I. INTRODUCTION

1. The Basic Ideas.

It is my aim in these notes to treat two basic and closely related problems in the theory of Gaussian processes: the question of sample path continuity and the distribution of the supremum of a Gaussian process over a fixed set in its parameter space.

The basic approach will follow the modern attitude that the precise geometrical structure of the parameter space of a Gaussian process is of relatively little importance in determining sample path properties. Of more importance is establishing the "size" of the parameter space when measured in terms of a metric based on the covariance function of the process. There are various ways to measure this size, the best known, and easiest to handle, being via the notion of *metric entropy*. The most recent, and most powerful tool however is the notion of *mesure majorante*, or *majorising measure*, introduced by Preston (1972), (albeit not under this name), developed by Fernique in a series of papers starting with Fernique (1974), and recently shown by Talagrand (1987) to be the only tool currently available that provides necessary, as well as sufficient, conditions for the a.s. continuity of a general Gaussian process.

In between regular metric entropy and majorising measures lie some other notions, retaining the ease of application of metric entropy while giving stronger results. We shall consider two of these as well, once in studying continuity, and once while looking at suprema distributions.

Before we can start anything, however, we need to settle on terminology and notation.

$(\Omega, \mathcal{F}, P)$ will be a complete probability space that will remain fixed throughout the notes. A Gaussian random variable X with mean $\mu \in \Re$ and variance $\sigma^2 \in \Re_+$ is a real valued random variable such that for each $\lambda \in \Re$

$$Ee^{i\lambda X} = e^{i\mu\lambda - \frac{1}{2}\sigma^2\lambda^2},$$

or, equivalently, the law of X has density $\sigma^{-1}\phi((x-\mu)/\sigma)$, where

$$\phi(x) := (2\pi)^{-\frac{1}{2}} \exp(-x^2/2).$$

If $\mu = 0$ we call X *centered*, and if we also have $\sigma = 1$, then X is called *standard normal*. Since the transition from centered to non-centered Gaussians is via the easy addition of a constant we shall treat, almost exclusively, only centered Gaussian variables.

A (centered) Gaussian process is a family $\{X_t\}_{t\in T}$ of random variables, indexed by a parameter set T, such that each linear combination $\sum \alpha_t X_t$ is (centered) Gaussian. Whereas the use of the letters t and T seem to

denote that the parameter is "time", the entire thrust of the modern theory of Gaussian processes is that T should be very general and neither theorems nor proofs should assume any particular structure for T. (In general we shall require only that T be a metric space, and that it be totally bounded – to be defined in a moment – in its metric.) For example, we shall consider examples in which T is a subset of $\Re^k$, $k > 1$ ("multiparameter time"), or the set of all squares within $[0,1]^k$, $k > 1$, or all convex sets in $[0,1]^k$, $k > 1$, etc. In none of these examples is T either one-dimensional, or ordered.

If X is a centered Gaussian process, then the above definition implies that the *covariance function* $R(s,t) = E(X_s X_t)$ on $T \times T$, which determines $E(\sum \alpha_t X_t)^2$, also determines the law of the entire process. It is easy to see that R must be positive semi-definite (i.e. $\sum_{s,t \in S} \alpha_s R(s,t) \alpha_t \geq 0$ for all $S \subset T$ and $\alpha_t \in \Re$). It is a basic result on Gaussian processes, following from the Kolmogorov extension theorem and the form of the multivariate Gaussian distribution, that given any positive semi-definite R on $T \times T$ there exists a centered Gaussian process on T with covariance function R.

Without further mention we shall always assume that T has a countable dense subset (which we shall often explicitly exhibit as part of an entropy calculation), and shall assume that X is a *separable* stochastic process. The importance of this assumption is that questions of continuity and boundedness of X over T then reduce to equivalent questions over the dense subset, and the fact that this is countable often makes for somewhat easier arguments.

This is all we need in the way of formalities. ■

By way of an introduction to what awaits us, consider the two problems of determining the a.s. continuity of X on T, and finding the distribution of $\sup_{t \in T} X_t$. Assume T is a metric space with metric τ, and consider

$$\rho^2(u) = \rho_\tau^2(u) = \sup_{\tau(s,t) \leq u} E(X_s - X_t)^2. \tag{1.1}$$

If X is to be continuous, then it is obvious that we must have $\rho_\tau(u) \to 0$ as $u \to 0$. In fact, this is a trivial necessary and sufficient condition for mean square continuity; c.f. Exercise 1.1. To obtain full, almost sure, sample function continuity we shall require more, and, in particular, we shall have to concentrate on the rate at which $\rho(u)$ tends to zero. The function ρ cannot, however, be of intrinsic importance, for we could change to an equivalent metric on T, without affecting the continuity or otherwise of X, but changing the specific behaviour of ρ.

Once we realise this fact, it is obvious that there is no good reason not to choose a metric on T that is equivalent to the basic metric, but easier to work with. An appropriate candidate is given via the covariance function of X as

$$d(s,t) = [E(X_s - X_t)^2]^{\frac{1}{2}} \tag{1.2}$$

in a notation that will henceforth remain fixed. Actually, d is only a pseudo-metric, since $d(s,t) = 0$ does not necessarily imply that $s = t$. Nevertheless, we shall call d the *canonical metric* for T and/or X. Since the question of a.s. continuity makes little sense if we are not prepared to accept that X is at least mean square continuous, which, as we have just noted, occurs if and only if the covariance function is continuous over $T \times T$, we shall henceforth assume R continuous. In this case it is easy to see that if (T, τ) is compact, then τ-continuity and d-continuity are equivalent. Since this fact is crucial to our approach, we give a simple proof.

Firstly, note that since R is continuous, if X is d-continuous on T then it is trivially τ-continuous. The other direction is not immediate. The problem comes from points that are distant in the τ-metric, but close in the d-metric. (For example, if we take perhaps the simplest of all non-trivial Gaussian processes,

$$X(t) \;=\; A\cos(t-\phi), \qquad\qquad t \in [0,T],$$

with A a Rayleigh random variable with probability density $a\exp(-a^2/2)$, $a \geq 0$, and independent of ϕ uniform on $[0, 2\pi)$, then since X is periodic $d(s,t) = 0$ whenever $\tau(s,t) := |t-s|$ is a multiple of 2π. Hence τ-continuity of X does not immediately imply d-continuity.)

Suppose, however, that X *is* τ-continuous. For $\eta \geq 0$, let

$$A_\eta \;=\; \{(s,t) \in T \times T\colon d(s,t) \leq \eta\}.$$

Since R is continuous, this is a τ-closed subset of $T \times T$. Furthermore, $\bigcap_{\eta > 0} A_\eta = A_0$. Fix $\epsilon > 0$. Then, by the τ-compactness of T, there is a finite set $B \subset A_0$ (the number of whose elements will in general depend on ϵ) such that

$$\bigcup_{(s',t') \in B} \{(s,t) \in T \times T\colon \max\left(\tau(s,s'), \tau(t,t')\right) \leq \epsilon\}$$

covers A_η for some $\eta = \eta(\epsilon) > 0$. That is, whenever $(s,t) \in A_\eta$ there is a $(s',t') \in B$ with $\tau(s,s'), \tau(t,t') \leq \epsilon$. Note that

$$|X_t - X_s| \;\leq\; |X_s - X_{s'}| + |X_{s'} - X_{t'}| + |X_{t'} - X_t|.$$

Since $(s',t') \in B \subset A_0$, we have $X(s') = X(t')$ a.s. Thus

$$\sup_{d(s,t) \leq \eta(\epsilon)} |X_t - X_s| \;\leq\; 2 \sup_{\tau(s,t) \leq \eta} |X_t - X_s|,$$

so that the τ-continuity of X implies its d-continuity. ■

Thus, at least for a while, we can concentrate on questions of continuity with respect to the canonical metric only, safe in the knowledge that if T is compact this is not a restriction.

It thus seems reasonable to now assume that continuity problems are going to be associated with the incremental covariance function $\rho_d(u)$ of (1.1). Life, however, is not quite so simple. Assume for the moment that T is continuous, in the sense that for all $u \leq \text{diam}(T)$, where

$$\text{diam}(T) = D(T) = \sup_{s,t \in T} d(s,t),$$

we have that there exists at least one pair of points $s,t \in T$ such that $d(s,t) = u$. Then it is immediate from the definition (1.1) that

$$\rho_d(u) = \sup_{d(s,t) \leq u} [E(X_s - X_t)^2]^{\frac{1}{2}} \equiv u.$$

Since this is independent of the covariance function it cannot be the correct measure with which to study continuity. The information which used to be in ρ has now somehow moved into the relationship between the canonical metric d and the parameter space T, and, in fact, is tied up in the "size" of T as measured in terms of d.

We denote by $N(\epsilon)$ the smallest number of closed d-balls of radius ϵ that cover T, and set $H(\epsilon) = \log N(\epsilon)$. The latter is called the metric entropy function for T (or X). We shall refer to any condition or result based on N or H as an entropy condition/result.

THROUGHOUT THESE NOTES WE SHALL ASSUME THAT THE FUNCTION $N(\epsilon)$ IS FINITE FOR ALL $\epsilon > 0$. THAT IS, WE ASSUME THAT T IS TOTALLY BOUNDED IN THE CANONICAL METRIC.

This is a basic assumption, without which many of the most elementary results following will not be true. However, it is so weak an assumption, that it never need worry us from the point of view of restricting applications of the general theory.

The first "general theory" result, is the following:

1.1 THEOREM. *A sufficient condition for the a.s. continuity of X on T is*

$$\int_0^\infty \left(\log N(\epsilon)\right)^{\frac{1}{2}} d\epsilon < \infty. \tag{1.3}$$

(This theorem has a long history, with the first published versions due to Dudley (1967) and Sudakov (1969). Dudley (1973) has a full historical account of the various early versions with their various proofs, both correct and "in doubt".)

Much more is actually true within the framework of Theorem 1.1 than we have stated here. For example, the function $\omega(\delta) = \int_0^\delta (\log N_\epsilon)^{\frac{1}{2}} d\epsilon$, serves,

up to a random multiplicative constant, as a modulus of continuity for X in terms of the canonical metric d. Consequently, investigation of the integral in (1.3) involves the study of a number of properties of X at once.

We shall see how to prove Theorem 1.1, along with a variety of related results, in Chapter 4. Note now, however, that when $\epsilon > \text{diam}(T)$ we have $N_\epsilon = 1$, so that the upper bound on the integral in (1.3) is really $\text{diam}(T)$. Furthermore, since $N(\epsilon)$ is clearly a decreasing function of ϵ, the issue of finiteness in (1.3) is at zero.

We shall motivate the need for an entropy type condition with a variety of examples in the following two sections.

Now, however, let us turn briefly to the question of the behaviour of $\sup_T X_t$. To be honest, we should point out that in general finding the distribution of this supremum is an almost impossible problem. For example, precise formulae for

$$P(\lambda) = P\{\sup_{t\in T} X_t \geq \lambda\} \tag{1.4}$$

are known for only six cases (i.e. six covariance functions) if X is required to be stationary, and in each one of these T is a finite interval in $\Re^1$. The best one can hope to get here are reasonable formulae for the asymptotic (as $\lambda \to \infty$) behaviour of $P(\lambda)$, and here there is a reasonably full theory.

It is easy to see (i.e. to guess) that the $\lambda \to \infty$ behaviour of $P(\lambda)$ will depend on two factors. The first, and most obvious, is any lack of homogeneity of X on T. If EX_t^2 is not constant over T, then it is reasonable to expect that the supremum will be achieved somewhere in the neighbourhood of the point (or points) of maximal variance. This is in fact the case.

The second factor will be the local smoothness, or lack thereof, of X. Again, it is reasonable to expect that the rougher X is near a point of maximal variance, the larger one can expect the supremum to be. Again this idea can be made rigorous and precise estimates and results proven. We shall return to this in detail in Chapter 5, after we have proven the rather surprising result that if $t_o \in T$ is such that

$$EX_{t_o}^2 = \sup_{t\in T} EX_t^2$$

then, at least for large λ, $P(\lambda)$ is not very different from $P\{X_{t_o} > \lambda\}$. Borell's inequality, Chapter 2, is the formal theorem, and in many ways is one of the building blocks of these notes. Some examples in Chapter 5 will show us that λ need not really be all that large for this result to be true. In fact, $\lambda \approx 1.5 \sup_{t\in T} (EX_t^2)^{1/2}$ is often large enough!

Now, however, the time has come to look at some examples. We start with a look at the continuity question for a specific family of centered Gaussian processes, with the aim of convincing you that you should, once and

for all, forget that in the prehistory of Gaussian processes "t" was indeed time, and to show that it is neither T that is important, nor X, but rather the relationship between them as measured by the canonical metric d. Not surprisingly, the family that we shall take for our first class of examples is based on what is perhaps the most important of all stochastic processes, Gaussian or not – the Brownian motion. The family is composed of some of the extensions of this process to parameter spaces richer and more complex than the real line. This family of examples forms the content of the following section, and is required reading, in terms of providing motivation for the general theory.

2. The Brownian Family of Processes.

Let $(E, \mathcal{E}, \nu)$ be a σ-finite measure space. (Usually it will be a Euclidean space with the Borel σ-algebra and Lebesgue measure). A *Gaussian white noise* based on ν is a random set function W on the sets $A \in \mathcal{E}$ of finite ν-measure such that

(1.5) $W(A)$ is centered Gaussian and $EW^2(A) = \nu(A)$.

(1.6) If $A \cap B = \phi$ then $W(A \cup B) = W(A) + W(B)$ a.s.

(1.7) If $A \cap B = \phi$ then $W(A)$ and $W(B)$ are independent.

Property (1.6) encourages one to think of W as a random (signed) measure, although it is not generally σ-finite. We describe (1.7) by saying that W has *independent increments.*

To see that such a process exists, think of $\mathcal{E}$ as a parameter space T, with a covariance function on $\mathcal{E} \times \mathcal{E}$ given by

$$R_\nu(A, B) \;=\; EW(A)W(B) \;=\; \nu(A \cap B).$$

Since for any $A_i \in \mathcal{E}$ and $\alpha_i \in \mathfrak{R}$

$$\begin{aligned} \sum_{i,j} \alpha_i R_\nu(A_i, A_j)\alpha_j \;&=\; \sum_{i,j} \alpha_i \alpha_j \int_E I_{A_i}(x) I_{A_j}(x)\, \nu(dx) \\ &= \int_E \Big(\sum_i \alpha_i I_{A_i}(x) \Big)^2 \nu(dx) \\ &\geq 0, \end{aligned}$$

R_ν is positive semi-definite. Thus W exists as a centered Gaussian process on $\mathcal{E}$. ∎

To see what the canonical metric d is in this case, note that

$$\begin{aligned} d^2(A,B) &= E(W(A)-W(B))^2 \\ &= \nu(A)+\nu(B)-2\nu(A\cap B) \\ &= \nu(AB^c \cup A^cB) \\ &= \nu(A\Delta B) \end{aligned}$$

where the symmetric difference $A\Delta B$ is made up of the points in either A or B, but not both.

We specialize to the case $E = \Re_+^k = \{(t_1,...,t_k): t_i \geq 0\}$, $\mathcal{E}$ = the Borel σ-algebra on $\Re_+^k$ and $\nu = \lambda$ = Lebesgue measure. At the risk of occasional confusion we denote a point $(t_1,\ldots,t_k)$ in $\Re_+^k$ simply by t. Let $(a,b] \subset \Re^k$ be the "half open" k-dimensional interval $\Pi_{i=1}^k (a_i,b_i]$, (with open, closed, and half-open from above intervals written and defined similarly), and set $W_t = W((0,t])$. W_t is called the *Brownian sheet.* Via this definition and the above, W is the centered Gaussian process with covariance

$$EW_sW_t = (s_1 \wedge t_1) \times \cdots \times (s_k \wedge t_k).$$

When $k = 1$, W is the standard Brownian motion. When $k > 1$, then if we think of $W(A)$ as a measure, W_t is the corresponding "distribution function". It vanishes on the axes, and if we fix $k-1$ of the indices, it is a scaled Brownian motion in the remaining free variable. (c.f. Exercise 2.1.)

The Brownian sheet holds roughly the same place in the theory of multi-parameter stochastic processes and multivariate statistics that the standard Brownian motion does in one dimension. It is a multi-parameter martingale (Cairoli and Walsh (1975), Wong and Zakai (1976)) and forms the basis of the multiparameter stochastic calculus. There is a nice review of its basic properties in Walsh (1986), where you can also discover its central rôle in the theory of stochastic partial differential equations, and in what sense it is valid to describe the derivative $\partial^k W(t_1,\ldots,t_k)/\partial t_1 \ldots \partial t_k$ as Gaussian white noise. Markovian properties, however, are somewhat missing, due to the fact that the parameter space is not totally ordered. (But see Section 5.5 for one way to get around this, and keep your eyes open for a forthcoming paper by John Walsh and Robert Dalang on this topic.) The lack of order in $\Re_+^k$ will not worry us, however, since from the Gaussian point of view we have already agreed to forgo using the special structure of specific parameter spaces.

The Brownian sheet arises in multivariate statistics in two main settings. If Z_+^k is the k-dimensional integer lattice in $\Re_+^k$ and $\{X_i\}_{i\in Z_+^k}$ a collection of centered, i.i.d. random variables with $EX_i^2 = 1$, then it is not too hard to show that $n^{-k/2}\sum_{i\in n(t)} X_i$ converges weakly to W_t as $n \to \infty$, $t \in [0,1]^k$, where $n(t) = \{i \in Z_+^k : 1 \leq i_j \leq [nt_j],\ j = 1,...,k\}$. ($[x]$ is the "integer

part" of x.) That is, W_t is basic to the k-dimensional functional central limit theorem.

Secondly, let $X_1, X_2, \ldots$, be a sequence of i.i.d. random variables, with common uniform distribution on $[0,1]^k$. Then the empirical distribution function $F_n(t)$ of the X_i is defined as

$$F_n(t) = \frac{\#\{i \le n: X_i \in [0,t]\}}{n}. \tag{1.8}$$

Similarly, the empirical measure is defined as

$$\nu_n(A) = \frac{\#\{i \le n: X_i \in A\}}{n}. \tag{1.9}$$

Then, as $n \to \infty$, the normalized empirical distribution function $G_n(t) := \sqrt{n}\{F_n(t) - F(t)\}$ converges weakly to the so-called *pinned Brownian sheet*, $\mathring{W}$, a version of which can be obtained as

$$\mathring{W}_t = W_t - |t| W_1, \qquad t \in [0,1]^k, \tag{1.10}$$

where $|t| = \prod_{i=1}^k t_i$. If we restrict A to a small enough (in terms of entropy) class of sets, then $n^{\frac{1}{2}}\{\nu_n(A) - \lambda(A)\}$ converges weakly to a process we shall call the *set indexed, pinned, Brownian sheet*, a version of which is given by

$$\mathring{W}(A) = W(A) - \lambda(A) W([0,1]^k).$$

(For details about the weak convergence, consult Dudley (1978a) and Pollard (1984). For more properties of the limit processes, be patient.)

Our first result is

1.2 PROPOSITION. *The Brownian sheet and pinned Brownian sheet are continuous on* $[0,1]^k$.

PROOF: Clearly, we need only prove the proposition for W. The continuity of $\mathring{W}$ then follows from that of W and (1.10).

We shall use the entropy condition of Theorem 1.1, for which we must calculate a bound for the minimal number of d-balls of radius ϵ needed to cover $[0,1]^k$.

Fix $t \in [0,1]^k$, and let $S(t,\delta)$ be the k-dimensional cube with lower left corner t and sides of length δ: i.e.

$$S(t,\delta) = \{s \in [0,1]^k : t_i \le s_i \le t_i + \delta,\ i = 1, \ldots, k\}.$$

Then it is easy to see that

$$\sup_{s \in S(t,\delta)} E(W_t - W_s)^2 = k\delta.$$

(Equality holds here if $t+\delta \in [0,1]^k$.) Hence,

$$\sup_{s \in S(t,\delta)} d(s,t) \leq \sqrt{k\delta},$$

and so one way to cover $[0,1]^k$ with d-balls of radius ϵ is with the $[k(1+\epsilon^{-2})]^k$ balls of radius ϵ^2/k centered at the points t of the form $(i_1\epsilon^2/k, \ldots, i_k\epsilon^2/k)$, $i_j = 0, 1, ..., [k\epsilon^{-2}]$. It thus follows that $N(\epsilon) < k^k(1+\epsilon^{-2})^k$, and so the entropy integral $\int_0^\infty (\log N(\epsilon))^{\frac{1}{2}} d\epsilon$ is finite with room to spare. This completes the proof. ■

The natural question to ask now is whether or not continuity is preserved when we look at Gaussian white noise indexed by more general classes of sets than the positive orthants that gave us the Brownian sheet. The motivation behind this question is that the weak convergence referred to above in the context of empirical processes requires that the limit process be continuous, and so the very *raison d'être* of our interest in the Brownian sheet (at least from this viewpoint) hinges on this question. The reason to look at more general index sets lies in the fact that the power of Kolmogorov-Smirnov tests based on statistics of the type

$$\sup_{A \in \mathcal{A}} n^{-\frac{1}{2}} |\nu_n(A) - \lambda(A)| \tag{1.11}$$

is substantially increased by taking richer classes of sets $\mathcal{A}$ other than merely lower orthants, (Pyke (1984, 1985)).

We shall give a number of examples of interesting classes $\mathcal{A}$ in the following section, and later in these notes. Now, however, we shall consider an example of a parameterising class of sets which is too rich to allow continuity.

Introduce a partial order on $\Re^k$ by writing $s < (\leq)\, t$ if $s_i < (\leq)\, t_i$ for all $i = 1, .., k$. Then a set A in $\Re^k$ is called a *lower layer* if $s \leq t$ and $t \in A$ implies $s \in A$. With some looseness of terminology, we shall refer to the Gaussian white noise based on Lebesgue measure on a class of sets $\mathcal{A}$ as the Brownian sheet on $\mathcal{A}$.

1.3 PROPOSITION. *The Brownian sheet on lower layers in* $[0,1]^2$ *is discontinuous and unbounded with probability one.*

PROOF: We start by constructing some examples of lower layers. Write a generic point in $[0,1]^2$ as (s,t) and let $T := T_{01}$ be the right triangle in which $s \leq 1$ and $t \leq 1 \leq s+t$. Let C_{01} be the square where $\frac{1}{2} < s \leq 1$ and $\frac{1}{2} \leq t \leq 1$.

For $n = 1, 2, \ldots,$ and $j = 1, ..., 2^n$, let T_{nj} be the right triangle defined by $s+t \geq 1$, $(j-1)2^{-n} \leq s < j2^{-n}$, and $1-j2^{-n} < t \leq 1-(j-1)2^{-n}$. Let C_{nj} be the square filling the upper right corner of T_{nj}, in which $(2j-1)2^{-(n+1)} \leq s < j2^{-n}$ and $1-(2j-1)2^{-(n+1)} \leq t < 1-(j-1)2^{-n}$.

structures of Figure 1.1, where each step comes from the horizontal and vertical sides of some T_{nj} with, perhaps, different n.

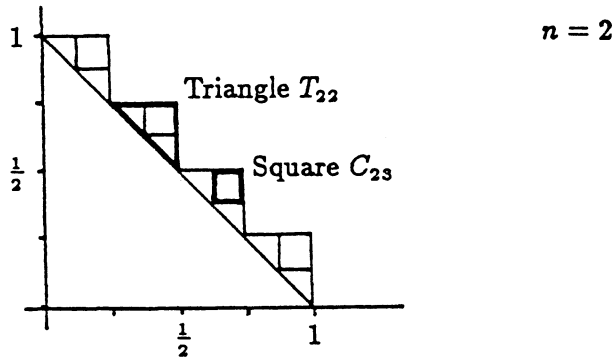

FIGURE 1.1. Construction of some lower layers.

Note that since the squares C_{nj} are disjoint for all n and j, the random variables $W(C_{nj})$ are independent. Also $\lambda(C_{nj}) = 4^{-(n+1)}$ for all n, j.

Let D be the negative diagonal $\{(s,t) \in [0,1]^2 : s + t = 1\}$, and $L_{nj} = D \cap T_{nj}$. For each $n \geq 1$, each point $p = (s,t) \in D$ belongs to exactly one such interval $L_{n,j(n,p)}$ for some unique $j(n,p)$.

For each $p \in D$ and $M < \infty$ the events

$$E_{np} := \{W(C_{n,j(n,p)}) > M2^{-(n+1)}\}$$

are independent for $n = 0, 1, 2, \ldots$, and have the same positive probability. (Since $W(C_{nj})/2^{-(n+1)}$ is standard normal for all n and j.) Thus, such an event occurs with probability one. Let $n(p) := n(p,\omega)$ be the least such n, defined and finite for almost all ω.

Since the events $E_{np}(\omega)$ are measurable jointly in p and ω, Fubini's theorem implies that, with probability one, for almost all $p \in D$ (with respect to Lebesgue measure on D) some E_{np} occurs, and $n(p) < \infty$. Let

$$\begin{aligned} V_\omega &= \bigcup_{p \in D} T_{n(p),j(n(p),p)}, \\ A_\omega &:= \{(s,t) : s + t \leq 1\} \cup V_\omega, \\ B_\omega &:= A_\omega \setminus \bigcup_{p \in D} C_{n(p),j(n(p),p)}. \end{aligned}$$

Then A_ω and B_ω are lower layers. Furthermore, almost all $p \in D$ belong to an interval of length $2^{\frac{1}{2}-n(p)}$ which is the hypothenuse of a triangle with the square $C_p = C_{n(p),j(n(p),p)}$ in its upper right corner, for which $2W(C_p) > M2^{-n(p)}$. Consequently,

$$W(A_\omega) - W(B_\omega) \geq \sum_p M2^{-n(p)}/2,$$

where the sum is over those $p \in D$ corresponding to distinct intervals $L_{n(p),j(n(p),p)}$. Since the union of the countably many such intervals is almost all of the diagonal, the sum of $\sum 2^{-n(p)}$ is precisely 1.

Hence $W(A_\omega) - W(B_\omega) \geq M/2$, implying that $\max\{|W(A_\omega)|, |W(B_\omega)|\} \geq M/4$. Sending $M \to \infty$ we see that W is unbounded and so, *a fortiori*, discontinuous with probability one over lower layers in $[0,1]^2$. ■

The above argument is due to Dudley (1978b), and a similar argument shows that W is unbounded over the convex subsets of $[0,1]^3$. Entropy arguments show that W is also unbounded over the convex subsets of $[0,1]^k$ for all $k \geq 4$, and (just to make sure that you don't confuse sample path properties with topological properties of the parameter space) that W *is* continuous over convex subsets of the unit square.

To see how entropy enters into this example, let us consider one further example, that is closely related to the above in terms of entropy, but is geometrically less appealing than lower layers.

Let γ be positive, and define the collection $\mathcal{A}_\gamma$ of subsets of $\Re^2$ as follows: Given $A_0 = [0,1]^2, A_1, ..., A_{n-1}$, let A_n be the closed rectangle whose left (vertical) side is the right side of A_{n-1}, (so that it has height 1) and with width $2^{n(1-\gamma)}$. Now divide each A_n into 2^n equal horizontal slices $A_{n1}, ..., A_{n2^n}$. The area of each A_{nk} is $2^{-n\gamma}$. The family $\mathcal{A}_\gamma$ is defined as the collection of all finite and countable unions of the sets $\{A_{nk}\}$.

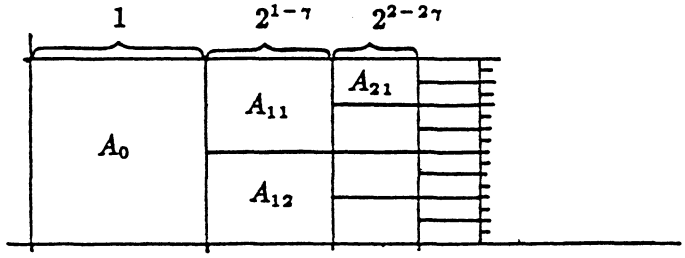

FIGURE 1.2. The partition used to create $\mathcal{A}_\gamma$.

We start by claiming that the entropy function $N(\epsilon)$ of $\mathcal{A}_\gamma$ satisfies, for all $\gamma > 1$

$$a_1 \exp(b_1 \epsilon^{-2/(\gamma-1)}) \leq N(\epsilon) \leq a_2 \exp(b_2 \epsilon^{-2/(\gamma-1)}), \tag{1.12}$$

for suitable constants a_i, b_i. It thus follows from Theorem 1.1 that since

$$\int_0 \left(\log N(\epsilon)\right)^{\frac{1}{2}} d\epsilon \leq K \int_0 \epsilon^{-1/(\gamma-1)}\, d\epsilon,$$

the Brownian sheet will be continuous on $\mathcal{A}_\gamma$ as long as $1/(\gamma-1) < 1$, i.e. if $\gamma > 2$. We shall prove half of (1.12) in a moment.

for suitable constants a_i, b_i. It thus follows from Theorem 1.1 that since

$$\int_0 \left(\log N(\epsilon)\right)^{\frac{1}{2}} d\epsilon \;\leq\; K \int_0 \epsilon^{-1/(\gamma-1)}\, d\epsilon,$$

the Brownian sheet will be continuous on $\mathcal{A}_\gamma$ as long as $1/(\gamma-1) < 1$, i.e. if $\gamma > 2$. We shall prove half of (1.12) in a moment.

On the other hand, since for $1 < \gamma \leq 2$ it follows from the lower bound in (1.12) that the entropy integral diverges, it would seem that in this case the Brownian sheet on $\mathcal{A}_\gamma$ is discontinuous. That this is in fact the case can be shown via an almost identical (but somewhat easier) proof to that used above to show unboundedness on lower layers. The details are left to Exercise 2.2.

The only case left undecided is $\gamma \leq 1$. In this case $\bigcup_{n\geq 0} A_n = [0,\infty) \times [0,1]$, and so the collection of sets in $\mathcal{A}_\gamma$ is unbounded, and thus uninteresting.

We now establish the upper bound in (1.12), which is all we need to prove continuity. (The lower bound requires considerably more work and is derived in Samorodnitsky (1987a, 1990).)

Choose $\epsilon = 2^{-m}$, $m > 1$ and set $S_\gamma = 2^\gamma/(2^\gamma - 2) = \lambda(\bigcup_{n,k} A_{nk})$ (since $\gamma > 1$). Define

$$M \;=\; M(\epsilon) \;:=\; 1 + [(2m + \log_2 S_\gamma)/(\gamma - 1)].$$

Let C_M be the set $C_M = \bigcup_{n=M}^{\infty} \bigcup_{k=1}^{2^n} A_{nk}$. Then

$$E\big(W(C_M)\big)^2 \;=\; \sum_{n=M}^{\infty} 2^{n(1-\gamma)} \leq 2^{-2m}.$$

Thus it follows that the d-distance between C_M and any subset of it composed of unions of A_{nk}, $n \geq M(\epsilon)$, $1 \leq k \leq 2^n$, is less than $2^{-m} = \epsilon$, and so all of these unions lie within a d-ball of radius ϵ of C_M.

We still need to find d-balls that cover all unions of A_{nk}, $n < M$, $1 \leq k \leq 2^n$. There are $1 + 2 + 2^2 + \cdots + 2^{M-1} = 2^M - 1$ such sets A_{nk}, and so $2^{2^M - 1}$ possible unions. Take each one of these as the center of a d-ball. Then

$$N(\epsilon) \;\leq\; 2^{2^M} \leq a_2 \exp(b_2 \epsilon^{-2/(\gamma-1)})$$

for appropriate a_2, b_2. This is what we wanted to prove. ■

To be certain that you understand this example, you should make certain that you understand why making γ larger makes $\mathcal{A}_\gamma$ "smaller". (After all, all $\mathcal{A}_\gamma$'s can be trivially mapped on to one another.)

The aim of the examples of this section was to convince you that the relationship between a Gaussian process and its parameter space is, as far

as continuity and boundedness are concerned, an important subject. Note that all the different types of Brownian process that we saw here, whether they were simple Brownian motions defined on $\Re_+$ or complicated set indexed processes defined on a class of sets $\mathcal{A}$, could be handled by a common analysis, based on the covariance function and the canonical metric d. When the parameter space was more complex, so was the analysis; but *in principle* it remained the same. This would not have been so had we attempted to analyse these processes via extensions of the simple Markovian properties exhibited by Brownian motion on $\Re_+$ – these do not extend in any simple fashion to more general parameter spaces.

We shall now look at a number of other examples, some of which will seem at the outset to be qualitatively beyond the range of processes we have set out to consider. That they can be handled within the framework of real valued Gaussian processes on a totally bounded metric space is part of the justification of the general approach that we take. At this stage you should definitely work through the example of processes on $\Re^1$, in order to have a concrete example on which to hang the theory of the following chapters, and that on generalised random fields, in order to have some idea why there is a need to build a general theory that is independent of the structure of the parameter space. Beyond that, the choice of how far to continue is up to you. Some readers criticized early versions of these notes, because there were not enough examples to enable one to appreciate how the general theory could be applied. Other readers criticized later versions, because there were too many examples, and working through them they lost sight of the forest because of the trees. The choice as to how much you should read now is therefore up to you.

3. A Collection of Examples.

1. PROCESSES ON $\Re^1$: Returning from the generalisations of the previous section and the abstraction of the first to "simple" processes on the real line, it is natural to expect that conditions for continuity and boundedness will become simple to both state and prove, and there will be little, if any, need to talk about concepts such as entropy.

This expectation is both true and false. 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 following important result is easy to state without specifically referring to any abstract notions. To state it, let X_t be a centered Gaussian process on a finite interval $[0,T]$, and define

$$p^2(u) = \sup_{|s-t|\le u} E|X_s - X_t|^2. \tag{1.13}$$

i.e. p is identical with the function ρ of (1.1), except that we are now using the usual Euclidean metric for the real line.

1.4 THEOREM. *If, for any* $\delta > 0$,

$$\int_0^\delta (-\log u)^{\frac{1}{2}}\, dp(u) < \infty, \tag{1.14}$$

then X *is continuous on* $[0,T]$ *with probability one. Furthermore, a sufficient condition for (1.14) to hold is for some* $0 < C < \infty$ *and* $\alpha, \eta > 0$

$$E|X_s - X_t|^2 \leq \frac{C}{|\log|s-t||^{1+\alpha}}, \tag{1.15}$$

for all s,t *with* $|s-t| < \eta$.

PROOF: The proof that (1.15) implies (1.14) (along with a similar, and somewhat easier condition to work with) is left to the exercises. Here we shall prove that the fact that (1.14) implies continuity is a trivial consequence of the entropy condition (1.3).

Note first that since $p(u)$ is obviously non-decreasing in u, the Riemann-Stieljes integral (1.14) is well defined, as is the inverse function $p^{-1}(u) := \sup\{t\colon p(t) \leq u\}$.

Note also the obvious fact that, for each $\epsilon > 0$, the interval $[0,T]$ can be covered by $1 + [(2p^{-1}(\epsilon))^{-1}T]$ intervals, each of which has radius ϵ in the canonical metric d, (c.f. (1.2) and (1.13)). Thus,

$$\begin{aligned}
\int_0^\infty \big(\log N(\epsilon)\big)^{\frac{1}{2}}\, d\epsilon &= \int_0^{p(T/2)} \big(\log N(\epsilon)\big)^{\frac{1}{2}}\, d\epsilon \\
&\leq \int_0^{p(T/2)} \Big(\log\big(1 + [T/(2p^{-1}(\epsilon))]\big)\Big)^{\frac{1}{2}}\, d\epsilon \\
&= \int_0^{T/2} \Big(\log\big(1 + T/(2u)\big)\Big)^{\frac{1}{2}}\, dp(u).
\end{aligned} \tag{1.16}$$

The question of the divergence of the last integral hinges on its behaviour in the neighbourhood of zero, and it is clear that (1.14) implies its convergence. Thus, by Theorem 1.1, we have the a.s. continuity of X. ∎

Whereas the integral condition (1.14) is sharper than the inequality (1.15), it is the latter, as one would expect, that is more commonly used. (In Chapter 4 we shall investigate the sharpness of (1.14) itself.) It is also condition (1.15) that captures, more clearly, the notion that the continuity of the process is a consequence of the smoothness of the covariance function at the origin.

It is important to note, however, that in the move from an explicit entropy condition like (1.3) to an implicit one of the form (1.14), we may have lost something. For stationary Gaussian processes on $\mathfrak{R}^1$, this is not the case (as we shall see later). In general however, if X_t is continuous (even stationary) on $T \subset \mathfrak{R}^1$ and satisfies (1.14), then it can be transformed by a homeomorphism f of T into a process $X_{f(t)}$ which may no longer satisfy (1.14) but will, of course, still be continuous. Such a transformation may not preserve stationarity, and so does not contradict the claim of the previous sentence. Furthermore, (as you should check for yourself – Exercise 3.1) a transformation of this kind has no effect on the entropy of X, and so its continuity can still be checked via entropy methods.

With the continuity question more or less settled, a natural question to ask at this stage is how to construct discontinuous processes on $\mathfrak{R}^1$. To do this, let us impose the restriction that X be stationary. In this case the function $p(u)$ simplifies to give

$$p^2(u) = \sup_{0 \le t \le u} E|X_t - X_0|^2 = 2 \sup_{0 \le t \le u} [R(0) - R(t)], \tag{1.17}$$

and the finiteness of the integral in (1.14) becomes a necessary, as well as sufficient, condition for sample path continuity. (See Chapter 4 for details.)

It is a straightforward exercise, given the results of Chapter 4 (see Exercise 3.3), to show that if

$$\frac{C_1}{|\log t|^{1+\alpha_1}} \le R(0) - R(t) \le \frac{C_2}{|\log t|^{1+\alpha_2}}, \tag{1.18}$$

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

For the sake of the reader who is familiar with the spectral theory of stationary processes, the continuity/discontinuity dichotomy can be most naturally stated in spectral terms.

Thus, let $F(\lambda)$ denote the spectral distribution function defined by

$$R(t) = \int_0^\infty \cos t\lambda \, dF(\lambda), \tag{1.19}$$

with spectral density $f(\lambda) = F'(\lambda)$, when F is absolutely continuous.

Via a reasonably standard Tauberian theorem (e.g. Cramér and Leadbetter (1967) pp. 174-177) it follows from (1.18) that if the integrals

$$\int_0^\infty \big(\log(1+\lambda)\big)^{1+\alpha} \, dF(\lambda) = \int_0^\infty \big(\log(1+\lambda)\big)^{1+\alpha} f(\lambda) \, d\lambda \tag{1.20}$$

converge for some $\alpha > 0$ then X is continuous, while if they diverge for some $\alpha < 0$ then X has discontinuous sample paths with probability one.

Thus, if you are familiar with the spectral approach, it is now obvious from (1.20) exactly how much "high frequency oscillation" is allowed before a stationary Gaussian process on the line becomes discontinuous. If you are not familiar with spectra, there is no need to worry. They will not appear again.

Exactly what happens when a Gaussian process is discontinuous is a topic we shall return to in part below, in the example treating generalised processes, and in both of the following chapters.

2. GAUSSIAN FOURIER SERIES: A problem studied in the 1930's by Paley and Zygmund, and at first seemingly unrelated to Gaussian processes, is the question of the uniform convergence of the random Fourier series

$$\sum_{n=0}^{\infty} a_n Y_n e^{int}, \qquad t \in [0, 2\pi], \tag{1.21}$$

where the a_n are real numbers satisfying $\sum_{n=0}^{\infty} a_n^2 = 1$ and the Y_n are a sequence of independent Rademacher random variables; i.e. $P\{Y_n = +1\} = P\{Y_n = -1\} = \frac{1}{2}$. An equivalent (this requires some work to show) question arises when the Y_n of (1.21) are taken to be standard normal.

Both the Rademacher and Gaussian versions of this problem have a number of interesting consequences in non-random harmonic analysis, and it was for this reason that they were posed. Furthermore, the summands in (1.21) can be treated in much greater generality, with the parameter t ranging over a compact Abelian or non-Abelian group, and the non-random a_n replaced by a product of random and non-random terms. You can read about all of this in depth in the monograph of Marcus and Pisier (1981), but for now we shall treat only the simple sum (1.21), and assume that the Y_n are standard normal.

Apart from being of independent mathematical interest, (1.21) is of course interesting as a way to build a stationary Gaussian process on $[0, 2\pi]$. Looked at this way, it is not surprising that the results related to Gaussian processes on the line have something to say about Paley and Zygmund's problem.

Consider the following related version of the original problem, which is a little easier to handle in the context of the *real*-valued Gaussian processes that we are dealing with.

1.5 THEOREM. *Let $\{Y_n\}_{n\geq 1}$ and $\{Y_n'\}_{n\geq 1}$ be two independent, infinite sequences of independent, standard normal random variables, and $\{a_n\}_{n\geq 1}$ a non-increasing real sequence. Then the sum*

$$X_t := \sum_{n=0}^{\infty} a_n (Y_n \cos nt + Y_n' \sin nt), \qquad t \in [0, 2\pi], \tag{1.22}$$

converges uniformly on $[0, 2\pi]$ *if, and only if, the following sum also converges:*

$$\sum_{j=2}^{\infty} \frac{\left(\sum_{n=j}^{\infty} a_n^2\right)^{1/2}}{j(\log j)^{1/2}} \,. \tag{1.23}$$

(An identical result, for the sum (1.21), can also be shown to be true. Such a result is actually more general than our formulation.)

PROOF: Actually, we shall only give a partial proof, as we shall only sketch why (1.23) is sufficient for uniform convergence. After you have read Chapter 3 – in particular Theorem 3.8 – you will be able to complete this part of the proof for yourself. (Exercise 2.5 of Chapter 3.) Despite the long wait, it is interesting to see, already at this stage, what is needed. The necessity is similar, but a little harder. You can find it in Marcus and Pisier (1981) or whichever of the references there to earlier, less abstract, results that you find most convenient.

Since we are assuming the convergence of (1.23), we have that the sequence $\{a_n\}_{n\geq 1}$ is square summable. Thus, by the three series theorem, the sum (1.22) converges for each *fixed* $t \in [0, 2\pi]$, with probability one. Let it do so on a countable dense subset of $[0, 2\pi]$, which, by separability, defines a limit process. The limit is clearly Gaussian, since the Y_n and Y_n' are both Gaussian. What we now have to show is that this convergence is uniform on the parameter space $[0, 2\pi]$. This would ordinarily require showing the existence of a common uniform modulus of continuity for the partial sums defining X_t. The results of Section 3.2, however, are rich in zero-one and related laws that tell us that all we need check is that the limit process, which we have just defined, is continuous. Uniform convergence then comes for free.

To prove the requisite continuity, we start by noting that an easy calculation shows the covariance function of the process X_t of the theorem to be

$$R(s,t) = R(t-s) = \sum_{n=0}^{\infty} a_n^2 \cos\big(n(t-s)\big).$$

The function p of (1.13) is thus given by

$$p^2(u) = 2 \sup_{0\leq t\leq u} \sum_{n=0}^{\infty} a_n^2 (1 - \cos nt) = 4 \sup_{0\leq t\leq u} \sum_{n=0}^{\infty} a_n^2 \sin^2\left(\frac{nt}{2}\right). \tag{1.24}$$

We now have to show that, with this formulation of p, the entropy integral (1.14) converges when the sum (1.23) does. Using Exercise 3.2, it is easy to see that the entropy integral converges if

$$\sum_{n=0}^{\infty} \frac{p(2^{-n})}{\sqrt{n}} < \infty. \tag{1.25}$$

Since $\sin^2 x < \max(1, x^2)$ for all x, we have from (1.24) that

$$p^2(2^{-n}) \leq 4\sum_{j=0}^{n} \frac{2^{2j}}{2^{2n}} A(2^j, 2^{j+1}) + B(2^n),$$

where

$$A(n,m) = \sum_{j=n+1}^{m} a_j^2, \qquad B(n) = A(n,\infty) = \sum_{j=n+1}^{\infty} a_j^2.$$

Using $(|x| + |y|)^{1/2} \leq |x|^{1/2} + |y|^{1/2}$ we see that the sum in (1.25) will converge if both

$$\sum_{n=0}^{\infty} \left(\frac{2^{-2n}}{n} \sum_{j=0}^{n} 2^{2j} A(2^j, 2^{j+1}) \right)^{\frac{1}{2}} < \infty, \tag{1.26}$$

and

$$\sum_{n=0}^{\infty} \left(\frac{B(2^n)}{n} \right)^{\frac{1}{2}} < \infty. \tag{1.27}$$

That (1.27) follows from the convergence of (1.23) is easy to check. It is also true that (1.26) follows from (1.27), but here the proof relies on an old inequality for series that takes a lot of time to prove and is of little independent interest as far as Gaussian processes are concerned. Hence we leave this out, and refer the reader to Marcus and Pisier (1981), once again, for full details. ■

Although we have perhaps left more unproven than proven, we now leave this example and turn to some others to continue building motivation for the general theory to come.

3. THE TALAGRAND EXPANSION: Whereas we initially posed the question of the convergence of the random Fourier series of the previous example as a natural problem in harmonic analysis, from the point of view of a probabilist it is far more natural to view (1.22), for example, as an orthogonal expansion of a Gaussian process on a finite interval. From this viewpoint, a natural question to ask is whether or not orthogonal expansions exist in general (i.e. when the parameter space T has no special structure). When X_t is continuous over T the answer to this question is in the affirmative, as we shall see in §3.2. Now, however, we consider a somewhat different expansion, due to Talagrand (1987).

Let $\{Y_n\}_{n \geq 0}$ be a sequence of (possibly dependent) centered Gaussian variables, with variances $\sigma^2(Y_n)$ satisfying

$$\lim_{n \to \infty} (\log n)^{\frac{1}{2}} \sigma(Y_n) = 0. \tag{1.28}$$

Thus, like the i.i.d. Y_n and Y_n' of (1.22), our new variables become smaller, stochastically, as $n \to \infty$. (To be precise, in the previous examples the size of the random summands was governed indirectly through the non-random coefficients a_n.)

Let T be a compact metric space, and $\{\alpha_n\}_{n\geq 0}$ a summable sequence of positive functions on T such that

$$\alpha_n(t) \geq 0, \qquad \sum_{n=0}^{\infty} \alpha_n(t) \leq 1, \tag{1.29}$$

for all $t \in T$. Define the centered Gaussian process

$$X_t = \sum_{n=0}^{\infty} \alpha_n(t) Y_n. \tag{1.30}$$

It is not too hard to show (Exercise 3.4) that, under the conditions we have specified, X_t is sample path continuous. This is half of the following theorem. What is rather surprising, however, is the converse, which is also much harder to prove.

1.6 THEOREM. *Let X be a centered Gaussian process on a compact metric space T. Then X is continuous if, and only if, it has a continuous covariance function, there exists a Gaussian sequence $\{Y_n\}_{n\geq 0}$ satisfying (1.28), and, for each $t \in T$, X_t can be written in the form (1.30), where the series converges in $\mathcal{L}^2(P)$ and the α_n satisfy (1.29).*

The practical power of this result is that it gives us a very simple, and very general, method for constructing continuous Gaussian processes, particularly in situations where it might be difficult to write down an orthonormal basis for functions on T. Furthermore, it tells us that *all* continuous processes can be built this way.

Two aspects of the theorem are of particular interest. Firstly, it is important to note that the Y_n need not be independent. This somewhat limits the usefulness of the converse part of the result. Secondly, nowhere have we made the (unnecessary) requirement that the α_n be continuous. For more details, including how to prove the easy half of Theorem 1.6, see Exercise 3.4.

4. GENERALISED RANDOM FIELDS:

The term "random field" is generally used to describe a stochastic process whose parameter space is either a k-dimensional Euclidean space or a k-dimensional lattice. Thus, a 1-dimensional random field is what is generally referred to as a "stochastic process". Problems related to the geometrical structure of the sample paths of random fields ($k \geq 1$) are often qualitatively

different, and substantially more difficult, than the corresponding problems when $k=1$: c.f. Adler (1981). Problems related to continuity, however, are generally dimension independent. For example, Theorem 1.4 remains true if we merely replace the domain of X by a compact subset of $\Re^k$, and the definition (1.13) of the function $p(u)$ by

$$p^2(u) = \sup_{\|s-t\| \le \sqrt{k}u} E|X_s - X_t|^2, \tag{1.31}$$

where $\| \ \|$ is the usual Euclidean norm (Exercise 3.5). Similar changes can be made to (1.15) and (1.20): i.e. simply make t or λ vectors and change moduli signs to Euclidean norms where appropriate. Formula (1.17) changes by taking the supremum over $\{t\colon \|t\| \le \sqrt{k}u\}$.

Our interest in this section, therefore, will be not with random fields of the above kind, but rather with a class of Gaussian processes defined on certain function spaces. Nevertheless, we shall use simple random fields to provide an example of the processes that do interest us.

Thus, let X_t be a centered, Gaussian random field on $\Re^k$, with covariance function $R(s,t)$. Let $\mathcal{F}$ be a family of functions on $\Re^k$, and for $\phi \in \mathcal{F}$ define

$$X(\phi) = \int_{\Re^k} \phi(t)\, X(t)\, dt. \tag{1.32}$$

We thus obtain a centered Gaussian process indexed by functions in $\mathcal{F}$, whose covariance function(al) is given by

$$R(\phi,\psi) = EX(\phi)\, X(\psi) = \int_{\Re^k}\int_{\Re^k} \phi(s)R(s,t)\psi(t)\, dsdt. \tag{1.33}$$

The main interest in (1.33) is that it often allows us to define function indexed Gaussian processes with covariance functional of this form, even when a point indexed process X_t with covariance function $R(s,t)$ does not exist. The examples that we shall consider here arise when $R(t,t) \equiv \infty$ for all t, and the divergence near the diagonal is bounded as follows:

$$R(s,t) \le \frac{C}{\|s-t\|^\alpha}, \qquad \text{for all} \quad \|s-t\| \le \delta, \tag{1.34}$$

for some $C < \infty$ and $\delta > 0$.

The main questions of interest here are on what families $\mathcal{F}$ can a process satisfying (1.33) be defined, and when is it continuous.

The answer to the first question is trivial: Given any positive definite function R, one can define a function indexed process on

$$\mathcal{F}_R = \{\phi\colon \int_{\Re^k}\int_{\Re^k} \phi(s)R(s,t)\phi(t)\, dsdt < \infty\}. \tag{1.35}$$

The answer to the second question involves a considerable amount of work, but you should follow the argument carefully, because it is one of the few places in these notes that we shall give an involved entropy calculation in a reasonably complicated example in reasonably complete detail.

As to why we have chosen covariances satisfying (1.34), and how these are related to wider classes of function (and measure) indexed functions, see both the comments at the end of this collection of examples, and the examples in the following subsection.

Let $T \subset \mathbb{R}^k$ be bounded, $q > 0$, and $p = [q]$. Let $C_0, \ldots, C_p$ and C_q be finite, positive constants, and let $\mathcal{F}^{(q)} = \mathcal{F}^{(q)}(T, C_0, \ldots, C_p, C_q)$ be the class of functions on T whose partial derivatives of orders $1, \ldots, p$ are bounded by $C_0, \ldots, C_p$, and for which the partial derivatives of order p satisfy Hölder conditions of order $q - p$ with constant C_q. Thus for each $\phi \in \mathcal{F}^{(q)}$ and $t, t + \tau \in T$

$$\phi(t+\tau) = \sum_{n=0}^{p} \frac{\Phi_n(t,\tau)}{n!} + \Delta(t,\tau), \tag{1.36}$$

where $\Phi_n(t,\tau)$ is a homogeneous polynomial of degree n in τ of the form

$$\Phi_n(t,\tau) = \Phi_n(t: \tau_1, \ldots, \tau_k) = \sum_{j_1=1}^{k} \cdots \sum_{j_n=1}^{k} \frac{\partial^n \phi(t)}{\partial t_{j_1} \ldots \partial t_{j_n}} \tau_{j_1} \cdots \tau_{j_n}, \tag{1.37}$$

and where

$$\sup_{t \in T} \left| \frac{\partial^n \phi(t)}{\partial t_{j_1} \ldots \partial t_{j_n}} \right| \leq C_n \quad \text{and} \quad |\Delta(t,\tau)| \leq C_q \|\tau\|^q. \tag{1.38}$$

Two things are obvious in the above setup, in which you should think of the dimension k as fixed. Firstly, the larger the α in (1.34) the rougher the process with this covariance will be. Secondly, the larger q is the smaller the family $\mathcal{F}^{(q)}$ will be, and thus the more likely that a Gaussian process defined on $\mathcal{F}^{(q)}$ will be continuous. Thus it seems reasonable to expect that a result of the following kind should be true. (For a partial result in the converse direction, see Example 3.6.)

1.7 THEOREM. *A centered Gaussian process with covariance function satisfying (1.33) and (1.34) will be continuous on $\mathcal{F}^{(q)}(T, C_0, \ldots, C_p, C_q)$ if $k > \alpha$ and if the following condition is satisfied:*

$$q > \frac{1 + \alpha - k}{2}.$$

REMARKS: Note that since we have not specified any other metric on $\mathcal{F}^{(q)}$, the continuity claim of the Theorem is with relation to the topology induced

by the canonical metric. There are, of course, more natural metrics on $\mathcal{F}^{(q)}$, but as long as we are interested in mere continuity and not in more detailed information such as moduli of continuity we need not bother with them.

If, despite the advice given above, you want to let k vary in the condition of the Theorem, while keeping α and q fixed, you will find that the larger k is, the less derivatives we require of our test functions to ensure continuity on $\mathcal{F}^{(q)}$. While at first this seems counter-intuitive, you should remember that as k increases the degree of the singularity in (1.34) decreases (for fixed α) and so the result is, in fact, reasonable.

While the reason for the assumption $k > \alpha$ should be obvious from the proof, it is worthwhile noting already that it is precisely this condition that gives us a process with finite variance, since, for $\phi \in \mathcal{F}^{(q)}$ with $\|\phi\|_\infty < M$

$$\begin{aligned} EX^2(\phi) &= \int_T \int_T \phi(s) R(s,t) \phi(t)\, ds dt \\ &\le M^2 \int_T \int_T R(s,t)\, ds dt \\ &\le CM^2 \int_T \int_T \|t-s\|^{-\alpha}\, ds dt. \end{aligned}$$

Since $T \subset \mathfrak{R}^k$ is compact, a transformation to polar coordinates easily shows that the last integral is finite only if $k > \alpha$.

PROOF: To make life notationally easier, we shall assume throughout that $T = [0,1]^k$.

The proof is, unfortunately, rather long. It will proceed via an entropy argument. Thus what we need to show, in the light of Theorem 1.1, is that the entropy of $\mathcal{F}^{(q)}$, as measured in the canonical metric d, where

$$d^2(\phi,\psi) = \int_{\mathfrak{R}^k} \int_{\mathfrak{R}^k} \big(\phi(s) - \psi(s)\big) R(s,t) \big(\phi(t) - \psi(t)\big)\, ds dt,$$

is small enough for the entropy integral (1.3) to converge. We shall do this by explicitly constructing, for each $\epsilon > 0$, a finite family $\mathcal{F}_\epsilon^{(q)}$ of functions that serve as an ϵ-net for $\mathcal{F}^{(q)}$ in the d metric.

We start by fixing $\epsilon > 0$ and defining

$$\delta = \delta(\epsilon) = \epsilon^{1/(q+(k-\alpha)/2)}. \tag{1.39}$$

Let Z_δ denote the grid of the $(1 + [\delta^{-1}])^k$ points in $[0,1]^k$ of the form

$$t_\eta = (\eta_1 \delta, \ldots, \eta_k \delta), \qquad \eta_i = 0, 1, \ldots, [\delta^{-1}], \quad i = 1, \ldots, k. \tag{1.40}$$

Define

$$\delta_n = \delta_n(\epsilon) = \delta^{q-n}, \qquad n = 0, \ldots, p,$$

and for each $\phi \in \mathcal{F}^{(q)}$, $n = 0, \ldots, p$, and t_η of the form (1.40) let $A_\eta^{(n)}(\phi)$ denote the vector formed by taking the integer part of δ_n^{-1} times the partial derivatives of ϕ of order n evaluated at the point t_η. (The index η here is, of course, k-dimensional.) Thus, a typical element of $A_\eta^{(n)}$ is of the form

$$(1.41) \qquad \left(A_\eta^{(n)}(\phi)\right)_i = \left[\frac{\phi^{(n_1,\ldots,n_k)}(t_\eta)}{\delta_n}\right], \qquad n_1 + \cdots + n_k = n,$$

where we have written $\phi^{(n_1,\ldots,n_k)}$ for the derivative $\partial^n \phi / \partial^{n_1} \ldots \partial^{n_k}$, and the index i runs from 1 to $\binom{n+k-1}{k-1}$.

Finally, for $\phi \in \mathcal{F}^{(q)}$, let $A_\delta = A_\delta(\phi)$ denote the $(k+1)$-dimensional matrix, a generic element of which is the vector $A_\eta^{(n)}$, $t_\eta \in Z_\delta$, $n = 0, \ldots, p$, (η is k-dimensional, n is one-dimensional) and let $F_{A(\delta)}$ denote the set of $\phi \in \mathcal{F}^{(q)}$ with fixed matrix A_δ. Our first task will be to show that the d-radius of $F_{A(\delta)}$ is not greater than $C\epsilon$, where C is a constant dependent only on q and k. All that will then remain will be to calculate how many different collections $F_{A(\delta)}$ are required to cover $\mathcal{F}^{(q)}$.

Thus, take $\phi_1, \phi_2 \in A_\delta$, and set

$$(1.42) \qquad \phi = \phi_1 - \phi_2.$$

Let $\| \ \|_d$ be the norm induced on $\mathcal{F}^{(q)}$ by the metric d, and $\| \ \|_\infty$ the usual sup norm. Thus,

$$\|\phi\|_d^2 = \int_{[0,1]^k} \int_{[0,1]^k} \phi(s) R(s,t) \phi(t)\, ds dt, \qquad \|\phi\|_\infty = \sup_{[0,1]^k} |\phi(t)|.$$

We have to show that the ϕ of (1.42) has d-norm less than $C\epsilon$.

Note first, however, that in view of the definition of the matrix A_δ via (1.41), we have that for each $t_\eta \in Z_\delta$ and each partial derivative $\phi^{(n_1,\ldots,n_k)}$ of such a ϕ of order $n_1 + \cdots + n_k = n \leq p$ that

$$\phi^{(n_1,\ldots,n_k)}(t_\eta) \leq \delta_n.$$

(Don't forget that the square brackets in (1.41) mean "integer part".)

Putting this inequality together with the Taylor expansion (1.36) and (1.37), we find that for all $t \in [0,1]^k$

$$\begin{aligned} |\phi(t)| &\leq \sum_{n=0}^{p} \frac{k^n}{n!} \delta_n \delta^n + C\delta^q \\ &= C(k,p)\delta^q, \end{aligned}$$

the last line following from the definition of the δ_n and the fact that each polynomial Φ_n of (1.37) has less than k^n distinct terms.

Thus, for ϕ of the form (1.42),

$$\|\phi\|_\infty \leq C\delta^q. \tag{1.43}$$

We now turn to $\|\phi\|_d$. With δ as above, set

$$D_\delta = \{(s,t) \subset [0,1]^k \times [0,1]^k : \max_{i=1,\ldots,k} |s_i - t_i| \leq \delta\}.$$

Then

$$\begin{aligned} \|\phi\|_d^2 &= \int_{[0,1]^k \times [0,1]^k} \phi(s)R(s,t)\phi(t)\,dsdt \\ &= \int_{D_\delta} \phi(s)R(s,t)\phi(t)\,dsdt \\ &\quad + \int_{([0,1]^k \times [0,1]^k)\setminus D_\delta} \phi(s)R(s,t)\phi(t)\,dsdt \qquad (1.44) \\ &= I_1(\delta) + I_2(\delta). \end{aligned}$$

Consider the first integral. Letting C change from line to line where necessary, we have from (1.34) and (1.43) that

$$\begin{aligned} I_1(\delta) &\leq C\,\delta^{2q} \int_{[0,1]^k} ds \int \cdots \int_{s_i - \delta \leq t_i \leq s_i + \delta} dt\, \|s-t\|^{-\alpha} \\ &\leq C\,\delta^{2q} \int_{[-\delta,\delta]^k} \|t\|^{-\alpha}\,dt \qquad (1.45) \\ &\leq C\,\delta^{(2q+k-\alpha)}, \end{aligned}$$

the last inequality coming from an evaluation via polar coordinates, and requiring the condition $k > \alpha$.

Similarly, again relying on the fact that $k > \alpha$, it is easy to check that $I_2(\delta)$ is also bounded above by $C\,\delta^{(2q+k-\alpha)}$. Substituting this fact and (1.45) into (1.44), and applying (1.39), we finally obtain that for ϕ satisfying (1.42)

$$\|\phi\|_d < C\,\delta^{\frac{1}{2}(2q+k-\alpha)} = C\epsilon. \tag{1.46}$$

That is, the d-radius of each set $A_\eta^{(n)}$ is no greater than a uniform constant times ϵ.

It remains to determine how many collections $F_{A(\delta)}$ are required to cover $\mathcal{F}^{(q)}$. Since this is a calculation that is now independent of both Gaussian processes in general, and the above covariance function in particular, we

shall only outline how this is done. The details, which require somewhat cumbersome notation, can be found in Kolmogorov and Tihomirov (1959).

Consider, for fixed δ, the $(k+1)$-dimensional matrix A_δ, parameterised, as in (1.40), by $\eta_i = 0,1,\dots,[\delta^{-1}]$, $i = 1,\dots,k$, and $n = 0,1,\dots,p$. Fix, for the moment, $\eta_2 = \cdots = \eta_k = 0$. It is clear from the restrictions (1.38), (1.41), the definition of δ_n, and the fact that each vector $A_\eta^{(n)}$ has no more than $\binom{n+k-1}{k-1}$ distinct elements, that there are no more than

$$O\Big(\frac{1}{\delta_0}\Big(\frac{1}{\delta_1}\Big)^{\binom{k}{k-1}} \cdots \Big(\frac{1}{\delta_p}\Big)^{\binom{p+k}{k-1}}\Big) = O(\delta^{-\xi}),$$

(for an appropriate and eventually unimportant ξ) ways to fill in the row of A_δ corresponding to $(n_1,\dots,n_k) = (0,\dots,0)$.

The main part of the proof is to now show that because of the rigid continuity conditions on the functions in $\mathcal{F}^{(q)}$, there exists an absolute constant $M = M(q, C_0, \dots, C_p, C_q)$, such that once this first row is determined, there are no more than M ways to complete the row corresponding to $(n_1,\dots,n_k) = (1,\dots,0)$, and similarly no more than M^2 ways to complete the row corresponding to $(n_1,\dots,n_k) = (2,\dots,0)$, etc. Thus, all told, there are no more than

$$O\big(\delta^{-\xi} \cdot M^{k(1+\delta^{-1})}\big) \tag{1.47}$$

ways to fill the matrix A_δ, and thus we have a bound for the number of different collections $F_{A(\delta)}$.

Modulo a constant, it now follows from (1.39), (1.46) and (1.47) that the logarithm of the entropy function for our process is bounded above by

$$C_1 + \frac{C_2\,\xi}{(q+(k-\alpha)/2)} \log\Big(\frac{1}{\epsilon}\Big) + C_3\Big(\frac{1}{\epsilon}\Big)^{1/(q+(k-\alpha)/2)}. \tag{1.48}$$

Substituting this into the entropy condition (1.3) (noting that it is really only the last term above that is important) is all that is now required to complete the proof of the theorem. ■

Before leaving this example, there are a number of comments that are worth making, that relate it to other problems both within and outside of the theory of Gaussian processes.

Firstly, in most of the literature pertaining to generalised Gaussian fields the parameter space used is the Schwartz space $\mathcal{S}$ of infinitely differentiable functions decaying faster than any polynomial at infinity. Since this is a very small class of functions (at least in comparison to the classes $\mathcal{F}^{(q)}$ that Theorem 1.7 deals with) continuity over $\mathcal{S}$ is automatically assured and therefore not often explicitly treated. However, considerations of continuity and

smaller parameter spaces are of relevence in the treatment of infinite dimensional diffusions, and associated infinite dimensional stochastic partial differential equations, in which solutions over very specific parameter spaces are often sought. For more on this see, for example, Itô (1983) and Watanabe (1984).

Secondly, some words on our choice of (1.34) as a condition on the covariance kernel $R(s,t)$. When $\alpha = k-2$, $k>2$ then the class of generalised fields that we are considering here includes the so called "free field" of Euclidean quantum field theory. (When $k=2$ the free field has a covariance kernel with a logarithmic singularity at 0, and when $k=1$ the free field is no longer generalised, but is the stationary Markov Gaussian process with covariance function $R(t)=e^{-\beta|t|}$, for some $\beta>0$.) This process, along with a large number of related generalised fields whose covariance kernels satisfy similar conditions, possesses a type of multi-dimensional Markov property. For details on this see, for example, Dynkin (1980, 1984), Adler and Epstein (1987), and references therein. For structural and renormalisation properties of generalised fields of this kind, presented among a much wider class of examples, see Dobrushin (1979), who also treats a large variety of non-Gaussian fields. We shall have a little more to say about some of these processes in Chapter 6, where we shall discuss how they are related to $\Re^d$-valued Markov processes on the real line.

Finally, it is worth noting that much of what has been said above regarding generalised fields – i.e. function indexed processes – can be easily extended to Gaussian processes indexed by a family of measures. For example, if we consider the function ϕ in (1.32) to be the (positive) density of a measure μ on $\Re^k$, then by analogy with (1.32) it makes sense to write

$$X(\mu) \;=\; \int_{\Re^k} X(t)\,\mu(dt),$$

with the corresponding covariance functional

$$R(\mu,\nu) \;=\; EX(\mu)X(\nu) \;=\; \int_{\Re^k}\int_{\Re^k} \mu(ds)R(s,t)\nu(dt).$$

Again, as was the case for generalised Gaussian fields, the process $X(\mu)$ may be well defined even if the covariance kernel R diverges on the diagonal. In fact, $X(\mu)$ will be well defined for all $\mu \in \mathcal{M}_R$, where

$$\mathcal{M}_R \;=\; \{\mu\colon \int_{\Re^k}\int_{\Re^k} \mu(ds)R(s,t)\mu(dt) \;<\; \infty\}.$$

Similar arguments to those used above to characterise the continuity of a family of Gaussian fields on $\mathcal{T}^{(q)}$ can be used to ascertain continuity

of measure indexed processes on suitably smooth classes of measures. We leave both the details and an attempt to formulate the appropriate results to the interested reader. (Actually, to the best of my knowledge, this specific example has never been treated in the literature, and so it would be a rather interesting problem to work out how to optimally, and naturally, formulate the requisite smoothness conditions. The general idea of how to proceed can be gleaned from the treatment of the following example.)

5. SET INDEXED PROCESSES:

We have already met some set indexed processes in dealing with the Brownian family of processes in the previous section. The motivation behind studying those processes lay in the development of multivariate Kolmogorov-Smirnov tests based on statistics of the type (1.11). Our aim now will be merely to introduce two new families $\mathcal{A}$ of subsets of $\Re^k$ that have proven useful as parameter spaces for the Gaussian white noise W satisfying (1.5)–(1.7), so that

$$EW(A)W(B) = R_\nu(A,B) = \nu(A \cap B), \qquad A, B \in \mathcal{A}, \tag{1.49}$$

where ν is a σ-finite measure on $\Re^k$.

Before we start in earnest, however, it is worthwhile taking a moment to consider the particularly simple family of special cases obtained when we take $\mathcal{A}$ to be a family of n-sided polygons in $\Re^2$. (A similar argument will apply to n-sided polyhedra in higher dimensions.)

To simplify the notation, let $\mathcal{A}$ be the family of all rectangular subsets of $[0,1]^2$. A typical element of $\mathcal{A}$ is $A = [s,t] \times [u,v]$. The geometry of W on this $\mathcal{A}$ is non-trivial, and can be conceptually simplified by defining a new, point indexed, random field X on the set

$$\mathcal{D} = \{(s,t,u,v) \in [0,1]^4 : s \le t,\ u \le v\}$$

by the correspondence

$$X(s,t,u,v) = W\big([s,t] \times [u,v]\big).$$

Thus, the questions of sample path continuity and boundedness of W on $\mathcal{A}$ reduce to the same questions for the point indexed random field X on $\mathcal{D}$, and we have already seen, at the beginning of the subsection on generalised random fields, how to treat such cases.

However, as simple as this approach may sound, it is not always quite so straightforward to implement. The main problem is that whereas the covariance function of W on $\mathcal{A} \times \mathcal{A}$ is natural and easy to work with, the covariance of X on $\mathcal{D} \times \mathcal{D}$ has very little in the way of "geometric symmetry", and specific calculations can often be difficult, depending on the structure of ν. A secondary problem is that some of the clever arguments that we

shall encounter in Chapter 5 for similarly indexed Brownian sheets – arguments that rely heavily on geometrical properties of $\mathcal{A}$ – break down on $\mathcal{D}$. Finally, of course, there is the problem that similar methods will not apply when a parameter space of sets does not have as simple a finite dimensional representation as do spaces of polyhedra.

It is to two such examples that we now turn.

The first family of sets is closely related to the family $\mathcal{F}^{(q)}$ of functions we have just studied in detail, and is, essentially, composed of those sets in $\mathfrak{R}^k$ that have q times differentiable boundaries. It is developed as follows:

Let S^{k-1} be the unit sphere in $\mathfrak{R}^k$, so that

$$S^{k-1} = \{t \in \mathfrak{R}^k : \|t\|^2 = 1\}.$$

Recall the basic result from differential topology that we can cover S^{k-1} by finitely many coordinate patches V_j, each represented as a C^∞ isomorphism $F_j : B^{k-1} \to V_j$ of the open ball $B^{k-1} = \{t \in \mathfrak{R}^{k-1} : \|t\|^2 < 1\}$.

Adapting slightly the notation of the previous example, let $\mathcal{F}^{(q)}(V_j, M)$ be the set of all real valued functions ϕ on V_j such that $\phi \circ F_j \in \mathcal{F}^{(q)}(B^{k-1}, M, \ldots, M)$ (c.f. (1.36)–(1.38)). Furthermore, let $\mathcal{F}^{(q)}(S^{k-1}, M)$ denote the set of all real valued functions ϕ on S^{k-1} such that the restriction of ϕ to V_j is in $\mathcal{F}^{(q)}(V_j, M)$. Taking the k-fold Cartesian product of copies of $\mathcal{F}^{(q)}(S^{k-1}, M)$, we obtain a family of functions from S^{k-1} to $\mathfrak{R}^k$, which we denote by $D(k, q, M)$, where the "D" stands for "Dudley", who introduced this family in Dudley (1974).

Each $\phi \in \mathcal{F}^{(q)}$ defines a $(k-1)$-dimensional surface in $\mathfrak{R}^k$, and a simple algebraic geometric construction enables one to "fill in" the interior of this surface to obtain a set I_ϕ. We shall denote the family of sets obtained in this fashion by $I(k, q, M)$, and call them the "Dudley sets with q-times differentiable boundaries".

Now we specialise to the special case in which the measure ν in (1.49) is Lebesgue measure, so that the process W there becomes the Brownian sheet. Calculations, whose nature we shall discuss in the proof outline below, then show that the log-entropy function for the Brownian sheet on $I(k, q, M) \cap T$, where $T \subset \mathfrak{R}^k$ is compact, satisfies

(1.50)

$$\log N\big(I(k, q, M), \epsilon\big) \leq \begin{cases} C_T \left(\frac{1}{\epsilon}\right)^{2(k-1)/(kq-k+1)} & (k-1)/k < q \leq 1, \\ C_T \left(\frac{1}{\epsilon}\right)^{2(k-1)/q} & 1 \leq q. \end{cases}$$

(Equivalent lower bounds for the log-entropy are also available for certain values of k and q.)

As an immediate conseqence of the above and Theorem 1.1 we have

1.8 PROPOSITION. *The Brownian sheet is continuous on a bounded collection of Dudley sets in $\Re^k$ with q times differentiable boundaries if $q > k-1 \geq 1$. If $k-1 > 1 \geq q > 0$ or if $k-1 > q \geq 1$ then the Brownian sheet is unbounded with probability one.*

OUTLINE OF PROOF: The proof of the unboundedness part of the result is beyond us at the moment. As far as the proof of continuity is concerned, all that really needs to be proven are the entropy inequalities above. These rely the "simple algebraic geometric construction" noted above, and so we shall not bring them in detail. The basic idea, however, requires little more than noting that there are basically as many sets in $I(k,q,M)$ as there are are functions in $D(k,q,M)$, and we have already seen, in the previous example, how to count the number of functions in $D(k,q,M)$. You can find the details in Dudley (1974), or try it for yourself in Example 3.7. ■

The astute reader will have noticed that whereas both in §1.2 and above we defined Gaussian white noise via any σ-finite measure ν (c.f. (1.49)) up until now we have only discussed the Brownian sheet based on Lebesgue measure. If we wish to extend the discussion to more general measures, and retain properties of continuity, we need to substantially restrict the class of sets $\mathcal{A}$ under consideration as the parameter space.

The most common classes of sets that arise in this context are the so called Vapnick-Červonenkis, or VC, sets, due, not surprisingly, to Vapnick and Červonenkis (1971). Since the arguments involved in entropy calculations are of an essentially combinatoric nature, and so of a somewhat different form to those we have met before, we shall now discuss them in some detail. (For a much more substantial treatment, including the importance of VC classes to the problem of finding "universal Donsker classes" in the theory of empirical processes, see, for example, Dudley (1978a).)

Let E be a subset of $\Re^k$, and ν a probability measure on E. Given a class $\mathcal{C}$ of subsets of E and a finite set $F \subset E$, let $\Delta^{\mathcal{C}}(F)$ be the number of different sets $C \cap F$ for $C \in \mathcal{C}$. For $n = 1, 2, \ldots$, let

$$m^{\mathcal{C}}(n) := \max\{\Delta^{\mathcal{C}}(F) : F \text{ has } n \text{ elements}\}.$$

Clearly, $m^{\mathcal{C}}(n) \leq 2^n$ for all n. Also, set

$$V(\mathcal{C}) = \begin{cases} \inf\{n : m^{\mathcal{C}}(n) < 2^n\}, & \text{if } m^{\mathcal{C}}(n) < 2^n \text{ for some } n. \\ \infty, & \text{if } m^{\mathcal{C}}(n) = 2^n \text{ for all } n. \end{cases} \tag{1.51}$$

The class $\mathcal{C}$ is called a Vapnik-Červonenkis class if $m^{\mathcal{C}}(n) < 2^n$ for *some* n; i.e. if $V(\mathcal{C}) < \infty$. The number $V(\mathcal{C})$ is called the VC index of $\mathcal{C}$.

Two extreme but easy examples which you can check for yourself are $E = \Re$ and $\mathcal{C}$ all half lines, for which $m^{\mathcal{C}}(n) = n+1$ and $V(\mathcal{C}) = 1$, and $E = [0,1]$ with $\mathcal{C}$ all the open sets in $[0,1]$. Here $m^{\mathcal{C}}(n) = 2^n$ for all n and so $V(\mathcal{C}) = \infty$ and $\mathcal{C}$ is not a VC class.

A more instructive example, that also leads into the general theory we are after, is $E = \Re^k$ and $\mathcal{C}$ is the collection of half-spaces of $\Re^k$. Let $\Phi(k,n)$ be the maximal number of components into which it is possible to partition $\Re^k$ via n hyperplanes. Then, by definition, $m^{\mathcal{C}}(n) = \Phi(k,n)$. It is not hard to see that Φ must satisfy the following recurrence relation:

$$\Phi(k,n) = \Phi(k,n-1) + \Phi(k-1,n-1), \tag{1.52}$$

with the boundary conditions $\Phi(0,n) = \Phi(k,0) = 1$. (If $\Re^k$ has already been partitioned into $\Phi(k,n-1)$ subsets via $n-1$ ($(k-1)$-dimensional) hyperplanes, $H_1,\ldots,H_{n-1}$, then adding one more hyperplane H_n will cut in half as many of these subsets as intersect H_n. There can be no more such subsets, however, than the maximal number of subsets formed on H_n by partitioning with the $n-1$ $(k-2)$-dimensional hyperplanes $H_1 \cap H_n,\ldots,H_{n-1} \cap H_n$; i.e. $\Phi(k-1,n-1)$. Hence (1.52).)

Induction then shows that

$$\Phi(k,n) = \begin{cases} \sum_{j=0}^{k} \binom{n}{j} & \text{if } n > k, \\ 2^n & \text{if } n \le k, \end{cases} \tag{1.53}$$

where we adopt the usual convention that $\binom{n}{j} = 0$ if $n<j$.

From either the above or (1.52) you can now check that

$$\Phi(k,n) \le n^k + 1, \qquad \text{for all } k,n > 0. \tag{1.54}$$

It thus follows, from (1.51), that the half-spaces of $\Re^k$ form a VC class for all k. ■

What is somewhat more surprising, however, is that an inequality akin to (1.54), which we developed only for this special example, holds in general.

1.9 LEMMA. *Let E be any set, $\mathcal{C}$ any collection of subsets of E, and $V(\mathcal{C}) \le v$. Then*

$$m^{\mathcal{C}}(n) < \Phi(v,n) \le n^v + 1, \qquad \textit{for all } n \ge v. \tag{1.55}$$

Since the proof of this result is combinatoric rather than probabilistic, and will be of no further interest to us, you are referred to either Vapnick and Červonenkis (1971) or Dudley (1978a) for a proof.

The importance of Lemma 1.9 is that it enables us to obtain bounds on the entropy function for the Gaussian white noise based on *any* probability measure ν, and that these bounds are independent of ν.

1.10 THEOREM. *Let W be the Gaussian white noise based on a probability measure ν on some measure space $(E, \mathcal{E}, \nu)$. Let $\mathcal{A}$ be a Vapnik-Červonenkis class of sets in $\mathcal{E}$ with $V(\mathcal{A}) = v$. Then there exists a constant $K = K(v)$ (not depending on ν) such that for $0 < \epsilon \leq \frac{1}{2}$, the entropy function for W satisfies*

$$N(\mathcal{A}, \epsilon) \leq K\epsilon^{-2v} |\log \epsilon|^v.$$

PROOF: We start with a little counting, and then turn to the entropy calculation proper. The counting argument is designed to tell us something about the maximum number of $\mathcal{A}$ sets, that are a certain minimum distance from one another, that can be packed into E.

Suppose $A_1, \ldots, A_m \in \mathcal{A}$, $m \geq 2$, and $\nu(A_i \Delta A_j) \geq \epsilon$ for $i \neq j$. We need an upper bound on m. Sampling with replacement, select n points at random from E. The ν-probability that at least one of the sets $A_i \Delta A_j$ contains none of these n points is at most

$$\binom{m}{2}(1-\epsilon)^n. \tag{1.56}$$

Choose $n = n(m, \epsilon)$ large enough so that this probability is less than 1. Then

$$P\{\text{all symmetric differences } A_i \Delta A_j \text{ are non-empty}\} > 0,$$

and so for at least one configuration of the n sample points the class $\mathcal{A}$ picks out at least m distinct subsets. (Since, with positive probability, given any two of the A_i there is at least one point not in both of them.) Thus, by (1.55),

$$m \leq m^{\mathcal{A}}(n) \leq n^v = \big(n(m,\epsilon)\big)^v. \tag{1.57}$$

Take now the smallest n for which (1.56) is less than 1. For this n we have $m^2(1-\epsilon)^{n-1} \geq 2$, so that

$$n - 1 \leq \frac{2\log m - \log 2}{|\log(1-\epsilon)|},$$

and $n \leq (2\log m)/\epsilon$. Furthermore, by (1.57), $m \leq (2\log m)^v \epsilon^{-v}$.

For some $m_0 = m_0(v) < \infty$, $(2\log m)^v \leq m^{1/(v+1)}$ for $m \geq m_0$, and then $m \leq \epsilon^{-v-1}$, so $\log m \leq (v+1)|\log \epsilon|$. Hence

$$m \leq K(v)\,\epsilon^{-v} |\log \epsilon|^v \qquad \text{for } 0 < \epsilon \leq \tfrac{1}{2} \tag{1.58}$$

if $K(v) = \max(m_0, 2^{v+1}(v+1)^v)$.

This concludes the counting part of the proof. We can now do the entropy calculation. Recall that the canonical distance between on sets of $\mathcal{E}$ is given by $d_\nu(A,B) = [\nu(A\Delta B)]^{\frac{1}{2}}$.

Fix $\epsilon > 0$, and suppose that $A_1, \dots, A_m$ are sets in $\mathcal{A}$ for which the union of ϵ-neighbourhoods of the A_i (in the d_ν metric) covers $\mathcal{A}$. By judicious trimming (see the construction of (4.15)–(4.16) for an example of the details) of the intersections of these neighbourhoods, one can assume that $d_\nu(A_i, A_j) \geq \epsilon$ for all i,j. Thus, by (1.58), $m \leq K(v)\epsilon^{2v}|2\log\epsilon|^v$, which proves the theorem. ∎

An immediate consequence of the entropy bound of Theorem 1.10 is

1.11 COROLLARY. *Let W be the Gaussian white noise based on a probability measure ν on some measure space $(E, \mathcal{E}, \nu)$. Then W is continuous over any Vapnik-Červonenkis class of sets in $\mathcal{E}$.*

With this we complete our two new examples of families of index sets for Gaussian processes.

You can find a lot more material on Vapnik-Červonenkis classes in the lecture notes by Dudley (1984), and some material for exercising your combinatorial skills in Examples 3.8–3.10. (If you find your combinatorial skills are not as sharp as you might like, look up Dudley's notes to find out how to solve the exercises.)

6. VECTOR VALUED PROCESSES:

Whereas the general theory espoused in the first two sections, and the examples we have looked at up until now, cover Gaussian processes on very general parameter spaces, all of our effort has been concentrated on real valued processes. It seems worthwhile to say at least a little about how well these results extend to Gaussian processes taking values in more general state spaces.

Thus, let $\mathbf{X}_t = \langle X_t^1, \dots, X_t^N \rangle$ be a $\Re^N$-valued Gaussian process on a metric space (T,τ); i.e. each X_t^i is a real valued Gaussian process on (T,τ). The covariance function that has formed the basis of all our analysis up until now must be replaced with a matrix valued function $\mathbf{R}(s,t)$ with generic element $R_{ij}(s,t) = EX_s^i X_t^j$. We shall assume, as usual, that the X_t^i have zero mean.

Despite this complication, the question of the continuity of $\mathbf{X}$ is no harder than before. Since $\mathbf{X}$ is continuous as a function from (T,τ) to $(\Re^N, \|\ \|)$ if, and only if, each X_t^i is continuous as a real valued function on (T,τ), the global continuity question can be treated as N separate, real valued, problems.

The problem of the behaviour of the supremum of $\mathbf{X}$ is not quite as simple. If we concentrate our attention on the random variable

$$\sup_{t\in T} \|\mathbf{X}_t\| = \sup_{t\in T} \left((X_t^1)^2 + \cdots + (X_t^N)^2\right)^{\frac{1}{2}}, \tag{1.59}$$

then it is clear that we are treating here the supremum of the non-Gaussian process $\|\mathbf{X}\|$. (If the N components of $\mathbf{X}$ are i.i.d. then $\|\mathbf{X}_t\|$ has a χ^2_N distribution. If they are independent but not identically distributed, then the distribution of $\|\mathbf{X}_t\|$ is non-central χ^2. In the general case, the distribution of $\|\mathbf{X}_t\|$ is algebraically complicated.) Thus there is no good reason to expect, *a priori*, that Gaussian techniques should work in handling $\sup_t \|\mathbf{X}_t\|$. Nevertheless, the same general approach does to work (at least in the case $T = \Re^k$ and i.i.d. components – see, for example, the treatment of χ^2 processes in Leadbetter, Lindgren and Rootzén (1983) and of χ^2 fields Adler (1981)) but the extension is not trivial and involves considerable hard work.

One way to get results while remaining in a purely Gaussian framework is to rewrite the Euclidean norm in (1.59) as an operator type norm: i.e.

$$\|\mathbf{X}_t\| := \sup_{\mathbf{a} \in S^N} \mathbf{a} \cdot \mathbf{X}_t, \tag{1.60}$$

where $S^N = \{\mathbf{a} \in \Re^N : \|\mathbf{a}\| = 1\}$ is the unit sphere in $\Re^N$.

If we now introduce a new, real valued Gaussian process $X(\mathbf{a}, t)$ on $S^N \times T$ by the correspondence

$$X(\mathbf{a}, t) = \mathbf{a} \cdot \mathbf{X}_t, \tag{1.61}$$

then

$$\sup_{t \in T} \|\mathbf{X}_t\| \equiv \sup_{(\mathbf{a}, t) \in S^N \times T} X(\mathbf{a}, t). \tag{1.62}$$

The righthand supremum here is of a real valued Gaussian process, albeit over a somewhat involved parameter space, and so in principle is amenable to the theory that we shall develop to handle such suprema in the following chapters.

In fact, it is easy to relate the canonical metric of $X(\mathbf{a}, t)$ to that of X_t, since

$$\begin{aligned} E\big(X(\mathbf{a}, s) - X(\mathbf{b}, t)\big)^2 &= E\big(\mathbf{a} \cdot \mathbf{X}_s - \mathbf{b} \cdot \mathbf{X}_t\big)^2 \\ &= E\big(\mathbf{a} \cdot (\mathbf{X}_s - \mathbf{X}_t) + (\mathbf{a} - \mathbf{b}) \cdot \mathbf{X}_t\big)^2. \end{aligned}$$

Consider the special case of i.i.d. components with $\sigma_T^2 = \sup_T EX_t^2$. Then the above gives us

$$\begin{aligned} d^2\big((\mathbf{a}, s), (\mathbf{b}, t)\big) &\leq \|\mathbf{a}\|^2 d^2(s, t) + \|\mathbf{a} - \mathbf{b}\|^2 \sigma_T^2 \\ &\qquad + 2\,\mathbf{a} \cdot (\mathbf{a} - \mathbf{b})\, E(X_s^i - X_t^i) X_t^i \\ &\leq \|\mathbf{a}\|^2 d^2(s, t) + \|\mathbf{a} - \mathbf{b}\|^2 \sigma_T^2 + 2\,\mathbf{a} \cdot (\mathbf{a} - \mathbf{b})\, d(s, t) \sigma_T, \end{aligned}$$

by Cauchy-Schwartz. Note $\|a\| = 1$ to obtain

$$d\big((\mathbf{a},s),(\mathbf{b},t)\big) \;\le\; d(s,t) \;+\; \|\mathbf{a}-\mathbf{b}\|\,\sigma_T\,.$$

It now follows from simple geometry on the N-dimensional sphere that if we write $N_{S^N,T}(\epsilon)$ for the entropy function of $X(\mathbf{a},t)$, $\mathbf{a}\in S^N$, $t\in T$, and $N_T(\epsilon)$ for that of each of the i.i.d. components X_t^i on T then

$$N_{S^N,T}(\epsilon) \;\le\; O\Big(\big(N_T(\epsilon)/\epsilon\big)^N\Big).$$

In the light of (1.62) this is all we need in order to handle those aspects of the distribution of $\sup_{t\in T}\|X_t\|$ that are amenable to entropy methods. ■

If the X_t^i are neither independent nor identically distributed then the above calculation may not be quite as simple, and this approach not as amenable. Nevertheless, this is often a good trick to remember. As we shall see in the next example, it also generalises to much more complex situations.

However, before we leave vector valued processes, it is worthwhile to note that a number of interesting questions arise for these processes, involving delicate interplays between the dimension N of the state space, the dimensionality of the parameter space, and measures of smoothness of the covariance function, that either do not arise, or are trivial, in the context of real valued processes.

For example, suppose the parameter space is $\mathfrak{R}^k$, the component processes X_i are independent, and for each $i = 1,\dots,N$,

$$E|X_s^i - X_t^i|^2 \;\asymp\; \|t-s\|^{2\beta_i} \qquad \text{as } \|t-s\|\to 0, \tag{1.63}$$

where of necessity $0<\beta_i\le 1$ for all i and we write $f(t)\asymp g(t)$ as $t\to t_0$ if there exist finite constants c and C such that

$$c \;\le\; \lim_{t\to t_0}\frac{f(t)}{g(t)} \;\le\; C. \tag{1.64}$$

A process of this form is called a (k,N) Gaussian field of index $\beta=(\beta_1,\dots,\beta_N)$.

Let $\dim A$ denote the Hausdorff dimension of a Euclidean set A. Then, rearranging indices, if necessary, so that $0<\beta_1\le\cdots\le\beta_N\le 1$,

(1.65)

$$\dim(\operatorname{Im}\mathbf{X}) \;=\; \min\Big[N,\;\frac{k+\sum_{i=1}^N(\beta_N-\beta_i)}{\beta_N}\Big],$$

(1.66)

$$\dim(\operatorname{Gr}\mathbf{X}) \;=\; \min\Big[k+\sum_{i=1}^N(1-\beta_i),\;\frac{k+\sum_{i=1}^N(\beta_N-\beta_i)}{\beta_N}\Big],$$

(1.67)

$$\dim(\mathbf{X}^{-1}(\mathbf{u})) \;=\; \max\Big[0,\;k-\sum_{i=1}^N\beta_i\Big],$$

where (1.65) and (1.66) hold with probability one, and (1.67) holds with positive probability for each $\mathbf{u} \in \mathfrak{R}^k$.

For more details on results of this kind, including a formal definition of Hausdorff dimension, see, for example, Chapter 8 of Adler (1981).

7. BANACH SPACE VALUED PROCESSES:

Banach space valued Gaussian processes arise in a number of different contexts, the two most common being in the areas of empirical processes and stochastic partial differential equations (SPDE's). Here I only want to show you that the basic questions of continuity and boundedness of these processes can be handled, if one so desires, in the framework of real valued processes. The argument is much the same as that we just saw for vector valued processes.

Thus, let B be a separable Banach space, and B^* its topological dual. A stochastic process X_t taking values in B is called Gaussian if for every $n \geq 1$, and all collections $t_1, \ldots, t_n \in T$ and $\phi_1, \ldots, \phi_n \in B^*$ the n-dimensional random variable

$$\big(\langle \phi_1, X_{t_1} \rangle, \ldots, \langle \phi_n, X_{t_n} \rangle\big)$$

has a multivariate Gaussian distribution. The covariance functional will be of the form

$$R_{st}(\phi, \psi) \;=\; E\,\langle \phi, X_s \rangle\, \langle \psi, X_t \rangle. \tag{1.68}$$

The continuity question for these processes is generally two-fold. Firstly, it is generally required that for fixed $t \in T$, $X_t(.)$ be a continuous linear functional on B^*. Secondly, for fixed $\phi \in B^*$, one wants that $\langle \phi, X_t \rangle$ be continuous in t. Since both of these problems relate to real valued processes, in the first case indexed by B^* and in the second by T, both fall within the realm of the general theory we have espoused so far.

If the continuity question is one of joint continuity in t and ϕ, then the approach developed in the previous example suggests defining the new (real valued) process

$$X(\phi, t) \;=\; \langle \phi, X_t \rangle,$$

with covariance function

$$R\big((\phi, s), (\psi, t)\big) \;=\; R_{st}(\phi, \psi),$$

and studying its continuity on $B^* \times T$.

This approach is of particular interest in the theory of SPDE's. There one is often interested in an equation which has a solution only in an appropriate family of functions (i.e. Banach space B) which must be specified in advance. In this case arguments of the kind used in the treatment of generalised Gaussian fields above often help one to limit, or at least to guess, the appropriate solution space in advance.

To be fair we should note that the trick of simplifying the state space at the cost of complicating the parameter space is not always easy to apply in practice, since entropy or similar calculations have to be made over $B^* \times T$, which will not generally be a nice space to work with. The most common problem is that T and B^* are very different kinds of spaces. An example of the application of this idea can be found in the paper by Iscoe *et. al.* (1989) in which the continuity of ℓ^2-valued Ornstein-Uhlenbeck processes is treated.

Before leaving this family of examples, however, it is worth noting that Banach space valued Gaussian processes are often not difficult to deal with directly, and, much as in the way that the geometric structure of the parameter space T does not enter the general theory of Gaussian processes, neither (usually) does the structure of the state space. Thus, many of the central results of these notes hold, and are almost as easy to prove, for Banach space instead of real valued processes. (There are, of course, too many annoying exceptions to this rule.) I chose to work with the real valued case in general – perhaps so that the "Introduction" part of the title of the notes will have some validity. Three recent papers with vector-valued results are Fernique (1987a,b, 1988).

8. NON-GAUSSIAN PROCESSES: A natural question to ask is whether or not the results and methods presented in these notes extend naturally to non-Gaussian stochastic processes. The answer, in general, must be somewhat equivocal.

In some cases there is an immediate transfer of results. To take a trivial but often useful example, for $F\colon \Re \to \Re$ and X_t Gaussian, the process $Y_t = F(X_t)$ is generally non-Gaussian, but continuity questions on Y are clearly simple, given enough information on X and F; i.e. the continuity question for the non-Gaussian Y can be reduced to a question on the Gaussian X. There are some questions, however, that even in this setting have surprising answers. For example, we shall see in the following chapter that a basic inequality that plays a central rôle in the theory of Gaussian processes (Slepian's inequality) fails to work for as simple a process as $|X_t|$.

A similar, but more interesting situation arises in the study of the so-called χ^2-processes. Given a sequence $X_1(t), \ldots, X_N(t)$ of independent copies of a Gaussian process, the corresponding χ^2-process is defined as

$$Z(t) = \sum_{i=1}^{N} X_i^2(t). \tag{1.69}$$

Again, it is clear that the continuity and boundedness questions for the non-Gaussian Z are equivalent to those for the Gaussian X_i, and no new methodolgy needs to be introduced. The comments of the preceeding paragraph, however, indicate that, *a fortiori*, Slepian's inequality cannot hold in this case either.

What if the process X_t, defined, as usual, on a metric space (T,τ), has no relationship whatsoever to a Gaussian process? (Note, that having left the Gaussian framework, we have also left the canonical metric d out of the picture.) Then Pisier (1980, 1981) and Fernique (1983) have various versions the following result, for which you need to know that a function $\phi\colon \mathfrak{R} \to \mathfrak{R}$ is called a Young function if it is even, continuous, convex, and satisfies

$$\lim_{x\to 0} \frac{\phi(x)}{x} = 0, \qquad \lim_{x\to\infty} \frac{\phi(x)}{x} = \infty.$$

1.12 THEOREM. *Let X_t be a Banach space valued stochastic process, defined on a metric space (T,τ), and assume that the process $\|X_t - X_s\|$ is separable. Let N_τ be the metric entropy function for T with respect to the metric τ. If there exists an $\alpha \in (0,1]$ and a Young function ϕ such that the following two conditions are satisfied, then X is continuous with probability one.*

$$E\phi\Big(\frac{\|X(t)-X(s)\|^\alpha}{\tau(s,t)}\Big) \le 1, \tag{1.70}$$

$$\int_{N_\tau(u)>1} \phi^{-1}\big(N_\tau(u)\big)\,du < \infty. \tag{1.71}$$

The power of this result is clear. One immediate implication is that much of the hard work that we have done in calculating entropy functions for Gaussian processes can be used with little or no extra effort to study the continuity properties of non-Gaussian processes as well. The only technical difficulty lies in finding a ϕ satisfying both (1.70) and (1.71). We shall look at only one, albeit very rich, class of examples – the so-called ξ-radial processes of Marcus (1987).

Recall that a symmetric, real valued, infinitely divisible random variable ξ is determined by a characteristic function of the form

$$Ee^{i\lambda\xi} = e^{-\Psi(|\lambda|)}, \tag{1.72}$$

where

$$\Psi(|\lambda|) = \int_0^\infty (\cos\lambda t - 1)\,d\nu[t,\infty), \tag{1.73}$$

and ν is a Lévy measure; i.e. a positive measure on $\mathfrak{R}_+ \setminus \{0\}$ satisfying

$$\int_0^\infty (1\wedge t^2)\,d\nu[t,\infty) < \infty.$$

When $\Psi(|\lambda|) = |\lambda|^\alpha$, then ξ is symmetric stable of index $\alpha \in (0,2]$. (The case $\alpha = 2$ is, of course, the Gaussian situation, and is somewhat different to the other cases. Nevertheless, we are certainly not lacking for a theory in this case.)

A real valued stochastic process X_t on T is called ξ-radial if there exists a probability measure m on the space of real valued functions $\beta(t)$ for which $\sup_{t\in T} |\beta(t)| = 1$, equipped with the cylindrical σ-algebra, such that for all measurable functions $\gamma(t)$

$$(1.74)\quad E\Big\{ \exp\Big(\sum_{t\in T} \gamma(t)X(t)\Big)\Big\} \;=\; \exp\Big\{ -\int \Psi\Big(\Big|\sum_{t\in T} \gamma(t)\beta(t)\Big|\Big)\, m(d\beta)\Big\}.$$

These processes are strictly stationary (why?), and the measure m is generally referred to as the spectral measure of X. Despite its name, however, it should not be confused with the usual spectral measure of a stationary stochastic process.

ξ-radial processes provide an interesting generalisation of Gaussian processes, and have, in one form or another, been of considerable interest over the past decade. In particular, while they are often highly non-Gaussian in their sample path behaviour, they are nevertheless amenable to a similar style of mathematical analysis. The stable case has been of particular interest, and you can find a good survey of these processes in Weron (1984). The forthcoming monograph of Samorodnitsky and Taqqu (199?) will provide the most wide ranging survey, however, along with many interesting and useful examples of stable processes beyond and including the ξ-radial case. Two of the simplest examples are given by the stable moving averages and harmonizable processes on $\Re$, representable, respectively, by integrals of the form

$$(1.75)\qquad \int_{\Re} H(t-s)\, Z(ds), \qquad\qquad \int_{\Re} e^{it\lambda}\, Z(dy),$$

where H is a nice function and Z a stable measure on $\Re$. (i.e. a stable process with independent increments.)

In the stable case it is not too hard to obtain the following corollary of Theorem 1.12. The proof (e.g. Fernique (1983)) relies, however, on various inequalities for stable variables and processes that I do not really want to develop here.

1.13 COROLLARY. *Let X_t be a real valued, symmetric stable process on (T,τ) of index $\alpha \in (0,1]$, and for $\beta \in (0,\alpha)$ let d_β be the distance function on T defined by*

$$(1.76)\qquad d_\beta(s,t) \;=\; E|X_t - X_s|^\beta.$$

Let N_β be the entropy function for (T, d_β). If, for some $\beta \in (0, \alpha)$

$$\int_0 \left(N_\beta\left(\epsilon^{1/\alpha}\right)\right)^{\beta/\alpha} \epsilon^{\beta-1}\, d\epsilon < \infty, \tag{1.77}$$

then X has continuous sample paths on T with probability one.

The paper of Marcus and Pisier (1984) and the more technical monograph of Marcus (1987) provide the deepest results on the general continuity problem for strictly stationary stable processes, and I shall do no more now than to refer you there to see not only how to improve on Theorem 1.12 for ξ-radial processes, but also to see how to go part of the way towards obtaining necessary, as well as sufficient, conditions for continuty.

It is worthwile noting at this stage, however, that whereas we shall soon see that for Gaussian processes it is possible to develop general necessary and sufficient conditions for sample path continuity, this is not a simple project for ξ-radial processes. Talagrand (1988b, 1989) has provided some necessary but not sufficient conditions for the boundedness and continuity of symmetric stable processes – in terms of majorising measures (see Chapter 4) – but this seems to be a problem whose solution is beyond our current means. (This is primarily because of the lack of one very specific tool – Slepian's inequality – that does not hold in the ξ-radial case but is extremely important in the necessity arguments in the Gaussian case.)

Finally, it is interesting to note that the discussion of Section 2 above, on the Brownian family of processes defined via Gaussian white noise, can be extended to function and set indexed processes defined via infinitely divisible white noise. In this case the resulting processes are rarely continuous, but often have cadlag versions (right continuous with left limits, in an appropriate sense). As was the case for continuity, entropy arguments can also be used to determine cadlaguity. For details see Adler and Feigin (1984) and Bass and Pyke (1984).

4. Exercises.

SECTION 1.1:

1.1 Show that any stochastic process has mean square continuous sample paths if and only if its covariance function R is continuous on $T \times T$. Check that this is equivalent to $\lim_{u \to 0} \rho(u) = 0$, where ρ is the function defined by (1.1). Show also that if R is continuous at diagonal points (t, t), then it is continuous everywhere.

SECTION 1.2:

2.1 Let W be a k-dimensional $(k > 1)$ Brownian sheet, and let $(i_1, \ldots, i_n)$, $n < k$, be a subset of $(1, 2, \ldots, k)$. Fix $s_{i_1}, \ldots, s_{i_n}$, $s_{i_j} \geq 0$ for all j, and let

Z be the process on $\mathfrak{R}_+^{k-n}$ defined by

$$Z(t_1,\ldots,t_{k-n}) = W(t_1,\ldots,t_{i_1-1},s_{i_1},t_{i_1+1}\ldots,t_{i_n-1},s_{i_n},t_{i_n+1},\ldots,t_{k-n}).$$

i.e. n of the k parameters of W are fixed by the s_{i_j}. Show that

$$|s_{i_1}\cdots s_{i_n}|^{-1/2}\, Z(t_1,\ldots,t_{k-n})$$

is a $(k-n)$-parameter Brownian sheet.

2.2 Let $\mathcal{A}_\gamma$ be the collection of subsets of $\mathfrak{R}^2$ described in §1.2 and whose construction is illustrated in Figure 1.2. Using an argument similar to that used to show that the Brownian sheet is unbounded on lower layers, show that it is also unbounded on $\mathcal{A}_\gamma$ if $1 < \gamma \leq 2$.

SECTION 1.3:

3.1 Let X be a continuous Gaussian process on $[0,1]$, f a homeomorphism of $[0,1]$, and Y a new process defined, also on $[0,1]$, by $Y_t = X_{f(t)}$. Show that X and Y have identical entropy functions. Show, by example, that it is possible for the covariance function of X to satisfy (1.14) while that of Y does not.

3.2 Show that the question of convergence or divergence of the integral (1.14) is equivalent to that of the integral

$$\int_K^\infty p(e^{-x^2})\,dx,$$

for arbitrary $K > 0$. Hence, or otherwise, show that (1.15) implies (1.14).

3.3 Assume for the moment that for a stationary Gaussian process on $\mathfrak{R}^1$ (1.14) is both necessary and sufficient for sample path continuity. Using the equivalence established in the previous exercise, show how the continuity/discontinuity question is answered by the relationship (1.18). The assumption will be justified in Chapter 4.

3.4 Prove the easy half of Theorem 1.6 concerning the Talagrand expansion. Let $\{Y_n\}_{n\geq 0}$ be a centered Gaussian sequence satisfying (1.28), and $\{\alpha_n\}_{n\geq 0}$ a sequence of functions satisfying (1.29) on a compact metric space T.
(i) Show that the process X_t defined by (1.30) is bounded with probability one. (You will need the inequality (2.1) for the proof, so use it freely for now. If you want a hint, see the argument following Theorem 4.6, which shows that entropy conditions don't always give tight conditions for continuity, and which therefore motivates half of these notes.)
(ii) It follows from the calculations used to prove (i) and the results of Chapter 4 that X is also continuous, but since we don't know these results

yet, show that the a.s. continuity of X will follow if we assume, in addition to the above, that the $\alpha_n(t)$ are continuous in t.

3.5 Show that if X_t is a centered Gaussian field on a compact subset of $\Re^k$, $k \geq 1$, and $p(u)$ is defined by (1.31), then the convergence of the entropy integral at (1.14) is still a sufficient condition for the continuity of X. Note that the proof for general k is really no different to that for $k=1$.

3.6 This is an exercise for the true lover of entropy methods: Theorem 1.7 gave sufficient conditions for the continuity of a Gaussian process indexed by a family of functions. Show that if the inequality in (1.34) also holds in the opposite direction (obviously with a different C) then the entropy calculations in the proof of Theorem 1.7 are sharp: i.e. the upper bound (1.48) also serves (obviously with different constants) as a lower bound. While this does not immediately imply that the conditions of the theorem are also necessary, it will give us some nice results when we come to talking about extrema distributions in Chapter 5.
(Hint: The route you will have to follow can be found on page 311 of Kolmogorov and Tihomirov (1959), which is the basic reference for all calculations of this kind.)

3.7 Complete the proof of Proposition 1.8.

3.8 Show that the finite union of Vapnik-Červonenkis classes is again a Vapnik-Červonenkis class.

3.9 Let $\mathcal{P}(k,n)$ be the set of n-sided polygons in $\Re^k$. Show that $\mathcal{P}(k,n)$ is a Vapnik-Červonenkis class of sets, and find its VC index.

3.10 If $\mathcal{F}$ is a family of functions on a space E, then for each $\phi \in \mathcal{F}$ set $\text{pos}(\phi) = \{t \in E: \phi(t) > 0\}$ and $\text{pos}(\mathcal{F}) = \{\text{pos}(\phi): \phi \in \mathcal{F}\}$. If $\mathcal{F}$ is a k-dimensional real vector space, such as polynomials of order $\leq k$, show that $\text{pos}(\mathcal{F})$ is a Vapnik-Červonenkis class with VC index k.