IV. BOUNDEDNESS AND CONTINUITY

1. Majorising Measures.

As usual, T is the parameter space of a centered Gaussian process X, equipped with and totally bounded in the canonical metric d. Let m be a probability measure on the Borel subsets of T, and, since we shall need it very often in the following, let $g\colon (0,1] \to \Re_+$ be the function defined by

$$g(t) = \big(\log(1/t)\big)^{\frac{1}{2}}, \qquad 0 < t \le 1. \tag{4.1}$$

Let $B(t,\epsilon)$ be an ϵ-ball in the d-metric about the point $t \in T$.

DEFINITION. *A probability measure m is called a majorising measure (for (T,d)) if*

$$\sup_{t\in T} \int_0^\infty g\big(m(B(t,\epsilon))\big)\, d\epsilon \ < \ \infty. \tag{4.2}$$

Majorising measures were formally introduced by Fernique (1974) (under the name "mesures majorantes"), although in essence they date back to a real variable lemma of Garsia, Rodemich and Rumsey (1970) and a paper of Preston (1972). Fernique proved (4.3) below, thus establishing that the existence of a majorising measure implied the boundedness of $\|X\| = \sup_{t\in T} X_t$. He continued his study of majorising measures in a number of papers, extending the ideas and coming very close to proving that they were the right tool to provide both necessary *and* sufficient conditions for continuity. For example, in Fernique (1978), he obtained both parts of Theorem 4.1 below under additional structural requirements on the space (T,d). In a pathbreaking paper, Talagrand (1987) completed Fernique's program by proving the second part of the following result in the most general case.

In this result, as with all others in this chapter, we shall assume without further comment that T has strictly positive diameter D. (Otherwise $m(B(t,\epsilon)) = 1$ for all ϵ, so that $g(m(B(t,\epsilon))) \equiv 0$, and the following result is trivial.)

4.1 THEOREM. *If m is any probability measure on (T,d), then*

$$E\|X\| \ \le \ K \sup_t \int_0^\infty g\big(m(B(t,\epsilon))\big)\, d\epsilon, \tag{4.3}$$

for some universal constant $K \in (1,\infty)$. Furthermore, K can be chosen such that if X is bounded with probability one, then there exists a probability measure m on (T,d) satisfying

$$K^{-1} \sup_t \int_0^\infty g\big(m(B(t,\epsilon)\big)\, d\epsilon \ \le \ E\|X\|. \tag{4.4}$$

That is, X is a.s. bounded on T if, and only if, (T,d) admits a majorising measure.

The proof is deferred to the following two sections. Note that, as was the case for the entropy integrals which we met in the Introduction, the upper limit of the integrals in (4.3) and (4.4) is really the d-diameter of T. Also note that while this result centers on the a.s. boundedness of X, we have already seen in the previous chapter that this is also the key to the continuity problem (c.f. Theorem 4.5).

The results of Theorem 4.1 are not intuitively immediate, nor are majorising measures easily understood. One thing is clear. For the integrals in (4.3) and (4.4) to be finite, it is necessary that m puts as much mass as possible on regions where the d-balls are small. Since small d-balls imply high incremental variance, these are the regions where the process has the most irregularities in its sample path, and so these are the regions where one would expect the supremum to be achieved. In other words, it is natural to expect that majorising measures are somehow related to the position of the supremum of X. However, this argument has not yet been made rigorous in general. (An obvious problem, of course, is there is no guarantee that in the general case the position of the supremum is uniquely defined.)

To see an example in which there is a strong relationship between majorising measures and the position of the supremum, take T finite and separated by d. Then there is, with probability one, a unique point $\tau(\omega) \in T$ such that $X_\tau(\omega) = \sup_{t\in T} X_t(\omega)$. Let μ_τ denote the law of τ. The following is true:

4.2 THEOREM. *With T, μ_τ, D as above, there exists a universal K such that*

$$(4.5) \qquad K^{-1}E\|X\| \;\leq\; D + \int_T \mu_\tau(dt) \int_0^\infty g\big(\mu_\tau(B(t,\epsilon))\big)\, d\epsilon \;\leq\; KE\|X\|.$$

Although we shall not prove this result, (the left-hand inequality can be deduced from results in Fernique (1976), the right-hand is due to Talagrand (1987), but neither are far removed from results that we *shall* prove)) it is worthwhile to try to understand it.

Note firstly that the finiteness of the double integral in (4.5) does not imply that μ_τ is a majorising measure. This would be true, however, if *a priori* we had removed from the finite space T all points of μ_τ-measure zero; i.e. points at which the supremum cannot be achieved. The proof of this is immediate once one thinks of the double integral as a finite double sum.

Another way of saying almost the same thing, after a fashion that at first looks as if it would generalise easily to countable parameter spaces and then perhaps to the most general separable ones, (but does not), is the following:

Choose $\alpha > 0$ and set

$$\widehat{T} = \widehat{T}(\alpha) = \Big\{t \in T\colon \int_0^\infty g\big(\mu_\tau(B(t,\epsilon))\big)\, d\epsilon \le \alpha K\, E\|X\|\Big\}.$$

By (4.5)

$$\begin{aligned} 1 - \mu_\tau(\widehat{T}) &= \mu_\tau(T \setminus \widehat{T}) \\ &= \int_{T\setminus\widehat{T}} \mu_\tau(dt) \\ &\le \int_{T\setminus\widehat{T}} \Big\{\frac{\int_0^\infty g\big(\mu_\tau(B(t,\epsilon))\big)\, d\epsilon}{\alpha K E\|X\|}\Big\}\, \mu_\tau(dt) \\ &\le \alpha^{-1}. \end{aligned}$$

Thus $\mu_\tau(\widehat{T}) \ge (1-\alpha^{-1})$, and on $\widehat{T}$ the measure μ_τ behaves like a majorising measure, with the unimportant restriction that it is not supported on $\widehat{T}$. We don't know very much about $\widehat{T}$, but, by taking α large, we see that $\widehat{T}$ contains most of the information about $\|X\|$, since

$$P\{\sup_{t\in\widehat{T}} X_t = \sup_{t\in T} X_t\} = \mu_\tau(\widehat{T}) \ge 1 - \alpha^{-1}.$$

Thus, from a heuristic viewpoint, Theorem 4.2 means that μ_τ is a majorising measure on a subset of T large enough to control $\sup_{t\in T} X_t$. ■

It would be interesting to extend Theorem 4.2 to the general case. David Pollard (1990) has recently shown that the supremum of a continuous Gaussian process is attained at a unique point in T. In terms of the results to follow, which will closely relate continuity to the integrals in (4.3) and (4.4), this seems to suggest that a result like Theorem 4.2 should hold in a setting more general than that of finite T. Nevertheless, this does not seem to be an easy extension to establish. ■

Although Theorem 4.2 covers only the case of finite T, it does have a partial extension to general parameter spaces. Take T general, and $\{T_n\}$ a sequence of finite subsets of T increasing to a countable dense subset. By separability and monotonicity

$$E\|X\|_T = \lim_{n\to\infty} E\|X\|_{T_n}.$$

But by Theorem 4.2 we have that, for each n,

$$K^{-1}E\|X\|_{T_n} \le D + \sup_{\mu\in\mathcal{P}(T)} \int_T \mu(dt) \int_0^\infty g\big(\mu(B(t,\epsilon))\big)\, d\epsilon,$$

where $\mathcal{P}(T)$ denotes the collection of all probability measures on T. The fact that the supremum is taken over all $\mathcal{P}(T)$ rather than $\mathcal{P}(T_n)$ only has the effect of making the upper bound larger.

By absorbing the factor of D into the integral – changing the constant if necessary and checking the argument around (4.30) if this bothers you – we have proven the upper bound in

4.3 THEOREM. *There exists a universal constant $K \in (1,\infty)$ such that for any centered Gaussian process X on (T,d)*

$$\begin{aligned}
K^{-1} \sup_{\mu \in \mathcal{P}(T)} & \int_T \mu(dt) \int_0^\infty g\big(\mu(B(t,\epsilon))\big)\, d\epsilon \\
&\le E\|X\| \\
&\qquad \le K \sup_{\mu \in \mathcal{P}(T)} \int_T \mu(dt) \int_0^\infty g\big(\mu(B(t,\epsilon))\big)\, d\epsilon.
\end{aligned}$$

The proof of the lower bound, which is the more interesting, is deferred until the following section.

There is one case in which the majorising measure for the lower bound (4.4) can be easily identified, so that Theorem 4.1 provides a useful necessary and sufficient condition for sample path boundedness. This is the case in which the process is stationary: i.e. T is an abelian group with an operation which shall denote as $+$ such that for all $s,t,\tau \in T$

$$E\,X(s)X(t) \;=\; E\,X(s+\tau)X(t+\tau).$$

4.4 THEOREM. *If X is stationary and T compact, then (4.3) and (4.4) hold with m taken to be normalised Haar measure on T. If T is a compact subset of an infinite group T', then (4.3) and (4.4) hold for the normalised restriction of Haar measure on T' to T.*

PROOF: The inequality (4.3) is true for any measure, and so for the Haar measures of the theorem. Thus we need only prove (4.4). For simplicity, we shall start with the case in which the full group T is compact.

Suppose there exists a majorising measure, m, satisfying (4.4) and define

$$D_m \;=\; \sup\big\{\eta : m\big(B(t,\eta)\big) \;<\; \tfrac{1}{2}, \text{ for all } t \in T\big\}.$$

Clearly

$$\int_0^{D_m} g\big(m(B(t,\epsilon)\big)\, d\epsilon \;\le\; KE\|X\|,$$

for all $t \in T$. Let μ denote normalised Haar measure on T, and let τ be a random variable with distribution μ; i.e. τ is uniform on T. For each $\epsilon > 0$, set $Z(\epsilon) = m(B(\tau, \epsilon))$. Then, for any $t_o \in T$

$$\begin{aligned} EZ(\epsilon) &= \int_T m\big(B(t,\epsilon)\big)\,\mu(dt) \\ &= \int_T m\big(t + B(t_o,\epsilon)\big)\,\mu(dt) \\ &= \int_T \mu\big(t + B(t_o,\epsilon)\big)\,m(dt) \\ &= \mu\big(B(t_o,\epsilon)\big), \end{aligned}$$

where the second equality comes from the stationarity of X and the third and fourth from the properties of Haar measures.

Now note that $g(x)$ is convex over $x \in (0, \frac{1}{2})$, so that it is possible to define a function $\widehat{g}$ that agrees with g on $(0, \frac{1}{2})$, is bounded on $(\frac{1}{2}, \infty)$, and convex on all of $\Re_+$. By Jensen's inequality,

$$\widehat{g}\big(EZ(\epsilon)\big) \le E\widehat{g}\big(Z(\epsilon)\big).$$

That is,

$$\widehat{g}\big(\mu\big(B(t_o,\epsilon)\big)\big) \le \int_T \widehat{g}\big(m(B(t,\epsilon)\big)\,\mu(dt).$$

With D_μ defined for μ as D_m was for m, set $\Lambda = \min(D_m, D_\mu)$. Then

$$\begin{aligned} \int_0^\Lambda \widehat{g}\big(\mu(B(t_o,\epsilon)\big)\,d\epsilon &\le \int_0^\Lambda d\epsilon \int_T \widehat{g}\big(m(B(t,\epsilon)\big)\,\mu(dt) \\ &= \int_T \int_0^\Lambda \mu(dt)\widehat{g}\big(m(B(t,\epsilon)\big)\,d\epsilon \\ &\le KE\|X\|. \end{aligned}$$

This is the crux of (4.4). The final stage of the proof, for this case, is left to you.

The case of T a compact subset of an infinite group T' is handled similarly. Let μ_T be the restriction of Haar measure on T' to T. For convenience, assume that $\mu_T(T) = 1$, so that we need not worry about the normalisation. Then argue as before, but replacing (4.6) by

$$\begin{aligned} EZ(\epsilon) &= \int_T m\big(B(t,\epsilon)\big)\,\mu_T(dt) \\ &\le \int_{T'} m\big(t + B(t_o,\epsilon)\big)\,\mu(dt) \\ &= \int_{T'} \mu\big(t + B(t_o,\epsilon)\big)\,m(dt) \\ &= \mu\big(B(t_o,\epsilon)\big). \end{aligned} \tag{4.6}$$

Now continue through the rest of the proof as for the compact case, making certain that the inequalities always go the right way. (Remember that g, and so $\widehat{g}$, are decreasing functions.) ■

The time has come to look at some applications of the above results. To start, we shall combine Theorem 4.1 with Theorem 3.3 to obtain a simple, applicable, condition for continuity.

For a probability measure m on (T, d) write

$$\gamma_m(\eta) \;=\; \sup_{t \in T} \int_0^\eta g\big(m(B(t,\epsilon))\big)\, d\epsilon.$$

4.5 THEOREM. *An centered Gaussian process on T is a.s. bounded and uniformly continuous if, and only if, there exists a probability measure m on (T, d) such that*

$$\lim_{\eta \to 0} \gamma_m(\eta) \;=\; 0. \tag{4.7}$$

REMARK: Of course you remember that implicit in this result, as throughout these notes, is the assumption that T is totally bounded. There is, in fact, a somewhat stronger version of Theorem 4.5, which states that X will be a.s. bounded and uniformly continuous on an arbitrary metric space (T, d) if, *and only if*, T is totally bounded in d and (4.7) holds. (Talagrand (1987), Theorem 4.5.) It is this extended result that permitted the claim, at the beginning of Chapter 1, that we lost no interesting cases by restricting ourselves to totally bounded parameter spaces from the start.

HALF PROOF: We shall prove only the sufficiency of (4.7). The necessity is somewhat harder.

As a first step we need to show that (4.7) implies that X is a.s. bounded. Take a $\eta > 0$ such that $\gamma(\eta) < \infty$. Then, by Theorem 4.1, X is a.s. bounded on each ball of radius η. Since T is totally bounded, it can be covered by finitely many such balls. Thus X is bounded on all of T.

To handle continuity, by Theorem 3.3 we need only show $\lim_{\eta \to 0} \phi_d(\eta) = 0$, where

$$\phi_d(\eta) \;=\; E \sup_{d(s,t) < \eta} (X_s - X_t). \tag{4.8}$$

The idea of the proof is to bound $\phi_d(\eta)$ via an integral involving a majorising measure, and to then derive convergence to zero from (4.7).

Let $U = \{(s,t) \in T \times T : d(s,t) \le \eta\}$. Provide $T \times T$ and its subspace U with the distance d' given by

$$d'\big((s,t),(s',t')\big) \;=\; \big(E[(X_s - X_t) - (X_{s'} - X_{t'})]^2\big)^{\frac{1}{2}}.$$

Note that $X_s - X_t$ is a centered Gaussian process on $T \times T$ with canonical metric d'. Theorem 4.1 should apply to this two-parameter process.

We start with some observations on the difference process. Firstly, note that for $(s,t) \in T \times T$ we have

$$B\big((s,t),\epsilon\big) \supset B(s,\epsilon/2) \times B(t,\epsilon/2), \tag{4.9}$$

where the first ball is a d'-ball and the last two d-balls.

Next, for $x = (s,t) \in T \times T$, let $\phi(x) \in U$ satisfy

$$d'\big(x,\phi(x)\big) = d'\big(x,U\big) = \inf_{y\in U} d'\big(x,y\big).$$

(U is closed, so the infimum is achieved.) Note that if $x \in U$ then we can, and shall, take $\phi(x) = x$. Furthermore, if $y \in U$ and $x \in T \times T$ then

$$\begin{aligned} d'\big(y,\phi(x)\big) &\le d'\big(y,x\big) + d'\big(x,\phi(x)\big) \\ &= d'\big(y,x\big) + d'\big(x,U\big) \\ &\le 2d'\big(x,y\big). \end{aligned}$$

Consequently, for all $y \in U$,

$$\phi\big(B_{d'}(y,\epsilon)\big) \subseteq U \cap B_{d'}\big(y,2\epsilon\big). \tag{4.10}$$

Now define a probability measure μ on U by setting $\mu(A) = (m \otimes m) \circ \phi^{-1}(A)$. In view of (4.9) and (4.10) we have, for all $(s,t) \in U$,

$$\begin{aligned} m\big(B(s,\tfrac{1}{2}\epsilon)\big)\, m\big(B(t,\tfrac{1}{2}\epsilon)\big) &= (m\otimes m)\,\big(B(s,\tfrac{1}{2}\epsilon) \times B(t,\tfrac{1}{2}\epsilon)\big) \\ &\le (m\otimes m)\,\big(B_{d'}((s,t),\epsilon)\big) \\ &\le (m\otimes m)\big(\phi^{-1}\big(U \cap (B_{d'}((s,t),2\epsilon)\big)\big) \\ &= \mu\,\big(U \cap (B_{d'}((s,t),2\epsilon)\big). \end{aligned}$$

It thus follows from Theorem 4.1 that for any measure m on T and any $\eta > 0$

$$\begin{aligned} \phi_d(\eta) &\le K \sup_{(s,t)\in U} \int_0^{2\eta} g\big(\mu\big(B_{d'}((s,t),\epsilon)\big)\big)\, d\epsilon \\ &\le K \sup_{(s,t)\in T\times T} \int_0^{2\eta} g\big(m\big(B_{d'}(s,\epsilon/4)\big)\, m\big(B_{d'}(t,\epsilon/4)\big)\big)\, d\epsilon \\ &\le 8K \sup_{t\in T} \int_0^{\eta/2} g\big(m\big(B(t,\epsilon)\big)\big)\, d\epsilon \\ &= 8K\gamma_m(\eta/2). \end{aligned}$$

This proves the theorem. ■

For necessity, and some further details related to this result, see Talagrand (1987).

As one might expect, majorising measures also feature in providing moduli of continuity, and the following result is the improvement on Corollary 3.4 promised in the previous chapter.

4.6 THEOREM. *Let X be a.s. bounded on T and τ be a metric on T such that the canonical metric d is τ-uniformly continuous. Let $\phi_\tau(\eta)$ be as at (4.8) and denote the τ-modulus of (uniform) continuity of X by*

$$W_\tau(\eta) = \sup_{\tau(s,t)<\eta} |X_s - X_t|.$$

If $\lim_{\eta\to 0} \phi_\tau(\eta) = 0$ then there exists an a.s. finite random variable $\delta = \delta(\omega)$ such that, for almost all ω,

$$W_\tau(\eta) \le \phi_\tau(\eta)$$

for all $\eta \le \delta(\omega)$. That is, $\phi_\tau(\cdot)$ is a uniform sample modulus for X in the metric τ.

This is, of course, Corollary 3.4 without the ϵ. An immediate corollary, in terms of the canonical metric d, is

4.7 COROLLARY. *Let m be a majorising measure for X, a.s. bounded on T. Then there exists a universal constant $K < \infty$ and an a.s. finite random variable $\delta(\omega)$ such that, for almost all ω,*

$$\sup_{s,t:\ d(s,t)<\eta} |X_t - X_s| \le K\gamma_m(\eta).$$

for all $\eta \le \delta(\omega)$.

PROOF OF THEOREM 4.6: Note firstly that if T is finite (in the sense that the number of pairs $s,t \in T$ for which $d(s,t) \neq 0$ is finite) the result is trivially true, by taking $\delta(\omega) = (1-\epsilon)\min\{\tau(s,t) \colon d(s,t) > 0\}$ for any $\epsilon > 0$. Thus assume that T is infinite.

As noted after the statement of Corollary 3.4, all we need to show is that

$$\lim_{\eta\to 0} \frac{\phi_\tau(\eta)}{d_\tau(\eta)} = \infty, \tag{4.11}$$

where $d_\tau(\eta) = \sup_{\tau(s,t)\le\eta} d(s,t)$.

By (4.4) and the argument applied in the proof of Theorem 4.5 there exists a probability measure on T such that a lower bound to the ratio in (4.11) is given by

$$\begin{aligned} &\frac{K \sup_{t\in T} \int_0^{d_\tau(\eta)} g\big(m(B(t,\epsilon))\big)\, d\epsilon}{d_\tau(\eta)} \\ &\ge K \sup_{t\in T} g\big(m(B(t, d_\tau(\eta)))\big) \\ &= K \sup_{t\in T} \sqrt{\log\left(\frac{1}{m\big(B(t,d_\tau(\eta))\big)}\right)}. \end{aligned}$$

Since d is τ-continuous, $d_\tau(\eta) \to 0$ as $\eta \to 0$, and so to complete the proof it suffices to show that there exists a $t \in T$ for which

$$m(B(t,\eta)) \to 0 \qquad \text{as } \eta \to 0. \tag{4.12}$$

Suppose that this is not true. That is, there exists a $\delta > 0$ such that for all $t \in T$ $\lim_{\eta \to 0} m(B(t,\eta)) \geq \delta$. It then follows that $m(\{t\}) \geq \delta$ for all $t \in T$. Since m is a probability measure this can only happen if T has at most δ^{-1} points. But T was assumed infinite. Thus (4.12) does hold, and the proof is complete. ∎

We now turn to the proof of Theorem 4.1. This is given in two parts, the upper bound (4.3) first, and then the lower bound (4.4). The upper bound proof is (comparatively) easy, the lower bound proof not so. I heartily recommend skipping both if this is your first reading of this chapter, and coming back to them only after you have read Section 4.4. It is only my childhood training as a mathematician and consequent feelings of guilt at skipping proofs that forces me to include them now and not later.

2. Upper Bound Proof.

The proof of (4.3) that we shall give is closely modelled on Fernique (1978), with some further input from Anderson *et. al.* (1988).

We start with an elementary observation, that we shall use routinely in what follows. The proof is left to you.

4.8 OBSERVATION. *If $f(t)$ is a positive decreasing function then*

$$\sum_{n=1}^{\infty} 2^{-n-1} f(2^{-n}) \leq \int_0^1 f(\epsilon)\, d\epsilon \leq \sum_{n=1}^{\infty} 2^{-n} f(2^{-n}). \tag{4.13}$$

We shall assume throughout this and the following section that $\operatorname{diam}(V) \leq 1$ so as to simplify some of the notational aspects of the proofs. To see that no loss of generality is involved through this assumption, let $D = \operatorname{diam}(V)$ and set $Y(t) = X(t)/D$. Note that balls in the canonical metrics of X and Y are then related by the fact that

$$B_X(t,\epsilon) = B_Y(t,\epsilon/D).$$

Suppose we succeed in proving (4.3) for the process Y. Then,

$$\begin{aligned}\frac{E\|X\|}{D} &= E\|Y\| \\ &\le K \sup_t \int_0^\infty g\big(m\big(B_Y(t,\epsilon)\big)\big)\,d\epsilon \\ &= K \sup_t \int_0^\infty g\big(m\big(B_X(t,D\epsilon)\big)\big)\,d\epsilon \\ &= \frac{K}{D} \sup_t \int_0^\infty g\big(m\big(B_X(t,\epsilon)\big)\big)\,d\epsilon.\end{aligned}$$

Thus (4.3) holds also for $\|X\|$, with the same constant K.

We start with an arbitrary probability measure μ and some geometry. Take $n \ge 1$ and let $\{t_{ni}\}_{i=1}^{r_n} \subset T$ be a finite set of distinct points of T such that

$$\sup_{t\in T} \inf_i d(t,t_{ni}) \le 2^{-n-3}.$$

By reordering, if necessary, we can assume that

$$\mu\big(B(t_{ni},2^{-n-2})\big) \ge \mu\big(B(t_{n(i+1)},2^{-n-2})\big), \qquad \text{for all } i \ge 1. \tag{4.14}$$

Define subsets $\{C_{ni}\}_{i=1}^{r_n}$ inductively as

$$C_{n1} = B\big(t_{n1},2^{-n-2}\big) \tag{4.15}$$

$$C_{ni} = \begin{cases} \emptyset, & \text{if } B(t_{ni},2^{-n-2}) \cap \bigcup_{j=1}^{i-1} C_{nj} \ne \emptyset, \\ B(t_{ni},2^{-n-2}), & \text{otherwise.} \end{cases} \tag{4.16}$$

Next, define a mapping

$$\pi_n : T \to \{t_{ni}\}_{i=1}^{r_n} \tag{4.17}$$

as follows:

Set $\pi_n(t)$ to be the first t_{ni} for which $d(t,t_{ni}) \le 2^{-n-3}$ and $C_{ni} \ne \emptyset$. If there is no such i such that the second of these conditions is satisfied, then choose the first t_{ni} such that $d(t,t_{ni}) \le 2^{-n-3}$, and note that since $C_{ni} = \emptyset$ there exists a maximal $j < i$ such that $C_{nj} \cap B(t_{ni},2^{-n-2}) \ne \emptyset$. Define $\pi_n(t)$ to be the corresponding t_{ni}. We denote the n-th such collection of $\{t_{ni}\}$ by $\mathcal{T}_n$, and set $\mathcal{T} = \bigcup_n \mathcal{T}_n$.

The first basic property of this construction that we shall use is that:

$$\begin{aligned} d(t,\pi_n(t)) &\le d(t,t_{ni}) + d(t_{ni},t_{nj}) \\ &\le 2^{-n-3} + 2\cdot 2^{-n-2} \\ &< 2^{-n}. \end{aligned} \tag{4.18}$$

Now note that since for each $t \in T$ and $n \ge 1$ there is a unique $C_{ni} := C_n(t)$, such that $\pi_n(t) = t_{ni} \in C_n(t)$ we can write, with some abuse of notation,

$$\mu_n(t) = \mu(C_n(t)).$$

This leads us to the second property of the construction: If t_{ni} is such that $d(t,t_{ni}) \le 2^{-n-3}$ and $C_{ni} \neq \emptyset$, then

$$\begin{aligned} \mu_n(t) &= \mu(B(t_{ni},2^{-n-2})) && \text{(by (4.15)–(4.16))} \\ &\ge \mu(B(t,2^{-n-3})). \end{aligned}$$

On the other hand, if t_{ni} is such that $d(t,t_{ni}) \le 2^{-n-3}$ and $C_{ni} = \emptyset$, then there is a $j < i$ such that $C_{nj} \cap B(t_{ni},2^{-n-2}) \neq \emptyset$, and in this case

$$\begin{aligned} \mu(B(t,2^{-n-3})) &\le \mu(B(t_{ni},2^{-n-2})) \\ &\le \mu(B(t_{nj},2^{-n-2})) && \text{(by (4.14))} \\ &= \mu_n(t). \end{aligned}$$

The above two inequalities yield that in all cases

$$\mu_n(t) \ge \mu(B(t,2^{-n-3})). \tag{4.19}$$

This is all the geometry that we shall require for the moment. We now build ourselves a new, somewhat simpler process than the original. We start with $\{\xi_t\}_{t\in\mathcal{T}}$, a collection of i.i.d. standard Gaussian variables on $\mathcal{T}$. (If a $t \in \mathcal{T}$ appears in more than one $\mathcal{T}_n$ choose a different ξ_t for each n, and change the following proof, where necessary, to account for this irritating phenomenon. We shall assume, however, that this never happens.) Define a new Gaussian process Y on T by

$$Y(t) = \sum_{n=1}^{\infty} 2^{-n}\,\xi(\pi_n(t)). \tag{4.20}$$

Note that although Y generates its own canonical metric on T, we shall treat it as defined on T equipped with the canonical metric of X. The following step in the proof is important enough to be singled out as an independent result.

4.9 LEMMA. *If Y is defined as at (4.20) then, for all $s,t \in T$,*

$$E(X_s - X_t)^2 \le 6E(Y_s - Y_t)^2.$$

PROOF: For $s,t \in T$, choose $N \ge 1$ such that

$$2^{-N} < d(s,t) = \left(E(X_t - X_s)^2\right)^{\frac{1}{2}} \le 2^{-N+1}. \tag{4.21}$$

By (4.18), it thus follows that for $n \ge N+1$, $\pi_n(t) \ne \pi_n(s)$. (For $n < N+1$ it is impossible to say whether $\pi_n(t) = \pi_n(s)$ or not. This is what makes it impossible to use this argument to obtain a proof of the lower bound as well.) Using the representation (4.20) of Y we thus deduce that

$$\begin{aligned} E(Y_t - Y_s)^2 &\ge \sum_{n=N+1}^{\infty} 2^{-2n} E\left[\xi(\pi_n(t)) - \xi(\pi_n(s))\right]^2 \\ &= 2\sum_{n=N+1}^{\infty} 2^{-2n} \\ &= 2^{-2N+2}/6. \end{aligned}$$

Comparing this to the upper bound in (4.21) proves the lemma. ∎

It now follows from Theorem 2.8 or 2.9 (the two forms of the Sudakov-Fernique inequality) that $E\|X\| \le 2\sqrt{6}\,E\|Y\|$, and so it suffices to prove (4.3) for the process Y.

To do this, let $\tau\colon \Omega \to T$ be a random point in T. If we can show that there exists an $M > 0$ such that for *every* such random variable τ

$$EY_\tau \le M, \tag{4.22}$$

then it must follow (by contradiction) that

$$E\|Y\| = E\sup_{t\in T} Y_t \le M. \tag{4.23}$$

Thus, consider (4.22) and let the distribution of τ on T be given by a probability measure ν. Since each T_n is finite, (4.20) gives us that

$$EY_\tau = \sum_{n=1}^{\infty} 2^{-n} \sum_{t\in T_n} E\{\xi_t I_{\{\pi_n(\tau)=t\}}\}. \tag{4.24}$$

Note the easy fact that for ξ standard normal

$$\begin{aligned} E\{\xi I_{\{\xi>\sqrt{2g(a)}\}}\} &= (2\pi)^{-\frac{1}{2}} \int_{\sqrt{2g(a)}}^{\infty} x e^{-\frac{1}{2}x^2}\,dx \\ &= (2\pi)^{-\frac{1}{2}} \left[-e^{-x^2/2}\right]_{\sqrt{2\log(1/a)}}^{\infty} \\ &= (2\pi)^{-\frac{1}{2}}\, a. \end{aligned} \tag{4.25}$$

Now note that each expectation in the summation in (4.24) can be written as

$$
\begin{aligned}
&E\{\xi_t I_{\{\pi_n(\tau)=t\}} I_{\{\xi_t>\sqrt{2}g(\mu_n(t))\}}\} + E\{\xi_t I_{\{\pi_n(\tau)=t\}} I_{\{\xi_t\le\sqrt{2}g(\mu_n(t))\}}\} \\
(4.26)\qquad &\le E\{\xi_t I_{\{\xi_t>\sqrt{2}g(\mu_n(t))\}}\} + \sqrt{2}g(\mu_n(t))E\{I_{\{\pi_n(\tau)=t\}}\} \\
&= (2\pi)^{-\frac{1}{2}}\mu_n(t) + \sqrt{2}g(\mu_n(t))\,P\{\pi_n(\tau)=t\},
\end{aligned}
$$

the last line following from (4.25) and the definition of μ.

Thus, substituting into (4.24), and applying (4.19) to obtain the third inequality below, we obtain

$$
\begin{aligned}
EY_\tau &\le \tfrac{1}{\sqrt{2\pi}} + \sum_{n=1}^{\infty} 2^{-n+\frac{1}{2}} \sum_{t\in T_n} g(\mu_n(t))\,P\{\pi_n(\tau)=t\} \\
(4.27)\qquad &\le \tfrac{1}{\sqrt{2\pi}} + \sum_{n=1}^{\infty} 2^{-n+\frac{1}{2}} \int_T g(\mu(C_n(t)))\,\nu(dt) \\
&\le \tfrac{1}{\sqrt{2\pi}} + \sum_{n=1}^{\infty} 2^{-n+\frac{1}{2}} \int_T g(\mu(B(t,2^{-n-3})))\,\nu(dt) \\
&\le \tfrac{1}{\sqrt{2\pi}} + \int_T \nu(dt) \sum_{n=1}^{\infty} 2^{-n+\frac{1}{2}} g(\mu(B(t,2^{-n-3}))).
\end{aligned}
$$

Replacing the sum over n by an integral over ϵ (c.f. Observation 4.8) gives us that there exists a constant K such that

$$
K^{-1}EY_\tau \le 1 + \int_T \nu(dt) \int_0^\infty g(\mu(B(t,\epsilon)))\,d\epsilon,
$$

and so, allowing K to change from line to line,

$$
\begin{aligned}
E\|X\| &\le K\sup_\tau EY_\tau \\
&= K\Big(1 + \sup_\nu \int_T \nu(dt) \int_0^\infty g(\mu(B(t,\epsilon)))\,d\epsilon\Big) \\
&\le K\Big(1 + \sup_t \int_0^\infty g(\mu(B(t,\epsilon)))\,d\epsilon\Big),
\end{aligned}
$$

the last inequality coming from the fact that ν is a probability measure.

I claim that $\sup_t \int_0^\infty g(\mu(B(t,\epsilon)))\,d\epsilon$ can be bounded from below by $\sqrt{\log 2}$. To see this, note that T has at least two distinct points (since we have assumed that $\mathrm{diam}(T) > 0$). Thus there must be at least one point $t \in T$ with $\mu(B(t,\epsilon)) \leq \frac{1}{2}$ for all $\epsilon < 1$. For this t

$$\int_0^\infty g\big(\mu\big(B(t,\epsilon)\big)\big)\,d\epsilon \;\geq\; \int_0^1 \sqrt{\log 2}\,d\epsilon \;=\; \sqrt{\log 2}, \tag{4.28}$$

as claimed.

Thus the 1 in the above bound can be absorbed into the integral, albeit with a slightly larger constant, and we are done. ■

Although this section is advertised as being devoted to upper bounds, while we have the appropriate notation at hand we shall also give the

PROOF OF THE LOWER BOUND OF THEOREM 4.3: We shall need two facts. The first, as you can verify with a little calculus, is that the function g satisfies

$$x\,g(x) \;\leq\; x\,g(y) + \frac{y}{\sqrt{2e}}, \tag{4.29}$$

for all x and y. The second is the rather useful inequality that

$$\sup_{\nu \in \mathcal{P}(T)} \int_T \nu(dt) \int_0^\infty g\big(\nu\big(B(t,\epsilon)\big)\big)\,d\epsilon \;\geq\; \sqrt{\log 2}, \tag{4.30}$$

which you can verify by choosing ν to be concentrated on two points $s, t \in T$ with $d(s,t) > 0$ and $\nu(\{s\}) = \nu(\{t\}) = \frac{1}{2}$. (This is the second, and last, time we need the assumption that T has non-zero diameter.)

In what follows, given a measure $\nu \in \mathcal{P}(T)$ we write τ_ν to denote a T-valued random variable with distribution ν. We shall use the notation of the previous proof without comment other than to point out now that when notation from there appears with a suffix m this is to indicate that the construction of the proof relates to the specific measure m. By (4.4), a proof of which is in the following section, there exists a universal K and a measure m such that

$$\begin{aligned} K\,E\|X\| \;&\geq\; \sup_{t\in T}\int_0^\infty g\big(m\big(B(t,\epsilon)\big)\big)\,d\epsilon \\ &=\; \sup_{\nu\in\mathcal{P}(T)} \int_T \nu(dt) \int_0^\infty g\big(m\big(B(t,\epsilon)\big)\big)\,d\epsilon \;-\; \eta, \end{aligned}$$

for any $\eta > 0$. To prove the above inequality simply take ν concentrated at a point t_o for which

$$\int_0^\infty g\big(m\big(B(t_o,\epsilon)\big)\big)\,d\epsilon \;\geq\; \sup_{t\in T}\int_0^\infty g\big(m\big(B(t,\epsilon)\big)\big)\,d\epsilon \;-\; \eta.$$

Thus

$$
\begin{aligned}
K\,E\|X\| + \eta &\geq \sup_{\nu \in \mathcal{P}(T)} \int_T \nu(dt) \sum_{n=1}^{\infty} 2^{-n-4} g\big(m\big(B(t, 2^{-n-3})\big)\big) \\
&\geq \sup_{\nu \in \mathcal{P}(T)} \int_T \nu(dt) \sum_{n=1}^{\infty} 2^{-n-4} g\big(m\big(C_n^m(t)\big)\big) \\
&= \sup_{\nu \in \mathcal{P}(T)} \sum_{n=1}^{\infty} \sum_{t \in T_n} 2^{-n-4} g\big(m\big(C_n^m(t)\big)\big)\, P\{\pi_n^m(\tau_\nu) = t\}.
\end{aligned}
$$

By (4.29) this is bounded below by

$$
\sup_{\nu \in \mathcal{P}(T)} \sum_{n=1}^{\infty} \sum_{t \in T_n} 2^{-n-4} \left[g\big(P\{\pi_n^m(\tau_\nu) = t\}\big)\, P\{\pi_n^m(\tau_\nu) = t\} - \frac{m\big(C_n^m(t)\big)}{\sqrt{2e}} \right]
$$

which, by (4.18), is in turn bounded below by

$$
\sup_{\nu \in \mathcal{P}(T)} \sum_{n=1}^{\infty} \sum_{t \in T_n} 2^{-n-4} g\big(\nu\big(B(t, 2^{-n})\big)\big)\, P\{\pi_n^m(\tau_\nu) = t\} - (16\sqrt{2e})^{-1}.
$$

The sum is equal to

$$
\begin{aligned}
&\sup_{\nu \in \mathcal{P}(T)} \sum_{n=1}^{\infty} 2^{-n-4} E\Big\{ g\big(\nu\big(B(\pi_n^m(\tau_\nu), 2^{-n})\big)\big) \Big\} \\
&\geq \sup_{\nu \in \mathcal{P}(T)} \sum_{n=1}^{\infty} 2^{-n-4} E\Big\{ g\big(\nu\big(B(\tau_\nu, 2^{-n+1})\big)\big) \Big\} \\
&\geq \sup_{\nu \in \mathcal{P}(T)} \frac{1}{16} \int_0^{\infty} E\Big\{ g\big(\nu\big(B(\tau_\nu, 2\epsilon)\big)\big) \Big\}\, d\epsilon \\
&\geq \sup_{\nu \in \mathcal{P}(T)} \frac{1}{32} \int_T \nu(dt) \int_0^{\infty} g\big(\nu\big(B(t, \epsilon)\big)\big)\, d\epsilon.
\end{aligned}
$$

Now apply (4.30) to find that, modulo a multiplicative constant, $E\|X\|$ is bounded below by

$$
\sup_{\nu \in \mathcal{P}(T)} \int_T \nu(dt) \int_0^{\infty} g\big(\nu\big(B(t, \epsilon)\big)\big)\, d\epsilon,
$$

which is precisely what we had to prove. ∎

3. Lower Bound Proof.

The time has now come to tackle the hardest proof in these notes, that of establishing the lower bound in Theorem 4.1.

In principle, the idea behind the proof is easy. Recall that the proof of the upper bound was based on comparing a given process X with a simpler process Y which dominated it, in the sense that $E(X_t - X_s)^2 \leq \text{const}\, E(Y_t - Y_s)^2$ for all $s, t \in T$. Since the structure of Y was simpler, it was possible to calculate a good upper bound for $E\|Y\|$, and so, by the Sudakov-Fernique inequality, obtain an upper bound for $E\|X\|$ as well.

We shall follow this path once again. What makes it somewhat harder to traverse this time, however, is that even after we have found a simple comparison process, the lower bound calculation is difficult. *C'est la vie.* What is somewhat more disappointing is that even after we shall have worked so hard to prove the result, we shall not have discovered how to built majorising measures. The proof is not constructive.

Before we start, we require the following standard result.

4.10 LEMMA. *Let $X_1, \ldots, X_N$ be a sequence of i.i.d. standard Gaussian variables. There exists a universal $K > 0$ such that*

$$E \sup_{1 \leq i \leq N} X_i \geq K\sqrt{\ln N}. \tag{4.31}$$

PROOF: Take $Z_1, \ldots, Z_N$ to be a sequence of i.i.d. random variables on $[1, \infty]$ with probability density $f(z) = z^{-2}$ and distribution function $F(z) = 1 - z^{-1}$.

The probability density of the $\min(Z_1, \ldots, Z_N)$ is thus given by

$$N f(z)(1 - F(z))^{N-1} = N z^{-N-1},$$

so that

$$E \min(Z_1, \ldots, Z_N) = \frac{N}{N-1}.$$

Note that $Z_i \stackrel{\mathcal{L}}{=} 1/\Psi(X_i)$, where Ψ, as usual, denotes the standard normal distribution function. Thus

$$\begin{aligned} \frac{N}{N-1} &= E \min(Z_1, \ldots, Z_N) \\ &= E \min\left\{\frac{1}{\Psi(X_1)}, \ldots, \frac{1}{\Psi(X_N)}\right\} \\ &= E\left(\Psi(\max(X_1, \ldots, X_N))\right)^{-1}. \end{aligned} \tag{4.32}$$

But $\Psi^{-1}(x)$ is a convex function, since

$$\begin{aligned}\frac{d^2\Psi^{-1}(x)}{dx^2} &= \frac{d}{dx}\Big(-\phi(x)\Psi^{-2}(x)\Big)\\ &= \phi(x)\Psi^{-3}(x)\,\big(2\phi(x) + x\Psi(x)\big),\end{aligned}$$

and the last expression is clearly positive for $x > 0$ and also positive for $x < 0$ since it is zero at $-\infty$ and increasing thereafter.

Thus Jensen's inequality and (4.32) give

$$\Psi\Big(E \max_{1\le i\le N} X_i\Big) \ \ge\ \frac{N-1}{N}.$$

Apply the inequalities of (2.1) to complete the proof. ∎

We now require some rather formidable notation, in which A will always be a subset of T, m a measure on T and $\mathcal{P}(A)$ the set of probability measures on A.

$$\begin{aligned}\gamma_m(A) &= \gamma(m,A) = \sup_{t\in A}\int_0^\infty g\big(m(B(t,\epsilon))\big)\,d\epsilon\\ \gamma(A) &= \inf_{m\in\mathcal{P}(A)}\gamma_m(A)\\ \alpha(A) &= \sup_{B\subseteq A}\gamma(B)\\ \alpha &= \sup\big\{\alpha(V) : \emptyset \ne V\subseteq T,\ V \text{ finite}\big\}.\end{aligned}$$

We have to establish the existence of a $m\in\mathcal{P}(T)$ such that for some universal constant K

$$\gamma_m(T) \ \le\ K\,E\|X\|_T\,.$$

Our first step towards this will be a little more modest.

4.11 LEMMA. *There exists a universal constant K such that for every non-empty, finite, $V\subseteq T$*

$$\alpha(V) \ \le\ K\,E\|X\|_V\,. \tag{4.33}$$

You should think of this lemma as an approximate version of our theorem. It is going to be easier to prove than the general result because of the fact that $\alpha(V)$ is determined by what X does on a finite subset of T. Before we prove this lemma, however, we shall see that it is in fact equivalent to the full result.

PROOF OF THE LOWER BOUND (4.3): Since (4.33) implies that $\alpha \le KE\|X\|_T$, we need only show the existence of a $m\in\mathcal{P}(T)$ for which $\gamma_m(T)\le$

$K\alpha$. This we shall do by a reasonably straightforward subsequencing argument; i.e. We shall use the definitions of the α's and γ's on finite subsets, where infima and suprema are achieved, to show that appropriate m's can be constructed there, and then take a weak limit of these as our subspaces grow towards the whole of T.

For each $n \geq 1$ let $T_n = \{t_{ki}\}_{i=1}^{r_n}$ be a finite subset of T such that

$$\sup_{t \in T} \inf_{1 \leq i \leq r_n} d(t, t_{ki}) \leq 2^{-n-4}.$$

Now we choose one of these sets, which we denote by T_k. (Be careful in following the subscripts in what follows. T_n and T_k are chosen from one large family of sets, but whereas T_k is (temporarily) fixed, T_n will vary.)

Since $\gamma(T_k) \leq \alpha(T_k) \leq \alpha$, there exists a $m_k \in \mathcal{P}(T_k)$ such that, by Observation 4.8,

$$\begin{aligned} \sup_{t \in T_k} \sum_{i=1}^{\infty} 2^{-i} g\big(m_k\big(B(t, 2^{-i})\big)\big) &\leq 2 \sup_{t \in T_k} \int_0^1 g\big(m_k\big(B(t, \epsilon)\big)\big)\, d\epsilon \\ &\leq 2\alpha. \end{aligned} \tag{4.34}$$

Consider m_k as a measure on T, by setting $m_k(A) = m_k(A \cap T_k)$ for $A \subset T$. Now follow the same construction as used in the proof of the upper bound – viz. (4.14)-(4.17) – to obtain a family of subsets $\{C_{ni}^k\}_{i=1}^{r(k,n)}$ of T and a mapping $\pi_n^k : T \to T_k$ satisfying (4.18) and such that for each $t \in T$ the unique $C_n^k(t) = C_{ni}^k$ with center $\pi_n^k(t) = t_{ni}^k$ satisfies

$$m_k\big(C_n^k(t)\big) \leq m_k\big(B(t, 2^{-n-3})\big). \tag{4.35}$$

(As before, we must assume that the $\{t_{ni}^k\}$ have been reordered to satisfy (4.14) for $\mu = m_k$, but, as before, this involves no loss of generality.)

Define $\mu_n^k \in \mathcal{P}(T_n)$ by

$$\mu_n^k\big(t_{ni}^k\big) = \frac{m_k\big(C_{ni}^k\big)}{\sum_{i=1}^{r_k} m_k\big(C_{ni}^k\big)}.$$

Now switch the sequencing, and hold n fixed while sending $k \to \infty$. Since n is fixed, T_n is finite, and so a standard argument shows that there is a subsequence $\{\mu_n^{k'}\}_{k'}$ of the above measures for which $\lim_{k' \to \infty} \mu_n^{k'} = \mu_n$ exists (pointwise) and is an element of $\mathcal{P}(T_n)$.

We are almost done. Fix $t \in T$ and choose a further subsequence such that $\lim_{k'' \to \infty} \pi_n^{k''}(t) = \pi_n(t)$ exists. Note from the finiteness of T_n that for large enough k'' we must have that $\pi_n^{k''}(t) \equiv \pi_n(t)$. Let $\tau_k(t)$ be the first t_{ki} such that $d(t,t_{ki}) \leq 2^{-k-4}$. Then, by (4.33), (4.34) and Fatou's lemma

$$\begin{aligned}
\sum_{i=1}^{\infty} 2^{-n}\, g\big(\mu_n(\pi_n(t))\big) &\leq \sum_{i=1}^{\infty} 2^{-n} \liminf_{k''\to\infty} g\Big(\frac{m_{k''}(C_{ni}^{k''}(t))}{\sum_{i=1}^{r_k} m_k(C_{ni}^{k})}\Big) \\
&\leq \sum_{i=1}^{\infty} 2^{-n} \liminf_{k''\to\infty} g\big(m_{k''}(C_n^{k''}(t))\big) \\
&\leq 16 \sum_{i=1}^{\infty} 2^{-n-4} \liminf_{k''\to\infty} g\big(m_{k''}(B(\tau_{k''}(t),2^{-n-4}))\big) \\
&\leq 32\,\alpha.
\end{aligned}$$

Finally, define $m \in \mathcal{P}(T)$ by $m(A) = \sum_{i=1}^{\infty} 2^{-n}\mu_n(A)$, Note that

$$\begin{aligned}
g\big(m(\{\pi_n(t)\})\big) &\leq g\big(2^{-n}\mu_n(\{\pi_n(t)\})\big) \\
&\leq g\big(2^{-n}\big) + g\big(\mu_n(\{\pi_n(t)\})\big)
\end{aligned}$$

and that by (4.18) and (4.35) $m(B(t,2^{-n})) \geq m(\{\pi_n(t)\})$. Thus

$$\begin{aligned}
\int_0^1 g\big(m(B(t,\epsilon))\big)\, d\epsilon &\leq \sum_{i=1}^{\infty} 2^{-n}\, g\big(m(B(t,2^{-n}))\big) \\
&\leq \sum_{i=1}^{\infty} 2^{-n}\, g\big(m(B(t,2^{-n}))\big) \\
&\leq \sum_{i=1}^{\infty} 2^{-n}\, g\big(\mu_n(\{\pi_n(t)\})\big) + \sum_{i=1}^{\infty} 2^{-n}\, g\big(2^{-n}\big) \\
&\leq 32\,\alpha + 2\sqrt{\log 2} \\
&\leq K\,\alpha.
\end{aligned}$$

Since t was arbitrary, the same is true for the supremum over t. This completes the proof of the majorising measure lower bound, assuming Lemma 4.11. ■

We must now show that Lemma 4.11 is valid. Before we can do so, however, we have two preliminary results to establish. These are actually the key to the proof, for they will show how to pack into the metric space (T,d) small balls that are separated from one another by an appropriate (even smaller) distance while at the same time ensuring that many side conditions are satisfied. The proof of Lemma 4.11 then proceeds by replacing X by a process defined, essentially, on the centers of these small balls, and studying the supremum of the new process. The fact that the values of the new process have a certain minimal correlation between them (i.e. the centers of these balls are a certain d-distance apart) is what makes the new process easier to handle.

If you are familiar with mixing arguments for proving central limit or Poisson limit theorems, in which one takes larger and larger time intervals, split into more and more large subintervals, themselves separated by yet more comparatively small but actually quite large subintervals, then you should think of what we are about to do as being a similar construction, but with "small" replacing "large" throughout.

4.12 LEMMA. *If (T,d) is a metric space of diameter D and $A_1,\dots,A_n$ a partition of T, then there exists a non-empty subset I of $\{1,\dots,n\}$ such that, for all $i \in I$,*

$$\alpha(A_i) \;\geq\; \alpha(T) \;-\; D\sqrt{2\log(1+|I|)},$$

where $|I|$ denotes the number of elements in I.

PROOF: Order the A_i so that $\alpha(A_1) \geq \alpha(A_2) \geq \dots$. Fix $S \subset T$, and choose $m_i \in \mathcal{P}(A_i \cap S)$, $i = 1,\dots,n$. Set $a_i = (i+1)^{-2}$, note $\sum_{i=1}^n a_i \leq 1$, and define $m \in \mathcal{P}(S)$ by

$$m(A) \;=\; \frac{\sum_{i=1}^n a_i m_i(A \cap A_i)}{\sum_{i=1}^n a_i m_i(S)}.$$

Then, for $t \in S \cap A_i$,

$$\begin{aligned}\int_0^\infty g\big(m(B(t,\epsilon))\big)\,d\epsilon \;&\leq\; \int_0^D g\big(a_i m_i(B(t,\epsilon))\big)\,d\epsilon \\ &\leq\; Dg(a_i) \;+\; \int_0^D g\big(m_i(B(t,\epsilon))\big)\,d\epsilon.\end{aligned}$$

Thus,

$$\begin{aligned}\gamma(m,S) \;&\leq\; \sup_{1\leq i\leq n}\Big\{Dg(a_i) \;+\; \gamma(m_i, A_i \cap S)\Big\} \\ &=\; Dg(a_{i_o}) \;+\; \gamma(m_{i_o}, A_{i_o} \cap S),\end{aligned}$$

for some $i_o \in \{1,\dots,n\}$. By taking infima first over $m_i \in \mathcal{P}(A_{i_o})$ and then over $m \in \mathcal{P}(S)$ it follows from the above and the definition of γ that $\gamma(S) \leq Dg(a_{i_o}) + \gamma(A_{i_o} \cap S)$. This, in turn, implies that

$$\alpha(T) \leq Dg(a_{i_o}) + \alpha(A_{i_o}).$$

Taking $I = \{1,\dots,i_o\}$, so that $|I| = i_o$, and noting the monotonicity of the $\alpha(A_i)$, completes the proof. ∎

The next lemma is the important one. To formulate it, for each $i \geq 1$ and $A \subseteq T$ set

$$\beta_i(A) \;=\; \alpha(A) \;-\; \sup_{t \in A} \alpha\bigl(A \cap B(t, 6^{-i-1})\bigr).$$

4.13 LEMMA. *If (T,d) is a finite metric space of diameter $D \leq 6^{-i}$ then there exists a non-empty subset I of $\{1,\ldots,|T|\}$ and subsets $\{B_k\}_{k \in I}$ of T such that*

$$\text{(4.36)} \qquad diam(B_k) \;\leq\; 6^{-i-1}$$

$$\text{(4.37)} \qquad d(B_j, B_k) \;\geq\; 6^{-i-2} \qquad \text{for } j \neq k,$$

$$\text{(4.38)} \qquad \alpha(B_k) + \beta_{i+1}(B_k) \;\geq\; \alpha(T) + \beta_i(T) - 6^{-i+1}\bigl(2 + \sqrt{\log|I|}\bigr).$$

PROOF: By induction over k, construct a sequence $\{A_k, B_k, T_k\}_k$ of triplets of subsets of T that satisfy

$$T_k = T \setminus \bigcup_{j<k} A_j, \qquad \emptyset \neq B_k \subseteq A_k \subseteq T_k,$$

$$\operatorname{diam}(A_k) \leq 6^{-i-1}, \qquad d(B_k, T_k \setminus A_k) \geq 6^{-i-2}$$

and such that for each k either

$$\text{(4.39)} \qquad \alpha(B_k) + \alpha(T_k) \;\geq\; 2\,\alpha(A_k),$$

or

$$\text{(4.40)} \qquad \alpha(B_k) + \beta_{i+1}(B_k) \;\geq\; \alpha(T_k).$$

Denote those k for which (4.39) is satisfied by I_1, and those for which (4.40) is satisfied by I_2.

The construction starts with $T_1 = T$. Given T_k, if

$$\text{(4.41)} \qquad \alpha\bigl(B(t, 6^{-i-2})\bigr) + \alpha(T_k) \;\geq\; 2\,\alpha\bigl(B(t, 2\cdot 6^{-i-2}) \cap T_k\bigr)$$

for some $t \in T_k$ then set $A_k = B(t, 2\cdot 6^{-i-2}) \cap T_k$ and $B_k = B(t, 6^{-i-2}) \cap T_k$. In this case $k \in I_1$. Otherwise, set $A_k = B(t_o, 3\cdot 6^{-i-2}) \cap T_k$ and $B_k = B(t_o, 2\cdot 6^{-i-2}) \cap T_k$, where $t_o \in T_k$ maximises $\alpha(B(t, 2\cdot 6^{-i-2}) \cap T_k)$. In this case, $k \in I_2$, since

$$\begin{aligned}
\beta_{i+1}(B_k) \;&\geq\; \alpha(B_k) - \sup_{t \in B_k} \alpha(B(t, 6^{-i-2}) \cap T_k) \\
&\geq\; \alpha(B_k) + \alpha(T_k) - 2 \sup_{t \in B_k} \alpha(B(t, 6^{-i-2}) \cap T_k) \\
&\geq\; \alpha(T_k) - \alpha(B_k),
\end{aligned}$$

the second inequality coming from (4.41) with the inequality sign reversed. The construction stops when $\alpha(T_k) < \alpha(T) - 2 \cdot 6^{-i}$, and let κ denote the first k for which this is true. Thus, in particular, $\alpha(T_k) \geq \alpha(T) - 2 \cdot 6^{-i}$ for all $k < \kappa$.

Since $\alpha(T_\kappa) < \alpha(T) - 6^{-i}\sqrt{2\log 3}$, the previous lemma, applied to the two set partition $\{\bigcup_{k<\kappa} A_k, T_\kappa\}$ of T, gives us that

$$\alpha\Big(\bigcup_{k<\kappa} A_k\Big) \;\geq\; \alpha(T) \;-\; 6^{-i}\sqrt{2\log 2}.$$

Another application of the previous lemma gives that there exists a non-empty $I \subseteq \{1,\ldots,\kappa-1\} = I_1 \cup I_2$ such that for $k \in I$

$$\begin{aligned}
\alpha(A_k) \;&\geq\; \alpha\Big(\bigcup_{k<\kappa} A_k\Big) \;-\; 6^{-i}\sqrt{2\log(1+|I|)} \\
&\geq\; \alpha(T) \;-\; 6^{-i}\Big(\sqrt{2\log 2} \;+\; \sqrt{2\log(1+|I|)}\Big) \\
&\geq\; \alpha(T) \;-\; 2\cdot 6^{-i}\Big(2 \;+\; \sqrt{\log(1+|I|)}\Big).
\end{aligned}$$

Finally, we obtain that for $k \in I \cap I_1$

$$\begin{aligned}
\alpha(B_k) \;+\; \beta_{i+1}(B_k) \;&\geq\; 2\,\alpha(A_k) \;-\; \alpha(T_k) \\
&\geq\; \alpha(T) \;+\; \beta_i(T) \;-\; 3\big(\alpha(T) - \alpha(A_k)\big) \\
&\geq\; \alpha(T) \;+\; \beta_i(T) \;-\; 6^{-i}\Big(2 \;+\; \sqrt{\log(1+|I|)}\Big).
\end{aligned}$$

Similarly, if $k \in I \cap I_2$ then

$$\begin{aligned}
\alpha(B_k) \;+\; \beta_{i+1}(B_k) \;&\geq\; \alpha(T) \;-\; 2\cdot 6^{-i} \\
&\geq\; 2\,\alpha(T) \;-\; \alpha(A_k) \;-\; 4\cdot 6^{-i}\Big(2 \;+\; \sqrt{\log(1+|I|)}\Big) \\
&\geq\; \alpha(T) \;+\; \beta_i(T) \;-\; 6^{-i}\Big(2 \;+\; \sqrt{\log(1+|I|)}\Big),
\end{aligned}$$

which proves the lemma. ∎

All that now remains is to provide a

PROOF OF LEMMA 4.11: Choose a finite $V \subset T$. Suppose $\alpha(V) \leq 6^3\,\mathrm{diam}(V)$, and take $s,t \in v$ with $d(s,t) = \mathrm{diam}(V)$. Since by Lemma 2.7 $E\max(X_s, X_t) = \mathrm{diam}(V)/\sqrt{2\pi}$, we have that (4.33) holds with $K = \sqrt{2\pi}\,6^3$.

Thus, we assume henceforth that $\alpha(V) > 6^3\,\mathrm{diam}(V)$. Recall that we have a global assumption, as discussed at the beginning of the previous section, that $\mathrm{diam}(V) \leq 1$. Set

$$\begin{aligned}
L \;&=\; \max\{j\colon \mathrm{diam}(V) \leq 6^{-j}\}, \\
M \;&=\; \min\Big\{m\colon \inf_{s,t\in V}\{d(s,t)\colon d(s,t) > 0\} \;>\; 2^{-m}\Big\}.
\end{aligned}$$

We now call on the construction of Lemma 4.13, which can be applied $M-L$ times to establish the existence on non-empty families $\mathcal{B}_L, \ldots, \mathcal{B}_M$ of non-empty subsets of V such that $\mathcal{B}_L = \{V\}$, $\sup_{B\in\mathcal{B}_i} \operatorname{diam}(B) \leq 6^{-i}$ and if $B \neq B'$ are both in $\mathcal{B}_i$ then $d(B, B') \geq 6^{-i-2}$. Indeed, a careful sequential application of Lemma 4.13 allows us to assume that for every $B \in \mathcal{B}_i$ and $j < i$ there is a $B' \in \mathcal{B}_j$ with $B \subseteq B'$.

Furthermore, if for $t \in V$ there is a (necessarily unique) $B \in \mathcal{B}_i$ with $t \in B$, then we denote this B by B_t^i and we shall write $N_t^i = |\{B \in \mathcal{B}_{i+1} : B \subseteq B_t^i\}|$. Then, by Lemma 4.13,

$$\alpha(B) + \beta_{i+1}(B) \geq \alpha(B_t^i) + \beta_i(B_t^i) - 6^{-i+1}\left(2 + \sqrt{\log N_t^1}\right), \tag{4.42}$$

for each $B \in \mathcal{B}_{i+1}$ such that $B \subset B_t^i$.

Now set $\widetilde{V} = \bigcap_{i=L}^{M} \bigcup_{B\in\mathcal{B}_i} B$, and $\Psi_t^k = \sum_{i=k}^{M-1} 6^{-i-1}\sqrt{\log N_t^i}$, for $t \in \widetilde{V}$. Note that $B_t^M = \{t\}$, and $B_t^L = V$, so that

$$\beta_M(B_t^M) = \alpha(B_t^M) = 0, \qquad \alpha(B_t^L) = \alpha(V), \qquad \beta_L(B_t^L) \geq 0.$$

This allows to use the following telescoping sum, to which we can apply (4.42):

$$\begin{aligned}
\alpha(V) &\leq \inf_{t\in\widetilde{V}} \sum_{i=L}^{M-1} \alpha(B_t^i) + \beta_i(B_t^i) - \alpha(B_t^{i+1}) - \beta_{i+1}(B_t^{i+1}) \\
&\leq 36 \inf_{t\in\widetilde{V}} \Psi_t^L + \tfrac{1}{2} 6^{-L+2} \\
&\leq 36 \inf_{t\in\widetilde{V}} \Psi_t^L + \tfrac{1}{2}\alpha(V),
\end{aligned}$$

so that

$$\alpha(V) \leq 72 \inf_{t\in\widetilde{V}} \Psi_t^L. \tag{4.43}$$

Now we define a family of comparison processes. Let $\{\{\xi_B\}_{B\in\mathcal{B}_i}\}_{i=L+1}^{M}$ be a collection of i.i.d. standard Gaussian variables, and for $t \in \widetilde{V}$ set $\xi_t^i = \xi_{B_t^i}$. Define, for $t \in \widetilde{V}$ and $L \leq k \leq M-1$,

$$Y_t^k = \sum_{i=k+1}^{M} 6^{-i}\xi_t^i. \tag{4.44}$$

For $s,t \in \widetilde{V}$ with $d(s,t) > 0$ we have that $6^{-i_o-1} < d(s,t) < 6^{-i_o}$ for some i_o, which puts s and t in the same element of $\mathcal{B}_i$ for all $i \le \max(L, i_o - 2)$ so that

$$
\begin{aligned}
E\big(Y_t^L - T_s^L\big)^2 &= E\Big(\sum_{i=\max(L+1,i_o-1)}^{M-1} 6^{-i}\big(\xi_t^i - \xi_s^i\big)\Big)^2 \\
&\le 2\sum_{i=i_o-1}^{\infty} 6^{-2i} \qquad\qquad (4.45)\\
&\le \frac{2\cdot 6^4}{35} E\big(X_t - X_s\big)^2.
\end{aligned}
$$

Set $\mathcal{C}_k = \{B \cap \widetilde{V} : B \in \mathcal{B}_k\}$. Then, from Lemma 4.10 and the definition of Ψ_t^k, we have that there is a universal constant K such that for $k = M-1$ and for each $B \in \mathcal{C}_k$

$$E \sup_{t\in B} Y_t^k \;\ge\; K \inf_{t\in B} \Psi_t^k. \tag{4.46}$$

If we could show that (4.46) was also true for $k = L$, we would be done, for then by (4.43), (4.45), Theorem 2.8 (the weak version of the Sudakov-Fernique inequality) and the fact that $\widetilde{V} \subseteq V$ we would have that

$$\alpha(V) \;\le\; K\, E\|Y^L\|_{\widetilde{V}} \;\le\; K'\, E\|X\|_{\widetilde{V}} \;\le\; E\|X\|_V,$$

as required.

Fortunately, we can get precisely what we need by backward induction. Thus, assume that (4.46) is true for some $L < k \le M-1$. We shall show that it is also true for $k-1$. This is all we need.

Thus, with k fixed, choose a $B \in \mathcal{C}_{k-1}$. Set $\mathcal{C}_k^B = \{C \in \mathcal{C}_k : C \subseteq B\}$, and let Ω_C denote the event that $Y_C > Y_{C'}$ for all $C' \ne C$ in $\mathcal{C}_k^B$. Because of the finiteness of the parameter spaces, there is a well defined random variable $\tau : \Omega \to \mathcal{C}_k^B$ defined by the relationship

$$Y_{\tau(\omega)}^k(\omega) \;=\; \sup_{t\in C} Y_t^k(\omega) \qquad \text{on } \omega \in \Omega_C. \tag{4.47}$$

Now use the independence of the ξ_t^k and (4.46) as follows:

$$\begin{aligned}
E \sup_{t\in B} Y_t^{k-1} &\geq E Y_\tau^{k-1} \\
&= \sum_{C\in \mathcal{C}_k^B} E\Big\{ I_{\Omega_C}\Big(\sup_{t\in C} Y_t^k + 6^{-k}\xi_C\Big)\Big\} \\
&= \frac{1}{|\mathcal{C}_k^B|}\sum_{C\in \mathcal{C}_k^B} E\sup_{t\in C} Y_t^k + 6^{-k} E \sup_{C\in \mathcal{C}_k^B} \xi_C \\
&\geq \frac{1}{|\mathcal{C}_k^B|}\sum_{C\in \mathcal{C}_k^B} \inf_{t\in C} \Psi_t^k + K 6^{-k}\sqrt{\log |\mathcal{C}_k^B|} \\
&\geq K \inf_{t\in B} \Psi_t^{k-1},
\end{aligned}$$

which is precisely what we had to show. ■

REMARK: Because this proof has been so very long and convoluted, you should think carefully about precisely what we have needed to make it work. One of the most important steps has been to move the problem from a general parameter space to a finite one. We have used the finiteness a number of times, but never more crucially than in the very last stage, in the definition of the random variable τ at (4.47). It is this step, more than any other, that once more seems to indicate that majorising measures are closely related to the distributions of the positions in T of suprema. Unfortunately, however, the proof relies on the position of the supremum of the comparison process Y, and not that of X itself, so that we cannot obtain this variable in the final result. (Again, I reiterate that in general we have no assurance that such a random variable need be uniquely defined, but that in the case of continuous X Pollard's (1990) result does give us this.) Thus the construction of majorising measures in the most general situations, remains, for the moment, an elusive goal.

4. Entropy.

We have now completed the main results associated with majorising measures. Since this is, however, a somewhat difficult tool to work with, we now return to the use of a somewhat easier, albeit not quite as efficient, concept.

For $\epsilon > 0$, let $N(\epsilon)$ as usual denote the minimal number of d-balls of radius ϵ needed to cover T. Then, as we noted in the Introduction, $H(\epsilon) = \log N(\epsilon)$ is called the (metric) entropy of T. We refer to any result or condition based on N or H as an entropy result/condition.

The fact that entropy results are generally implied by results involving majorising measures is a consequence of the following important lemma.

4.14 LEMMA. *If $\int_0^\infty (\log N(\epsilon))^{\frac{1}{2}} d\epsilon < \infty$, then there exists a majorising measure m and a universal constant K such that*

$$\sup_{t\in T}\int_0^\eta g\big(m(B(t,\epsilon))\big)\,d\epsilon \;<\; K\Big(\eta|\log\eta| \;+\; \int_0^\eta \big(\log N(\epsilon)\big)^{\frac{1}{2}}\,d\epsilon\Big), \tag{4.48}$$

for all $\eta > 0$.

PROOF: As usual, we can and so shall assume that $\operatorname{diam}(T) = 1$. For $n \geq 0$, let $\{A_{n,1},\ldots, A_{n,N(2^{-n})}\}$ be a minimal family of d-balls of radius 2^{-n} which cover T. Set

$$B_{n,k} \;=\; A_{n,k}\setminus \bigcup_{j<k} A_{n,j}, \tag{4.49}$$

so that $\mathcal{B}_n = \{B_{n,1},\ldots, B_{n,N(2^{-n})}\}$ is a partition of T and each B_i is contained in a d-ball of radius 2^{-n}. Define a probability measure m on T by

$$m(A) \;=\; \tfrac{1}{2}\sum_{n=0}^{\infty} 2^{-n}\,\big(N(2^{-n})\big)^{-1}\sum_{B\in\mathcal{B}_n}\chi(A\cap B),$$

where $\chi(A) = 1$ if $A \neq \emptyset$ and $\chi(A) = 0$ otherwise. For every $t \in T$ we have that if $\epsilon \in (2^{-(n+1)}, 2^{-n}]$, then

$$m\big(B(t,\epsilon)\big) \;\geq\; \big(2^{n+1}N(2^{-(n+1)})\big)^{-1}.$$

This implies that, for all $t \in T$ and all $n \geq 0$,

$$\begin{aligned}
\int_0^{2^{-n}} &\big(\log\big(1/m(B(t,\epsilon))\big)\big)^{\frac{1}{2}}\,d\epsilon\\
&\leq\; \sum_{k=n+1}^{\infty} 2^{-k}\big(\log\big(2^kN(2^{-k})\big)\big)^{\frac{1}{2}}\\
&\leq\; \sum_{k=n+1}^{\infty} 2^{-k}\big(k\log 2\big)^{\frac{1}{2}} + 2\int_0^{2^{-n}}\big(\log(N(\epsilon)\big)^{\frac{1}{2}}\,d\epsilon\\
&\leq\; (n+2)2^{-n}\sqrt{\log 2} + 2\int_0^{2^{-n}}\big(\log(N(\epsilon)\big)^{\frac{1}{2}}\,d\epsilon,
\end{aligned}$$

the last line following from a little elementary algebra.

It is now easy to complete the proof, using monotonicity arguments to pass from dyadic η to all η. ■

An immediate consequence of Theorem 4.1 and Lemma 4.14 is the following:

4.15 COROLLARY. *A sufficient condition for the continuity of a centered Gaussian process on T is that $\int_0^\infty (\log(N(\epsilon))^{\frac{1}{2}}\, d\epsilon < \infty$. Furthermore, there exists a universal constant K such that*

$$E\|X\| \leq K \int_0^\infty \big(\log N(\epsilon)\big)^{\frac{1}{2}}\, d\epsilon. \tag{4.50}$$

PROOF: To establish continuity we need only, in the notation of, and by, Theorem 4.5 show that $\gamma_m(\eta) \to 0$ as $\eta \to 0$ for some m. This, however, is immediate from Lemma 4.14 and the finiteness of $\int_0^\infty (\log(N(\epsilon))^{\frac{1}{2}}\, d\epsilon$.

The inequality (4.50) is an immediate consequence (in the case $D = 1$, which is sufficiently general) of setting $\eta = 1$ in (4.48). ■

There is also a lower bound to $E\|X\|$ involving entropy conditions, dating back to Sudakov (1971), viz.

$$K \sup_{\epsilon > 0} \epsilon\big(\log N(\epsilon)\big)^{\frac{1}{2}} \leq E\|X\|.$$

Given what we have behind us, this inequality is easy to prove. Details of how to proceed appear in Exercise 4.4.

Since entropy calculations are easier to make than are those based on majorising measures, a natural question to ask is whether or not there is a converse to the above corollary.

When X is *stationary*, the answer is positive, and in 1975 Fernique established the following result for processes on $\Re^d$, $d \geq 1$. We can easily obtain a somewhat more general result from what we have already laboured to prove.

4.16 THEOREM. *Let X be a stationary Gaussian process on a compact group, or a compact subset of an infinite group. Then the following three conditions are equivalent:*

(i) X *is a.s. continuous on* T,

(ii) X *is a.s. bounded on* T,

(iii) $$\int_0^\infty (\log(N(\epsilon)))^{\frac{1}{2}}\, d\epsilon < \infty.$$

PROOF: That (i) implies (ii) is obvious. That (iii) implies (i) is Corollary 4.15. Thus it suffices to show that (ii) implies (iii), which we shall now do. Note firstly that by Theorem 4.4 we know that

$$\sup_{t \in T} \int_0^\infty g\big(m(B(t,\epsilon))\big)\, d\epsilon < \infty \tag{4.51}$$

for m either normalised Haar measure on T, in the compact case, or the normalised restriction of infinite Haar measure in the second case. Furthermore, by stationarity, the value of the integral must be independent of t.

For $\epsilon \in (0,1)$ let $M(\epsilon)$ be the maximal number of points $\{t_k\}_{k=1}^{M(\epsilon)}$ in T for which

$$\min_{1 \le j,k \le M(\epsilon)} d(t_j, t_k) \;>\; \epsilon.$$

It is easy to check (Exercise 4.3) that

$$N(\epsilon) \;\le\; M(\epsilon) \;\le\; N(\epsilon/2).$$

Thus, since m has total mass unity, we must have

$$m\big(B(t,\epsilon)\big) \;\le\; \big(N(2\epsilon)\big)^{-1}.$$

Consequently, by (4.51) and, in particular, its independence on t

$$\begin{aligned} \infty > \int_0^\infty g\big(m\big(B(t,\epsilon)\big)\big)\, d\epsilon \;&\ge\; \int_0^\infty \big(\log N(2\epsilon)\big)^{\frac{1}{2}}\, d\epsilon \\ &=\; 2\int_0^\infty \big(\log N(\epsilon)\big)^{\frac{1}{2}}\, d\epsilon, \end{aligned}$$

which proves the theorem. ∎

That an entropy condition cannot in general provide a necessary and sufficient condition for continuity can be seen by counterexamples. Here is one.

4.17 EXAMPLE. *Let $\{X_n\}_{n\ge 1}$ a sequence of i.i.d. Gaussian variables such that*

$$\sigma_n \;=\; \sigma(X_n) \;=\; (EX_n^2)^{\frac{1}{2}} \;=\; (1+\log n)^{-\frac{1}{2}}. \tag{4.52}$$

Then the sequence $\{X_n\}$ is a.s. bounded but does not satisfy the "finite entropy" condition.

PROOF: Note that for each $n \ge 1$ and λ large enough ($\lambda \ge 2$ will do)

$$\begin{aligned} P\{X_n \ge \lambda\} \;&\le\; e^{-\lambda^2/2\sigma_n^2} \\ &\le\; e^{-\frac{1}{2}\lambda^2(1+\log n)} \\ &\le\; K n^{-\frac{1}{2}\lambda^2}\, e^{-\frac{1}{2}\lambda^2}. \end{aligned}$$

Thus

$$\begin{aligned} P\{\sup_n |X_n| > \lambda\} \;&=\; P\{\exists n \ge 1\colon\ |X_n| > \lambda\} \\ &\le\; K e^{-\frac{1}{2}\lambda^2} \sum_{n=1}^{\infty} n^{-\frac{1}{2}\lambda^2} \\ &\le\; K e^{-\frac{1}{2}\lambda^2}. \end{aligned} \tag{4.53}$$

It is now a trivial calculation to check (Borel-Cantelli) that $\sup_n X_n$ is a.s. bounded.

Note that so far we have not had to use the fact that the X_n are independent to prove boundedness of the supremum.

The sequence $\{X_n\}_{n\geq 1}$ does not necessarily give a finite entropy integral however. Now use the fact that the X_n are independent, and take $\epsilon > 0$. Then, for $n < n_\epsilon = \exp(-1 + 1/(2\epsilon^2))$, we have $\sigma(X_n) > \epsilon\sqrt{2}$. Thus,

$$d(n,m) \;>\; 2\epsilon \qquad \text{for all } \; n,m < n_\epsilon .$$

Since this means that n and m cannot belong to the same ϵ ball if $n,m < n_\epsilon$, it follows that $N(\epsilon) \geq n_\epsilon - 1$, so that

$$\inf_{\epsilon>0} \epsilon(\log N_\epsilon)^{\frac{1}{2}} > 0,$$

and so the metric entropy integral $\int (\log N(\epsilon))^{\frac{1}{2}}\, d\epsilon$ cannot be finite. ■

Despite the lack of a finite entropy integral, however, there must be an appropriate majorising measure. In this case it is not too hard to find, and an appropriate measure is

$$m(\{n\}) \;=\; \frac{K_\alpha}{n(\log n)^{1+\alpha}} \tag{4.54}$$

for any $\alpha > 0$, where K_α is the appropriate normalising constant. There are many counter-examples of this kind. All are based, as is the above, on being forced to use too many balls to cover parts of the parameter space that are somehow too "thin" for anything important to happen.

One way to get around this, while remaining with metric entropy, is to try to divide the space up in two stages, firstly by looking to see which parts are most likely to be problematic, and only then doing entropy calculations. This is an approach that has turned out to be useful in getting good bounds for the supremum probabilities $P\{\|X\| > \lambda\}$ and we shall employ it to much advantage in the following chapter.

A result based on this approach is the following, which goes part of the way towards closing the gap between the entropy and majorising measure conditions for continuity. It is due to Samorodnitsky (1988). The proof we give is based on a personal communication of Talagrand.

Firstly, note that by Theorem 3.6 it is enough to study a.s. continuity at each point of T individually in order to determine full sample path continuity. Let $t_o \in T$ be fixed, and let $B(t_o,\delta)$ be the d-ball around t_o of radius δ. Furthermore, for each $\eta \in (0,\delta)$, set

$$A(t_o,\delta,\eta) \;=\; \{t \in T : \eta < d(t,t_o) \leq \delta\},$$

so that if $\eta < \delta$

$$B(t_o,\delta) \;=\; B(t_o,\eta) \cup A(t_o,\delta,\eta).$$

For a general set $C \subset T$, let $N(C,\epsilon)$ be the minimal number of d-balls of radius ϵ required to cover C. Then the following is true.

4.18 THEOREM. *A sufficient condition for the a.s. continuity of X at the point $t_o \in T$ is the existence of a function $H(s,t)$ satisfying the following two conditions:*

$$\lim_{s \to 0} \int_0^s H^{\frac{1}{2}}(s,t)\, dt = 0, \tag{4.55}$$

$$\log N\big(B(t_o, \eta, \delta), \epsilon\big) \le H(\eta, \epsilon). \tag{4.56}$$

for all $\epsilon > 0$ and any $0 < \eta < \delta < \Delta$, Δ some fixed constant.

It is easy to see that Theorem 4.18 gives weaker conditions for continuity than does Corollary 4.15, and there are examples covered by the weaker conditions only. Counterexamples show, however, that even the use of a "two-parameter entropy" cannot give necessary *and* sufficient conditions for continuity.

The proof of this result is a very nice example of how judicious use of Borell's inequality, can easily lead to powerful results.

PROOF OF THEOREM 4.18: Consider the process $Y_t = X_t - X_{t_o}$. Clearly Y induces the same canonical metric on T that X induces, and all entropies remain unchanged.

Let $C_n = \{t \in T\colon 2^{-n} \le d(t, t_o) \le 2^{-n+1}\}$. It clearly suffices to show that for any sequence $\epsilon_n \downarrow 0$ the sequence $P\{\sup_{t \in C_n} Y_t > \epsilon_n\}$ is summable. Note firstly that

$$\int_0^{D(C_n)} \big(\log N(C_n, \epsilon)\big)^{\frac{1}{2}}\, d\epsilon \le 4 \int_0^{2^{-n}} \big(H(2^{-n}, \epsilon)\big)^{\frac{1}{2}}\, d\epsilon,$$

by (4.56), so that by Corollary 4.15 and 4.55

$$\lim_{n \to \infty} E \sup_{t \in C_n} Y_t = 0.$$

Let $\eta_n = E \sup_{t \in C_n} Y_t$. By Borell's inequality

$$P\{\sup_{t \in C_n} Y_t > \epsilon_n\} \le 2e^{-\frac{1}{2}(\epsilon_n - \eta_n)^2/2^{-n}}.$$

By letting ϵ_n decrease slowly enough we obtain a summable series and so complete the proof. ■

5. Ultrametricity and discrete majorising measures.

If you have ever so much as browsed through the papers of Fernique and Talagrand on Gaussian processes (I would hope that by now you have done more than just browsed) you will have noticed that there is one glaring omission in the current notes. The word "ultrametric", which is so very prevalent there, has not appeared here at all.

A metric space (T,d) is called ultrametric if for all $r,s,t \in T$ we have

$$d(s,t) \leq \max\big(d(s,r),\, d(r,t)\big), \tag{4.57}$$

and a Gaussian process X on a space T is called ultrametric if T, together with the canonical metric generated by X, is ultrametric.

The most important aspect of ultrametricity, from the point of view of Gaussian processes, is that two balls of the same radius are either identical or disjoint. We have already seen two very important examples of ultrametric spaces in this chapter, but before I tell you where they were, let us look at a very general way to construct examples.

Let $\{T_n\}_{n\geq 1}$ be a collection of finite sets, and $\{\pi_m^n : T_n \to T_m\}_{1\leq m\leq n}$ a collection of mappings such that for $m \leq n \leq p$

$$\pi_m^p = \pi_m^n \circ \pi_n^p,$$

and, for $t \in T_n$, $\pi_n^n(t) = t$. If T is the projective limit of the $\{T_n, \pi_m^n\}$, furnished with the natural metric, then T is a compact ultrametric space.

To define a ultrametric Gaussian process on T we argue as follows: Firstly, let π_n denote the projection of T onto T_n. Then to every point $t \in \bigcup_n T_n$ assign a standard Gaussian random variable $\xi(t)$. Take the $\xi(t)$ to be mutually independent. Choose $0 < q < 1$, and for $t \in T$ define

$$X(t) = \sum_{n=1}^{\infty} q^n\, \xi\big(\pi_n(t)\big). \tag{4.58}$$

This process is ultrametric, as you can check for yourself (Exercise 5.1).

Now you should remember where we have used ultrametric spaces in these notes. The proofs of both the upper and lower bounds of Theorem 4.1 relied on comparing a given process to one whose structure was somewhat simpler – simpler enough to permit calculation. In both cases, the processes constructed were ultrametric.

In fact, we could have rewritten our proofs somewhat (thus taking them closer to the style of Talagrand's) by first proving what we needed for ultrametric processes, and afterwards mapping the results across to the general case. (Historically, this is precisely how the decade from the mid 'seventies until the mid 'eighties saw most of the majorising measure results derived.) Whether one works this way, or as we have, is very much a matter of taste. Both routes involve the same amount of work – there are no free lunches here.

To be fair, however, I should point out that there are certain advantages to taking the route of first proving things for ultrametric processes and then generalising. The first is that since the first stage of the proof is easier, it gives you a better chance of checking whether or not a new result may or may not be correct and/or provable. Secondly, it has certain distinct advantages when dealing with non-Gaussian processes.

To see how useful ultrametricity can be in terms of simplifying proofs, you should look at a series of papers by Evans (1988a,b, 1989) that are concerned with the sample path properties and extrema of Gaussian processes indexed by local fields. Although Evans does not use the notion of ultrametricity directly, the structure of his parameter space gives him properties of this kind for free, and so allows for neater proofs and occasionally more powerful results than one can obtain in a general setting. ■

We can now turn to the notion of discrete majorising measures, which, as with ultrametricity, can sometimes save one some work.

If (T,d) is a metric space, then a probability measure m is called a *discrete majorising measure* for T if there exists a countable set $S \subset T$ which supports m and a sequence of mappings $\{\pi_n\}_{n\geq 1}$ from T into S such that

$$d(t,\pi_n(t)) \leq 2^{-n}, \qquad t\in T,\ n\geq 1, \tag{4.59}$$

and

$$\sup_{t\in T}\sum_{n=1}^{\infty} 2^{-n} g\big(m(\pi_n(t))\big) < \infty. \tag{4.60}$$

To see an nice example of why it is sometimes easier to use discrete majorising measures than their regular counterparts, you should look at Andersen *et. al.* (1988), which has a detailed and powerful treatment of central limit theorems and laws of the iterated logarithm for empirical processes on very general parameter spaces. At this point I shall content myself by just quoting the main results from there which link the two types of majorising measures. The proof is left as an exercise.

4.19 THEOREM. *If (T,d) has admits a majorising measure μ, then it also admits a discrete majorising measure m. If μ satisfies*

$$\lim_{\delta\to 0}\sup_{t\in T}\int_0^{\delta} g\big(\mu(B(t,\epsilon))\big)\, d\epsilon = 0 \tag{4.61}$$

then m can be chosen to satisfy

$$\lim_{k\to\infty}\sup_{t\in T}\sum_{n=k}^{\infty} 2^{-n}\, g\big(m(\{\pi_n(t)\})\big) = 0. \tag{4.62}$$

This result, along with what has gone before, is sufficient to establish that essentially all the results that we have stated relating boundedness and continuity of Gaussian processes to majorising measures have a corresponding discrete majorising measure form.

For example, if you look again at the upper bound proof for Theorem 4.1, you will see that we have actually proven there that the comparison process Y of (4.20) is a.s. bounded on T when there exists a discrete majorising measure on T. In fact, we showed that the upper bound in (4.3) can be replaced by

$$E\|X\| \ \leq \ K \sup_{t\in T} \sum_{n=k}^{\infty} 2^{-n}\, g\big(m(\{\pi_n(t)\})\big). \tag{4.63}$$

You should work through the details yourself in Exercise 5.3. In view of Theorem 4.19, a corresponding lower bound also holds.

Again, for more information on discrete majorising measures see Anderson *et. al.* (1988).

6. Discontinuous Processes.

Up until this point we have concentrated on finding conditions under which a Gaussian process X is continuous, and, once it is known to be continuous, measuring its smoothness in terms of moduli of continuity. It seems only reasonable to make a slight detour to see what happens when X is discontinuous.

Recall first that by the zero-one law of Theorem 3.12, if the probability that X is continuous on T is less than one then this probability must be zero, and so X is actually discontinuous with probability one. It then follows from Theorem 3.6 that, for *each* $t \in T$,

$$P\{\lim_{s\to t} X_s \neq X_t\} \ = 1.$$

Thus X is discontinuous with probability one at every point in a dense subset of T, and so discontinuous at every point in T.

Like the little girl with the curl in the middle of her forehead, when X is good (continuous) it is very very good, but when it is bad, it is horrid.

My task now, as an adherent of the Gaussian school, is to show you that things aren't quite as bad as they seem, as long as you look at them the right way. The way to do this will be to set $T = [0,1]$ (although any compact $T \subset \Re^k$ would do) and to describe some very elegant results due to Don Geman, written up in an unpublished manuscript, Geman (198?). (If you want details, you will have to write to Geman himself. If enough of you do this, this may convince him to publish the paper, as should have been done long ago!) We start with the concept of Lebesgue density.

Let $A \subseteq [0,1]$ be Lebesgue measurable, and let

$$D_0(A) := \lim_{t\downarrow 0} \frac{\lambda(A\cap(0,t))}{t}$$

be the Lebesgue density of A at zero. (λ is, as usual, Lebesgue measure.) $D_0(A)$ is a measure of how "thick" A is in the vicinity of $t=0$. The basic idea is going to be that any set that has Lebesgue density zero at 0 is so thin that we are going to be prepared to throw it away when we consider the local behaviour of X near 0.

Thus, if ϕ is a Borel measurable function on $[0,1]$, we define

$$\operatorname{ap}\limsup_{t\downarrow 0} \phi(t) = \inf\{a\colon D_0(\{s\colon \phi(s) > a\}) = 0\},$$

and

$$\operatorname{ap}\liminf_{t\downarrow 0} \phi(t) = \sup\{a\colon D_0(\{s\colon \phi(s) < a\}) = 0\},$$

where the "ap" here is to be read as "approximate". Clearly, if $\operatorname{ap}\liminf_{t\downarrow 0} \phi_t = \operatorname{ap}\limsup_{t\downarrow 0} \phi_t$, we call the resulting number the approximate limit of ϕ at zero, and write it as $\operatorname{ap}\lim_{t\downarrow 0} \phi_t$.

One more definition will suffice for the main result. We shall call a non-negative function $\psi = \psi_\phi$ an *approximate upper function* for ϕ at 0 if

$$\operatorname{ap}\lim_{t\downarrow 0} \frac{|\phi(t)-\phi(0)|}{\psi(t)} \le 1.$$

Approximate lower functions are similarly defined. Here is Geman's main result.

4.20 THEOREM. *Let X be centered, Gaussian on $[0,1]$, and set*

$$p(u) = \left(E|X_u - X_0|^2\right)^{\frac{1}{2}}.$$

If X is stationary, p increasing, and if

$$\lim_{u\downarrow 0} \frac{p(u/|\log u|)}{p(u)} = 1, \tag{4.64}$$

then

$$\operatorname{ap}\limsup_{t\downarrow 0} \frac{|X_t - X_0|}{p(t)\sqrt{\log|\log t|}} \le 1$$

with probability one. That is, $p(t)\sqrt{\log|\log t|}$ is an a.s. approximate upper function for X at 0.

If X is not stationary, then set

$$(4.65)\qquad \tilde{\rho}(u) = \sup\left\{\frac{E(X_t - X_0)(X_s - X_0)}{p(s)p(t)} : \frac{s}{t} = u, \quad s,t \in (0,1)\right\},$$

and assume that

$$(4.66)\qquad \int_0^1 \frac{ds}{(1-\tilde{\rho}(s))^{\frac{1}{2}}} < \infty.$$

Then $p(t)\sqrt{2\log|\log t|}$ is an a.s. approximate upper function for X at 0.

I shall not even attempt to prove this result, other than to note that it is based on the non-probabilistic result that a set A has density 0 at $t = 0$ if, and only if,

$$\int_A \Psi\left(\frac{m(A\cap(0,t))}{t}\right)\frac{dt}{t} < \infty,$$

for a some continuous, strictly increasing Ψ on $[0,1]$ with $\Psi(0) = 0$.

You can find precursors to Theorem 4.20, along with full proofs, in Geman (1977, 1979) and Geman and Zinn (1978), but only Geman (198?) has all the details.

The interest in this theorem is that it covers very many processes that are discontinous in the usual sense, but continuous in the "approximate" sense. In fact, there is a (very reasonable) conjecture that this is true of *all* Gaussian processes on $[0,1]$. Don Geman described it to me as follows: "If, after we have drawn a discontinuous, unbounded, Gaussian sample path on the blackboard, we were to step far enough backwards so that we could no longer see sets of zero Lebesgue density, the sample path would become continuous and bounded".

There are examples in the Exercises.

7. Exercises.

SECTION 4.1:

1.1. Let $m_1, \ldots, m_n$ be a sequence of majorising measures on (T,d), and $\alpha_1, \ldots, \alpha_n$ a sequence of non-negative numbers summing to 1. Show that $\sum_{k=1}^n \alpha_k m_k$ is also a majorising measure. That is, convex combinations of majorising measures yield majorising measures.

SECTION 4.4:

4.1. Let T be an compact metric space. Construct a centered Gaussian process on T whose entropy integral diverges, but whose sample paths are continuous with probability one. (Hint. Use the construction of Theorem 1.6 together with the arguments of §4.4.)

4.2. You are now going to prove Corollary 4.15 directly, following Dudley's (1973) proof.

Let $N(\epsilon)$ be the entropy function for X on T, set $H(\epsilon) = \log N(\epsilon)$ and $f(\eta) = \int_0^\eta (H(\epsilon))^{\frac{1}{2}}\, d\epsilon$. We are going to show that f is a modulus of continuity for X in the canonical metric. This will imply Corollary 4.15.
(i) Define sequences $\delta_n \downarrow 0$ and $\epsilon_n \downarrow 0$ inductively as follows: $\epsilon_1 = 1$. Given $\epsilon_1, \ldots, \epsilon_n$, let

$$\begin{aligned} \delta_n &= 2\inf\{\epsilon : H(\epsilon) \le H(\epsilon_n)\}, \\ \epsilon_{n+1} &= \min(\epsilon_n/3, \delta_n). \end{aligned}$$

Show that

$$\tfrac{2}{3} \sum_{m=n}^{\infty} H^{\frac{1}{2}}(\epsilon_m)\epsilon_m \le f(\epsilon_n) \le 4 \sum_{m=n}^{\infty} H^{\frac{1}{2}}(\epsilon_m)\epsilon_m .$$

(ii) Note that for each $n \ge 1$ there is a set $\mathcal{A}_n \subset T$ such that, for any $t \in T$, $\inf_{s\in\mathcal{A}_n} d(s,t) \le 2\delta_n$. Set $\mathcal{G}_n = \{s - t : s,t \in \mathcal{A}_{n-1} \cup \mathcal{A}_n\}$, and let

$$P_n = P\Big\{ \max_{t\in\mathcal{G}_n} \frac{X(t)}{E[X^2(t)]^{1/2}} \ge 3H^{\frac{1}{2}}(\epsilon_n)\Big\}.$$

By using entropy arguments to bound the number of elements of $\mathcal{G}_n$, along with standard Gaussian inequalities, show that

$$P_n \le 4\exp\{-\tfrac{1}{2}H(\epsilon_n)\}.$$

(iii) For any $t \in T$, set $A_0(t) = t$, and for $n \ge 1$ let $A_n(t) \in \mathcal{A}_n$ be such that $d(t, A_n(t)) \le 2\delta_n$. Write X_t as

$$X_t = \sum_{n=0}^{\infty} \big(X(A_n(t)) - X(A_{n+1}(t))\big),$$

and use this representation plus the results of (i) and (ii) to show that f is a modulus of continuity for X.

4.3. For $\epsilon \in (0,1)$ let $M(\epsilon)$ be the maximal number of points $\{t_k\}_{k=1}^{M(\epsilon)}$ in T for which

$$\min_{1\le j,k\le M(\epsilon)} d(t_j, t_k) > \epsilon.$$

Show that

$$N(\epsilon) \leq M(\epsilon) \leq N(\epsilon/2).$$

4.4. To give a direct proof of Sudakov's lower bound

$$K \sup_{\epsilon > 0} \epsilon\big(\log N(\epsilon)\big)^{\frac{1}{2}} \leq E\|X\|,$$

proceed as follows:

Fix $\epsilon > 0$ and choose a maximal set of points, T_ϵ, say, as in the above exercise. Choose i.i.d. zero mean Gaussian variables $\{Y_n\}_{n=1}^{M(\epsilon)}$ with variance ϵ^2. Apply the Sudakov-Fernique inequality of Theorem 2.8 to compare $E\|X\|_{T_\epsilon}$ to $E\|Y_n\|$. Then bound the latter from below by Lemma 4.10. Now send $\epsilon \to 0$ to get a result involving $E\|X\|_T$. If you did it all properly, you will have proven Sudakov's lower bound.

4.5. Find an example satisfying the entropy condition of Theorem 4.18 but not that of Corollary 4.15.

SECTION 4.5:

5.1. Show that the process defined by (4.58) is in fact ultrametric.

5.2 Prove Theorem 4.19.

(Hint: Use the construction we made in the proof of the upper bounds on $E\|X\|$, defining π_n as there, and setting $\mu_n(t) = m(\pi_n(t))$. This result is due, in this explicit formulation, to Anderson *et. al.* (1988).)

5.3. Prove the upper bound to $E\|X\|$ of (4.63). (Note that since T is finite all integrals involving majorising measures are in fact sums.)

SECTION 4.6:

6.1 Let X_t be stationary on $[0,1]$ and assume $E|X_u - X_0|^2 \sim |\log u|^{-\beta}$, $0 < \beta < \infty$. Show that X is continuous for $\beta > 1$ and discontinuous otherwise. In the former case find the regular modulus of continuity at 0. In both cases, calculate the approximate modulus of continuity.

6.2 Let X_t be an index-β process on $[0,1]$. Show that (4.40) is satisfied for these processes. Calculate regular and approximate upper functions at 0 and compare them.