

$$\mathbb{E}\left\{N_u(T)\right\} = \int_0^T \mathbb{E}\left\{ \dot{f}(t) \, \mathbb{1}_{\{\dot{f}(t)>0\}} \big| f(t) = u \right\} p_t(u) \, dt,$$

where $p_t(u)$ is the density of $f(t)$.
(ii) Assume that f is stationary and Gaussian with zero mean and unit vari-
     ance and show that

$$\mathbb{E}\left\{N_u(T)\right\} = T \frac{\lambda_2^{1/2}}{2\pi} e^{-u^2/2}, \tag{2.8.4}$$

where λ_2 is the second spectral moment

$$\lambda_2 = \mathbb{E}\left\{ \dot{f}(t)^2 \right\} = \int_{\mathbb{R}} \lambda^2 \, \nu(d\lambda)$$

and ν is the spectral measure (2.4.2).
(iii) Assume that f is Gaussian with zero mean and constant unit variance.
      Show that

$$\mathbb{E}\left\{N_u(T)\right\} = \frac{e^{-u^2/2}}{2\pi} \int_0^T \lambda_t^{1/2}(t) \, dt,$$

where $\lambda_t = \mathbb{E}\left\{ \dot{f}(t)^2 \right\}$.

Exercise 2.8.22. Let f be the cosine process of (2.3.3). Show, from first
principles, (i.e. without using the results of Section 2.7) that, for $T \leq \pi/\lambda$,
the Rice formula (2.8.4) holds for this process.

Exercise 2.8.23. Let f be a C^2 process defined on the real line, and let
$M_u(T)$ be the number of local maxima of f above the level u in $[0, T]$, viz.

$$M_u(T) = \# \left\{ t \in [0, T] : \dot{f}(t) = 0, \ \ddot{f}(t) < 0, \ f(t) \geq u \right\}.$$

Also, let $M(T) = M_{-\infty}(T)$ be the total number of local maxima in $[0, T]$.

(i) Apply Rice's formula to show that if f is stationary and Gaussian with mean 0 and variance 1 then

$$\mathbb{E}\left\{M(T)\right\} = T\frac{\lambda_4^{1/2}}{2\pi\lambda_2^{1/2}},$$

where λ_4 is the fourth order spectral moment

$$\lambda_4 = \mathbb{E}\left\{\ddot{f}(t)^4\right\} = \int_{\mathbb{R}} \lambda^4\, \nu(d\lambda),$$

and ν is the spectral measure (2.4.2).

(ii) For general f, use Theorem 2.7.1 to show that

$$\mathbb{E}\left\{M_u(T)\right\} = \int_0^T \mathbb{E}\left\{-\ddot{f}(t)\,\mathbb{1}_{\{\ddot{f}(t)<0\}}\,\mathbb{1}_{\{f(t)>u\}}\,\Big|\,\dot{f}(t)=0\right\}\dot{p}_t(0)\,dt,$$

where $\dot{p}_t$ is the density of $\dot{f}(t)$.

(iii) Again assuming that f is stationary and Gaussian with mean 0 and variance 1, and using the above, show that

$$\mathbb{E}\left\{M_u(T)\right\} = T\frac{\lambda_4^{1/2}}{2\pi\lambda_2^{1/2}}\Psi\left(\frac{\lambda_4^{1/2}u}{\Delta^{1/2}}\right) - T\frac{\lambda_2^{1/2}}{\sqrt{2\pi}}\varphi(u)\Phi\left(\frac{\lambda_2 u}{\Delta^{1/2}}\right),$$

where $\Delta = \lambda_4 - \lambda_2^2$.

(iv) Using the above and Rice's formula show that, for C^2, stationary, zero mean and unit variance Gaussian processes on the real line,

$$\lim_{u\to\infty} \frac{\mathbb{E}\left\{M_u(T)\right\}}{\mathbb{E}\left\{N_u(T)\right\}} = 1.$$

Exercise 2.8.24. Suppose that $T = [0,1]^N$. Show that under conditions (a) and (b) of Theorem 2.7.1 there are no points $t \in \partial T$ satisfying $f(t) = u$.