Regularity Theory

In this chapter we treat two questions about Gaussian processes simultanteously. Namely, "when is a Gaussian process continuous?"; and "when is a Gaussian process bounded"? We begin by discussing a sufficient condition for questions such as continuity and boundedness. That condition is based on a very general principle about abstract stochastic processes, and involves the notion of metric entropy.

1. Metric Entropy

Let $(T\,,d)$ be a non-empty, compact metric space and define $B(t\,,r)$ to be the closed d-ball of radius $r > 0$ about $t \in T$; that is,

$$B(t\,,r) := \{s \in T : d(s\,,t) \leqslant r\} \qquad [t \in T,\ r > 0].$$

By default, for every $\varepsilon > 0$ there exists an integer $n(\varepsilon) \geqslant 1$ and points $t_1, \ldots, t_{n(\varepsilon)} \in T$ such that $T = \cup_{j=1}^{n(\varepsilon)} B(t_j\,,\varepsilon)$.

Definition 1.1. We write $N_T(\varepsilon)$ for the smallest such integer $n(\varepsilon)$. The function N_T is called the *metric entropy* of $(T\,,d)$.

The function $N_T : (0\,,\infty) \to \mathbb{Z}_+$ is non-increasing, and the behavior of $N_T(\varepsilon)$ for $\varepsilon \approx 0$ quantifies the "size" of the compact set T. For example, we can see easily that T is a finite set if and only if $\lim_{\varepsilon\downarrow 0} N_T(\varepsilon) < \infty$. And if $\lim_{\varepsilon\downarrow 0} N_T(\varepsilon) < \infty$, then the rate at which $N_T(\varepsilon)$ diverges can yield information about the geometry of T. The following is one way in which this statement can be quantified.

Definition 1.2. The *Minkowski dimension* of $(T\,,d)$ is defined as

$$\dim_{\mathrm{M}}(T) := \limsup_{\varepsilon\downarrow 0} \frac{\log N_T(\varepsilon)}{\log(1/\varepsilon)}.$$

Some authors refer to $\dim_{\mathrm{M}}(T)$ as the "fractal dimension" of T. See Mandelbrot XXX, for instance.

ex:dimM:1 **Example 1.3.** Here are a few simple examples that show how the behavior of N_T near zero can describe the "size" of T.

(1) Consider the set $T := \prod_{i=1}^n [a_i\, b_i]$ where $a_i < b_i$ are real numbers. We endow T with the Euclidean metric, $d(s\,,t) := \|s-t\|$ for all $s,t \in T$. Then it is not hard to see that $\dim_{\mathrm{M}}(T) = n$. In other words, the Minkowski dimension of T agrees with any reasonable topological notion of dimension for T.

(2) Let T denote the standard ternary Cantor set in $\mathbb{R}_+$, and endow T with the Euclidean metric, $d(s\,,t) := |t-s|$ for all $s,t \in T$. Then it is possible to verify that $\dim_{\mathrm{M}}(T) = \log 2/\log 3$, which ought to be a familiar computation to you.

(3) If T is a finite set then $\dim_{\mathrm{M}}(T) = 0$. However, there are countable spaces that have non-zero Minkowski dimension. For instance, consider $T := \{1\,,1/2\,,1/3\,,1/4\,,\ldots\}$, endowed with the Euclidean metric, $d(s\,,t) := |t-s|$ for all $s,t \in T$. It is a good exercise to prove that $\dim_{\mathrm{M}}(T) = 1/2$.

(4) There are also many metric spaces of infinite Minkowski dimension. An example is the space T of all continuous, real-valued functions on $[0\,,1]$, endowed with the usual metric,

$$d(s\,,t) := \sup_{0\leqslant y\leqslant 1} |s(y)-t(y)| \qquad \text{for all } s,t \in T.$$

Or one can consider $T = L^p[0\,,1]$ for any $1 \leqslant p \leqslant \infty$, endowed with $d(s\,,t) := \|s-t\|_{L^p[0,1]}$ for all $s,t \in L^p[0\,,1]$.

The main result of this section is a careful version of the assertion that a T-indexed stochastic process is continuous if the index set T is "not too big," as understood, in one fashion or another, via the behavior of N_T near zero.

Now let $\{X_t\}_{t\in T}$ be a real-valued stochastic process, indexed by a set T, where $(T\,,d)$ is a metric space. Define

$$\Psi_X(u) := \sup_{s,t\in T} \mathbb{P}\left\{|X_t - X_s| > d(s\,,t)u\right\} \qquad [u>0].$$

Also, introduce the "tail" functions $\{\mathcal{T}_{p,X}\}_{p\geqslant 1}$ as follows.

$$\mathcal{T}_{p,X}(\lambda) := p\int_0^\infty u^{p-1}\left(\lambda\Psi_X(u)\wedge 1\right)\mathrm{d}u \qquad [\lambda>0, p\geqslant 1].$$

The goal of this section is to prove the following result about increments of general T-indexed stochastic processes. In the next section we will work out examples that highlight some of the uses of such a theorem.

th:entropy

Theorem 1.4. *For every finite set $S\subset T$ and all $p\geqslant 1$,*

$$\mathbb{E}\left(\max_{\substack{s,t\in S:\\ d(s,t)\leqslant\delta}}|X_t-X_s|^p\right)\leqslant\left\{12\int_0^\delta\left[\mathcal{T}_{p,X}\left([N_S(r)]^2\right)\right]^{1/p}\mathrm{d}r\right\}^p,$$

for every $0<\delta\leqslant\triangle(T)$, where $\triangle(S) := \sup_{s,t\in S} d(s\,,t)$ denotes the d-diameter of S. If $T\subset\mathbb{R}^M$ for some $M\geqslant 1$ and is endowed with any Euclidean metric d, then there exists a finite constant $L\geqslant 1$—depending only on $(M\,,d)$—such that, for every finite set $S\subset T$ and all $p\geqslant 1$,

$$\mathbb{E}\left(\max_{\substack{s,t\in S:\\ d(s,t)\leqslant\delta}}|X_t-X_s|^p\right)\leqslant\left\{12\int_0^\delta\left[\mathcal{T}_{p,X}\left(L\cdot N_S(r)\right)\right]^{1/p}\mathrm{d}r\right\}^p,$$

for every $0<\delta\leqslant\triangle(T)$.

Remark 1.5. The second part of the previous theorem holds not only when $(T\,,d)$ is a metric subset of a Euclidean space, but in fact whenever $(T\,,d)$ is a metric subset of a metric space $(A\,,d)$ to which the Besicovitch covering theorem applies; for more information on this topic consult Füredi and Loeb XXX.

There are many variations on Theorem 1.4 XXX. This formulation of Theorem 1.4 is particularly elegant, and might well be new. The essence of the proof can be traced back to an unpublished manuscript of Kolmogorov, with non-trivial extensions due to Preston XXX and particularly Dudley XXX. The argument rests on the following simple *a priori* estimate.

lem:entropy:1

Lemma 1.6. *Let $A\subset T\times T$ be a finite set of cardinality $|A|$. Then,*

$$\mathbb{E}\left[\max_{(s,t)\in A}|X_t-X_s|^p\right]\leqslant\mathcal{T}_{p,X}(|A|)\cdot\max_{(s,t)\in A}[d(s\,,t)]^p \qquad \text{for all } p\geqslant 1.$$

Proof. For every $u>0$,

$$\mathbb{P}\left\{\max_{(s,t)\in A}\left|\frac{X_t-X_s}{d(s\,,t)}\right|\geqslant u\right\}\leqslant\sum_{(s,t)\in A}\mathbb{P}\left\{\left|\frac{X_t-X_s}{d(s\,,t)}\right|\geqslant u\right\}\leqslant|A|\Psi_X(u)\wedge 1.$$

Integrate $[pu^{p-1}\,\mathrm{d}u]$ to see that

$$\mathbb{E}\left(\max_{(s,t)\in A}\left|\frac{X_t - X_s}{d(s\,,t)}\right|^p\right) \leqslant \mathcal{T}_{p,X}(|A|).$$

This implies the lemma. □

Proof of Theorem 1.4. Let us first consider the general case; the Euclidean case requires making small adjustments and is deferred to the end of the proof.

We can restrict attention to $\{X_s\}_{s\in S}$ in order to see that we can assume—without loss in generality—that $S = T$ is a finite set.

Define

$$K_n := N_T\left(2^{-n}\triangle(T)\right) \qquad [n \geqslant 0].$$

Then clearly, $1 = K_0 \leqslant K_1 \leqslant K_2 \leqslant \ldots.$

The definition of metric entropy ensures that for every integer $n \geqslant 0$ we can find a finite set $T_n \subset T$ such that:

(1) $|T_n| = K_n$;
(2) $\inf_{s\in T_n} d(t\,,s) \leqslant 2^{-n}\triangle(T)$ for all $t \in T$; and
(3) There exists an integer $M \geqslant 1$ such that $T_n = T$ for all $n \geqslant M$.

Therefore, among other things, we can associate to every point $t \in T$ a unique point $\pi_n(t) \in T_n$ such that:

(P.a) $d(t\,,\pi_n(t)) \leqslant 2^{-n}\triangle(T)$;[1] and
(P.b) $\pi_n(s) = \pi_n(t)$ whenever $d(s\,,t) \leqslant 2^{-n}\triangle(T)$.

Since $K_0 = 1$, we can see from the definition of $\triangle(T)$ that we can choose $T_0 := \{t_0\}$, where $t_0 \in T$ is an arbitrary point. Having observed this we then write $X_t - X_{t_0} = \sum_{n=0}^{\infty}(X_{\pi_{n+1}(t)} - X_{\pi_n(t)})$ for every $t \in T$, all the time noticing that the sum is composed of not more than $M+1$ nonzero terms, and hence is absolutely convergent a.s. In particular,

$$|X_t - X_s| \leqslant \sum_{n=0}^{\infty}\left|X_{\pi_{n+1}(t)} - X_{\pi_n(t)} - X_{\pi_{n+1}(s)} + X_{\pi_n(s)}\right| \qquad [s,t \in T],$$

where the summands are zero for all but a finite number of integers n. If, in addition, there exists an integer $m \geqslant 0$ such that $d(s\,,t) \leqslant 2^{-m}\triangle(T)$, then $\pi_k(t) = \pi_k(s)$ for all $k \geqslant m$. From here it follows readily that if $d(s,t) \leqslant 2^{-m}\triangle(T)$, then

$$|X_t - X_s| \leqslant \sum_{n=m}^{\infty}\left|X_{\pi_{n+1}(t)} - X_{\pi_n(t)}\right| + \sum_{n=m}^{\infty}\left|X_{\pi_{n+1}(s)} - X_{\pi_n(s)}\right|.$$

[1] There can be more than one such point; if so, then choose any one of the possibles one. Since T is finite, this can even be done algorithmically, if needed.

Now, $d(\pi_{n+1}(t)\,,\pi_n(t)) \leqslant d(\pi_n(t)\,,t) + d(\pi_{n+1}(t)\,,t) \leqslant \frac{3}{2}\cdot 2^{-n}\triangle(T)$, and a similar bound holds for $d(\pi_{n+1}(s)\,,\pi_n(s))$ as well. Therefore,

$$\max_{\substack{s,t\in T\\ d(s,t)\leqslant 2^{-m}\triangle(T)}} |X_t - X_s| \leqslant 2\sum_{n=m}^{\infty} \max_{\substack{u\in T_{n+1},v\in T_n\\ d(u,v)\leqslant\frac{3}{2}\cdot 2^{-n}\triangle(T)}} |X_u - X_v|.$$

By Minkowski's inequality,

$$\left\| \max_{\substack{s,t\in T\\ d(s,t)\leqslant 2^{-m}\triangle(T)}} |X_t - X_s| \right\|_{L^p(\mathbb{P})} \leqslant 2\sum_{n=m}^{\infty} \left\| \max_{\substack{u\in T_{n+1},v\in T_n\\ d(u,v)\leqslant\frac{3}{2}\cdot 2^{-n}\triangle(T)}} |X_u - X_v| \right\|_{L^p(\mathbb{P})},$$

for all real numbers $p \geqslant 1$. The cardinality of T_n is $K_n \leqslant K_{n+1}$ and the cardinality of T_{n+1} is K_{n+1}. Therefore, Lemma 1.6 implies that

$$\begin{aligned}\left\| \max_{\substack{s\in T_{n+1},t\in T_n\\ d(s,t)\leqslant\frac{3}{2}\cdot 2^{-n}\triangle(T)}} |X_t - X_s| \right\|_{L^p(\mathbb{P})} &\leqslant 3\cdot 2^{-n}\triangle(T)\left[\mathcal{T}_{p,X}\left(K_{n+1}^2\right)\right]^{1/p} \qquad (7.1)\\ &= 6\cdot 2^{-1-n}\triangle(T)\left[\mathcal{T}_{p,X}\left(\left[N_T\left(2^{-1-n}\triangle(T)\right)\right]^2\right)\right]^{1/p},\end{aligned}$$

eq:for:KCT

for every $n \geqslant m$. Sum over all $n \geqslant m$ in order to see that

$$\begin{aligned}&\left\| \max_{\substack{s,t\in T\\ d(s,t)\leqslant 2^{-m}\triangle(T)}} |X_t - X_{t_0}| \right\|_{L^p(\mathbb{P})}\\ &\qquad\leqslant 6\sum_{n=m}^{\infty} 2^{-1-n}\triangle(T)\left[\mathcal{T}_{p,X}\left(\left[N_T\left(2^{-1-n}\triangle(T)\right)\right]^2\right)\right]^{1/p}\\ &\qquad= 12\sum_{n=m}^{\infty}\int_{2^{-2-n}\triangle(T)}^{2^{-1-n}\triangle(T)}\left[\mathcal{T}_{p,X}\left(\left[N_T\left(2^{-1-n}\triangle(T)\right)\right]^2\right)\right]^{1/p}\,\mathrm{d}\varepsilon\\ &\qquad\leqslant 12\int_0^{2^{-1-m}\triangle(T)}\left[\mathcal{T}_{p,X}\left([N_T(\varepsilon)]^2\right)\right]^{1/p}\,\mathrm{d}\varepsilon. \qquad (7.2)\end{aligned}$$

eq:for:KCT:1

If $0 < \delta \leqslant \triangle(T)$, then we can find always a unique integer $m \geqslant 0$ such that $2^{-m-1}\triangle(T) < \delta \leqslant 2^{-m}\triangle(T)$, whence

$$\begin{aligned}&\left\| \max_{\substack{s,t\in T\\ d(s,t)\leqslant\delta}} |X_t - X_{t_0}| \right\|_{L^p(\mathbb{P})} \leqslant \left\| \max_{\substack{s,t\in T\\ d(s,t)\leqslant 2^{-m}\triangle(T)}} |X_t - X_{t_0}| \right\|_{L^p(\mathbb{P})}\\ &\leqslant 12\int_0^{2^{-1-m}\triangle(T)}\left[\mathcal{T}_{p,X}\left([N_T(\varepsilon)]^2\right)\right]^{1/p}\,\mathrm{d}\varepsilon \leqslant 12\int_0^{\delta}\left[\mathcal{T}_{p,X}\left([N_T(\varepsilon)]^2\right)\right]^{1/p}\,\mathrm{d}\varepsilon,\end{aligned}$$

by the preceding portion of the proof. The general form of the theorem follows from this.

The Euclidean form is proved similarly; except we notice that by the Besicovitch covering theorem XXX,

$$L = L(M\,,d) := \sup_{n\geqslant 0}\max_{v\in T_n}\left|\left\{u\in T_{n+1} : d(u\,,v)\leqslant \tfrac{3}{2}\cdot 2^{-n}\triangle(T)\right\}\right| < \infty.$$

[For instance, $L\leqslant 2^{1+M}$ when $d(s\,,t) := \max_{i\leqslant M}|s_i - t_i|$.] Therefore, (7.1) can be improved to the following:

$$\begin{aligned}\left\| \max_{\substack{s\in T_{n+1}, t\in T_n\\ d(s,t)\leqslant\frac{3}{2}\cdot 2^{-n}\triangle(T)}} |X_t - X_s| \right\|_{L^p(\mathbb{P})} &\leqslant 6\cdot 2^{-1-n}\triangle(T)\left[\mathcal{T}_{p,X}\left(LK_{n+1}\right)\right]^{1/p}\\ &= 6\cdot 2^{-1-n}\triangle(T)\left[\mathcal{T}_{p,X}\left(L\cdot N_T\left(2^{-1-n}\triangle(T)\right)\right)\right]^{1/p}.\end{aligned}$$

The remainder of the proof is unchanged. □

2. Continuity Theorems

§1. Kolmogorov's Continuity Theorem. Among other things, Theorem 1.4 and its variations can be used to sometimes show that a stochastic process $\{X_t\}_{t\in T}$ can be constructed in a nice way, where as before $(T\,,d)$ is a metric space.

Definition 2.1. Let $X := \{X_t\}_{t\in T}$ and $Y := \{Y_t\}_{t\in T}$ be two stochastic processes. We say that X is a *modification*—sometimes also called a *version*—of Y if $\mathbb{P}\{X_t = Y_t\} = 1$ for all $t\in T$.

Of course, if X is a version of Y, then in turn Y is a version of X as well. What the preceding really says is that X and Y have the same distributions in the sense that

$$\mathbb{P}\left\{X_{t_1}\in A_1\,,\ldots,X_{t_k}\in A_k\right\} = \mathbb{P}\left\{Y_{t_1}\in A_1\,,\ldots,Y_{t_k}\in A_k\right\},$$

for all Borel sets $A_1,\ldots,A_k\subset\mathbb{R}$ and all $t_1,\ldots,t_k\in T$. In particular, all computable probabilities for X are the same as their counterparts for the process Y. In this sense, if X and Y are modifications of one other, then they are "stochastically indistinguishable."

th:entropy:1 **Theorem 2.2.** *Let $X := \{X_t\}_{t\in T}$ be a stochastic process and suppose there exists a real number $p\geqslant 1$ and a compact separable set $S\subset T$ such that*

$$\int_0^{\triangle(S)}\left[\mathcal{T}_{p,X}\left([N_S(r)]^2\right)\right]^{1/p}\,\mathrm{d}r < \infty,$$

for some $p \geqslant 1$. *Then,* $\{X\}_{t\in S}$ *has a continuous version* $\{Y_t\}_{t\in S}$ *which satisfies*

$$\mathbb{E}\left(\sup_{\substack{s,t\in S:\\ d(s,t)\leqslant\delta}} |Y_t - Y_s|^p\right) \leqslant \left\{12\int_0^\delta \left[\mathcal{T}_{p,X}\left([N_S(r)]^2\right)\right]^{1/p} \mathrm{d}r\right\}^p,$$

for every $0 < \delta \leqslant \triangle(S)$. *If* $(T\,,d)$ *is a metric subset of* $\mathbb{R}^M$*, then we can replace* $[N_S(r)]^2$ *everywhere by* $L \cdot N_S(r)$ *in both of the preceding two displays, where* $L = L(M\,,d) \geqslant 1$ *is a finite universal number.*

Theorem 2.2 is a restatement of Theorem 1.4. But this particular formulation has the following useful consequence.

th:KCT **Theorem 2.3** (Kolmogorov's Continuity Theorem)**.** *Suppose that* $(T\,,d)$ *is compact and separable, and that* $\{X_t\}_{t\in T}$ *is a stochastic process for which we can finite constants* $A,\beta > 0$ *and* $p > 2\dim_{\mathrm{M}}(T)$ *such that*

$$\mathbb{E}\left(|X_t - X_s|^p\right) \leqslant A[d(s\,,t)]^{p+\beta} \qquad \textit{for all } s,t\in T. \tag{7.3}$$ eq:KCT

Then X *has a continuous version* Y *which satisfies the following: For every* $1 \leqslant q < p$*, and* $0 < \alpha < (p - 2\dim_{\mathrm{M}}(T))/q$*,*

$$\mathbb{E}\left(\sup_{\substack{s,t\in T:\\ s\neq t}} \left|\frac{Y_t - Y_s}{[d(s\,,t)]^\alpha}\right|^q\right) < \infty. \tag{7.4}$$ goal:KCT

If $(T\,,d)$ *is a compact subset of* $\mathbb{R}^M$ *for some integer* $M \geqslant 1$*, then the conditions for* p *and* α *can be respectively improved to* $p > \dim_{\mathrm{M}}(T)$ *and* $0 < \alpha < (p - \dim_{\mathrm{M}}(T))/q$*.*

This is an excellent time to go back and deduce, from Theorem 2.3, Proposition 2.2 on page 86. I will leave the few remaining details to you, and verify Theorem 2.3 only.

Proof. I will prove the general case; the Euclidean case is proved similarly. Let us apply Chebyshev's inequality to see that for all $u > 0$,

$$\Psi(u) = \sup_{s,t\in T} \mathbb{P}\left\{|X_t - X_s| > d(s\,,t)u\right\} \leqslant \sup_{s,t\in T} \frac{A[d(s\,,t)]^\beta}{u^p} = Bu^{-p},$$

where $B := A[\triangle(T)]^\beta$. This shows that whenever $q\in(0\,,p)$ and $\lambda > 0$,

$$\mathcal{T}_{q,X}(\lambda) \leqslant q\int_0^\infty u^q\left(\frac{B\lambda}{u^p}\wedge 1\right)\frac{\mathrm{d}u}{u} = C\lambda^{q/p},$$

where C is finite and equal to the same integral as the previous one, but with λ replaced by 1 [change of variables]. The definition of Minkowski dimension tells us that $N_t(\varepsilon) \leqslant \varepsilon^{-\dim_{\mathrm{M}}(T)+o(1)}$ as $\varepsilon \downarrow 0$, and that the inequality is an identity for infinitely-many values of $\varepsilon \downarrow 0$. In particular, for every $\varrho \in (\dim_{\mathrm{M}}(T)\,, p/2)$ we can find a finite constant D such that $N_t(r) \leqslant Dr^{-\varrho}$ for all $0 < r \leqslant \triangle(T)$. Consequently,

$$\int_0^\delta \left[\mathcal{T}_{q,X}\left([N_T(r)]^2\right)\right]^{1/q} \mathrm{d}r \leqslant \text{const}\cdot\int_0^\delta r^{-2\varrho/p}\,\mathrm{d}r \leqslant \text{const}\cdot\delta^{(p-2\varrho)/p},$$

uniformly for all $0 < \delta \leqslant \triangle(T)$. Theorem 2.2 now yields

$$\mathrm{E}\left(\sup_{\substack{s,t\in S:\\ d(s,t)\leqslant 2^{-m}\triangle(T)}} |X_t - X_s|^q\right) \leqslant \text{const}\cdot 2^{-m(p-2\varrho)q/p},$$

uniformly for all integers $m \geqslant 0$ and all finite [hence also countable] $S \subset T$. In particular,

$$\mathrm{E}\left(\sup_{\substack{s,t\in S:\\ 2^{-m-1}\triangle(T)<d(s,t)\leqslant 2^{-m}\triangle(T)}} \left|\frac{X_t - X_s}{[d(s\,,t)]^\alpha}\right|^q\right) \leqslant \text{const}\cdot 2^{-m(p-2\varrho-\alpha q)q/p},$$

uniformly for all integers $m \geqslant 0$ and real numbers $\alpha > 0$. We sum the preceding expectation over all $m \geqslant 0$ in order to see that, if in addition we ask that $p - 2\varrho - \alpha q > 0$, then the following slightly weaker version of (7.4) holds:

$$\mathrm{E}\left(\sup_{\substack{s,t\in S:\\ s\neq t}} \left|\frac{X_t - X_s}{[d(s\,,t)]^\alpha}\right|^q\right) < \infty. \tag{7.5}$$

goal:KCT:S

We apply the preceding with S being a dense subset of T, and define

$$Y_t := \limsup_{\substack{s\to t:\\ s\in S}} X_s \qquad \text{for all } t \in T.$$

Clearly, $\mathrm{P}\{Y_s = X_s\} = 1$ for all $s \in S$. And (7.3) shows that X is continuous in probability; that is, $X_s \to X_t$ in probability as $s \to t$. Therefore, we can deduce from the density of S in T that Y is a modification of X. Eq. (7.5), the construction of Y, and Fatou's lemma together imply (7.4) Since we can have ϱ arbitrarily close to $\dim_{\mathrm{M}}(T)$, this proves the theorem. □

§2. Application to Gaussian Processes. Let $X := \{X_t\}_{t\in T}$ denote a mean-zero Gaussian process, indexed by an arbitrary set T. Define

$$d(s\,,t) := \sqrt{\mathrm{E}\left(|X_t - X_s|^2\right)} \qquad [s, t \in T].$$

It is easy to see that $d(s\,,t) \leqslant d(s\,,u) + d(u\,,t)$ for all $s, t, u \in T$, and $d(s\,,t) = d(t\,,s)$. That is, d is a pseudo-metric on T. Let us write $s \sim t$

if $d(s,t) = 0$, and $[t] := \{s \in T : s \sim t\}$. Clearly, $\sim$ is an equivalence relation on T and $[t] \in T/\sim$ denotes the equivalence class of $t \in T$.

We can define $\bar{X}_{[t]} := X_s$ for all $[t] \in T/\sim$ as $\bar{X}_{[t]} := X_s$ for any and every $s \in [t]$. Then it follows that $\bar{X}$ is a mean-zero Gaussian process, indexed by $T/\sim := \{[t] : t \in T\}$, and with the same "finite-dimensional distributions" as X. In this way we can assume without loss of generality that (T,d) is a metric space; otherwise we study $\bar{X}$ in place of X, using the same methods.

With the preceding in mind, we can now see that Theorem 1.4 and its consequences imply sufficient conditions for X to have a continuous modification. In order to identify the details, we first develop two estimates.

lem:Psi:Gauss **Lemma 2.4.** $\Psi_X(u) < \exp(-u^2/2)$ *for all* $u > 0$.

Proof. The usual proof of this sort of fact yields twice the stated upper bound. The argument for this slight improvement is even easier, and borrowed from Khoshnevisan, XXX. Let U have a $\mathrm{N}(0,1)$ distribution on the line and observe that

$$\begin{aligned}\mathbb{P}\{U > u\} &= \frac{1}{\sqrt{2\pi}}\int_u^\infty \mathrm{e}^{-x^2/2}\,\mathrm{d}x = \frac{\mathrm{e}^{-u^2/2}}{\sqrt{2\pi}}\int_u^\infty \exp\left(-\frac{(x-u)(x+u)}{2}\right)\mathrm{d}x\\ &= \frac{\mathrm{e}^{-u^2/2}}{\sqrt{2\pi}}\int_0^\infty \exp\left(-\frac{y(y+2u)}{2}\right)\mathrm{d}y.\end{aligned}$$

If $u > 0$, then $y + 2u > y$ whence follows $\mathbb{P}\{U > u\} < \frac{1}{2}\exp(-u^2/2)$. This proves the lemma because $\Psi_X(u) = \mathbb{P}\{|U| > u\} = 2\mathbb{P}\{U > u\}$. $\square$

Lemma 2.5. $\mathcal{T}_{1,X}(\lambda) \leqslant 3\sqrt{\log(\lambda \vee \mathrm{e})}$ *for all* $\lambda > 0$.

Proof. If $0 < \lambda \leqslant \mathrm{e}$, then by Lemma 2.4, $\mathcal{T}_{1,X}(\lambda) \leqslant \int_0^\infty \exp(-u^2/2)\,\mathrm{d}u = \sqrt{\pi/2} \leqslant \sqrt{8}$. If $\lambda > \mathrm{e}$, then we write

$$\begin{aligned}\mathcal{T}_{1,X}(\lambda) &\leqslant \sqrt{2\log\lambda} + \lambda\int_{\sqrt{2\log\lambda}}^\infty \mathrm{e}^{-u^2/2}\,\mathrm{d}u = \sqrt{2\log\lambda} + \sqrt{2\pi}\,\lambda\mathbb{P}\left\{U > \sqrt{2\log\lambda}\right\}\\ &< \sqrt{2\log\lambda} + \sqrt{\frac{\pi}{2}} < 3\sqrt{\log\lambda},\end{aligned}$$

using the fact that $\mathbb{P}\{U > u\} \leqslant \frac{1}{2}\exp(-u^2/2)$ for all $u > 0$ [see proof of Lemma 2.4]. This has the desired consequence. $\square$

We can now appeal to the previous lemma and Theorem 1.4 in order to deduce the following, which is essentially due to Dudley XXX.

th:Dudley **Theorem 2.6** (Dudley, XXX). *For every finite subset $S \subset T$,*

$$\mathbb{E}\left(\max_{\substack{s,t\in S:\\ d(s,t)\leqslant\delta}} |X_t - X_s|\right) \leqslant 36\int_0^\delta \sqrt{\log(N_S(r)\vee \mathrm{e})}\,\mathrm{d}r \qquad [0<\delta\leqslant \triangle(T)].$$

If $(T\,,d)$ is compact and separable, and $\int_0^{\triangle(T)} \sqrt{\log(N_T(r))}\,\mathrm{d}r < \infty$, then X has a continuous modification Y, and

$$\mathbb{E}\left(\sup_{\substack{s,t\in T:\\ d(s,t)\leqslant\delta}} |Y_t - Y_s|\right) \leqslant 36\int_0^\delta \sqrt{\log(N_T(r)\vee \mathrm{e})}\,\mathrm{d}r \qquad [0<\delta\leqslant \triangle(T)].$$

Remark 2.7. Since $\int_{0+}\sqrt{\log(1/r)}\,\mathrm{d}r < \infty$, Theorem 2.6 implies that X has a continuous version whenever the Minkowski dimension of T is finite.

It is now easy to obtain bounds for the expectation of $\sup_{t\in S} X_t$.

co:Dudley **Corollary 2.8.** *For all finite sets $S \subset T$,*

$$\mathbb{E}\left[\max_{t\in S} X_t\right] \leqslant 36\int_0^{\triangle(S)} \sqrt{\log(N_S(r)\vee \mathrm{e})}\,\mathrm{d}r.$$

Therefore, if U is a denumerable subset of T, then $\sup_{t\in U}|X_t|$ is finite [a.s.] if and only if $\mathbb{E}[\sup_{t\in U} X_t] < \infty$.

Proof. Without loss of generality, we may and will assume that $S = T$ is finite. Choose and fix an arbitrary $t_0 \in T$ and use the fact that $\mathbb{E}(X_{t_0}) = 0$ to write $\mathbb{E}[\max_{t\in T} X_t] = \mathbb{E}[\max_{t\in T}(X_t - X_{t_0})]$. Therefore, Dudley's Theorem implies the inequality of the corollary. For the remainder, let $U \subset T$ be a countable set [if T is finite, then there is nothing to prove]. By the Borell, Sudakov–Tsirelson inequality [Theorem 2.1, page 55],

$$\mathbb{P}\left\{\left|\max_{t\in S}|X_t| - \mathbb{E}\left[\max_{t\in S}|X_t|\right]\right| > z\right\} \leqslant 2\mathrm{e}^{-z^2/2} \qquad [z>0],$$

for every finite $S \subset U$. Among other things, this inequality implies that $\sup_{t\in U}|X_t| < \infty$ iff $\mathbb{E}[\sup_{t\in U}|X_t|] < \infty$. Because X and $-X$ have the same law,

$$\mathbb{E}\left[\sup_{t\in U} X_t\right] \leqslant \mathbb{E}\left[\sup_{t\in U}|X_t|\right] \leqslant 2\,\mathbb{E}\left[\sup_{t\in U} X_t\right].$$

Therefore, the corollary follows. □

Example 2.9. Set $T := \{1\,,\ldots,n\}$ and define $\{Z_t\}_{t\in T}$ to be a sequence of i.i.d. N$(0\,,1)$ random variables. Because

$$d(s\,,t) = \begin{cases} \sqrt{2} & \text{if } s\neq t,\\ 0 & \text{if } s = t,\end{cases}$$

we can see that $\triangle(T) = \sqrt{2}$ and $N_T(r) = n$ for all $r > 0$. Corollary 2.8 yields

$$\mu(n) = \mathrm{E}\left[\max_{1\leqslant i\leqslant n} Z_i\right] \leqslant \text{const}\cdot\sqrt{\log n},$$

which we know is sharp to leading order [Proposition 1.3, page 7] .

We conclude this section by inspecting a classical condition for the continuity of a "stationary Gaussian process."

Example 2.10. Suppose $T = [0\,,1]$ and $\{X_t\}_{0\leqslant t\leqslant 1}$ is a stationary Gaussian process with $\mathrm{E}[X_t] = 0$ and $\mathrm{E}[X_tX_s] = \varrho(|t-s|)$ for a symmetric, strictly decreasing and continuous function $\varrho : \mathbb{R}_+ \to \mathbb{R}$ such that $\varrho(0) = 1$.[2] Because

$$d(s\,,t) = \sqrt{2\,(1-\varrho(|s-t|))} \qquad [0 \leqslant s, t \leqslant 1],$$

it follows that: (a) $\triangle(T) < \sqrt{2}$; (b) $(T\,,d)$ is compact; and (c) Every ball in $(T\,,d)$ is also a Euclidean ball. In fact,

$$B(t\,,r) = \left\{ s \geqslant 0 : |s-t| < \varrho^{-1}\left(1 - \frac{r^2}{2}\right)\right\} \qquad [0 \leqslant t \leqslant 1,\, 0 < r \leqslant \sqrt{2}].$$

From this, and Dudley's theorem, we can conclude that

$$N_{[0,1]}(\varepsilon) = \frac{(1+o(1))}{\varrho^{-1}\left(1-\frac{1}{2}\varepsilon^2\right)} \qquad \text{as } \varepsilon \downarrow 0.$$

This leads to the following sufficient condition XXX for the continuity of the process X,

$$\int_{0+} \sqrt{\log\left(\frac{1}{\varrho^{-1}\left(1-\frac{1}{2}r^2\right)}\right)}\,\mathrm{d}r < \infty.$$

As it turns out, this is also a necessary condition for both continuity and boundedness; see Theorem 3.7 on page 108 below.

Example 2.11. Let $X := \{X_t\}_{t\geqslant 0}$ denote Brownian motion on $\mathbb{R}$. In this example,

$$d(s\,,t) = \sqrt{\mathrm{E}\left(|X_s - X_t|^2\right)} = \sqrt{|t-s|} \qquad [s, t \geqslant 0].$$

Let us consider the restriction $\{X_t\}_{t\in T}$ of Brownian motion to the index set $T := [0\,,1]$. Then, $(T\,,d)$ is a compact metric space and $N_T(\varepsilon) \leqslant \varepsilon^{-2}$

---

[2] The linear Ornstein–Uhlenbeck process [p. 83] is an example of such a process with $\varrho(z) = \exp(-z)$.

for all $\varepsilon \in (0\,,1)$. In particular, Dudley's theorem yields that X has a continuous version W which satisfies

$$\mathrm{E}\left(\sup_{\substack{0\leqslant s,t\leqslant 1\\ |t-s|\leqslant\delta}} |W_t - W_s|\right) \leqslant \text{const}\cdot\sqrt{\delta\log(1/\delta)},$$

for all $0 < \delta < 1$. Now apply the proof of Kolmogorov's continuity theorem [Theorem 2.3] to see that[3]

$$\mathrm{E}\left(\sup_{\substack{0\leqslant s,t\leqslant 1\\ s\neq t}} \frac{|W_t - W_s|}{\sqrt{|t-s|\log(1/|t-s|)}}\right) < \infty.$$

In particular, the random variable under the expectation is finite a.s. This statement is sharp. For instance, a theorem of Lévy XXX states that

$$\lim_{\delta\to 0}\sup_{\substack{0\leqslant s,t\leqslant 1\\ 0<|t-s|\leqslant\delta}} \frac{|W_t - W_s|}{\sqrt{|t-s|\log(1/|t-s|)}} = \sqrt{2} \qquad \text{a.s.}$$

§3. An Infinite-Dimensional Example. Among other things, Dudley's theorem has found many applications in the theory of empirical processes and its connections to machine learning, etc. The following example is the sort that arises naturally in empirical-process theory (see, Dudley XXX for instance).

Let $\mathbb{H} := L^2[0\,,1]$ and consider white noise $X := \{X(h)\}_{h\in\mathbb{H}}$. Recall that X is a mean-zero Gaussian process with

$$\mathrm{Cov}[X(f)\,,X(g)] = \int_0^1 f(x)g(x)\,\mathrm{d}x \qquad [f,g\in\mathbb{H}].$$

It is easy to see that the random function X is unbounded on $\mathbb{H}$. For instance, choose and fix an arbitrary orthonormal family $\{\phi_k\}_{k=1}^\infty$ in $\mathbb{H}$—such as $\phi_k(x) = \sin(2\pi kx)$ for all $k \geqslant 0$—and note that $X(\phi_1), X(\phi_2),\ldots$ are i.i.d. $\mathrm{N}(0\,,1)$ random variables, and hence are unbounded. In fact, we have seen already that $\limsup_{k\to\infty} X(\phi_k)/\sqrt{2\log k} = 1$ a.s.; see (5.2) on page 55. Still, there are many infinite-dimensional subsets $T \subset \mathbb{H}$ for which $\{X(f)\}_{f\in T}$ defines a bounded random function. The following furnishes one such example.

Proposition 2.12. *Let T denote the collection of all Lipschitz-continuous functions $f : [0\,,1] \to \mathbb{R}$ such that $f(0) = 0$ and $\mathrm{Lip}(f) \leqslant 1$. Then, $\{X(f)\}_{f\in T}$ has a bounded version.*

[3] This is a little sharper than what we obtain from Kolmogorov's continuity theorem [Theorem 2.3].

Proof. Since T is not a closed subspace of $\mathbb{H}$, it is helpful to instead metrize T via $\varrho(f\,,g) := \sup_{0\leqslant x\leqslant 1}|f(x)-g(x)|$ for $f,g\in T$. I will prove that $\{X(f)\}_{f\in T}$ has a version Y that is continuous when T is metrized by ϱ rather than the courser metric $d(f\,,g) := \sqrt{\mathrm{E}(|X(f)-X(g)|^2)} = \|f-g\|_{\mathbb{H}}$ of $\mathbb{H}$. This enterprise will immediately yield the a.s.-boundedness of $f\mapsto Y(f)$, for example, as well as the measurability of $\sup_{f\in T}Y(f)$, $\sup_{f\in T}|Y(f)|$, etc.

The Arzela–Ascoli theorem XXX ensures that $(T\,,\varrho)$ is compact; i.e., T is closed and $N_{(T,\varrho)}(\varepsilon)<\infty$ for all $\varepsilon>0$. We wish to understand the behavior of $N_{(T,\varrho)}(\varepsilon)$ near $\varepsilon=0$.

For every integer $n\geqslant 0$ define T_n to be the collection of all continuous, piecewise linear, functions $g:[0\,,1]\to\mathbb{R}$ such that:

(1) g is linear on $[j/n\,,(j+1)/n]$ for every $0\leqslant j\leqslant n-1$;

(2) $|g((j+1)/n)-g(j/n)| = 1/n$.

It is easy to see that:

(i) For every $f\in T$ there exists $g\in T_n$ such that $\varrho(f\,,g)\leqslant n^{-1}$; and

(ii) $|T_n|\leqslant 2^n$.

Thus, $N_{(T,\varrho)}(1/n)\leqslant 2^n$, whence $\log N_{(T,\varrho)}(1/n)\leqslant n\log 2$. For all $\varepsilon\in(0\,,1)$ we can find an integer $n\geqslant 1$ such that $(n+1)^{-1}<\varepsilon\leqslant n^{-1}$. For this choice of n, we find that $\log N_{(T,\varrho)}(\varepsilon)\leqslant\log N_{(T,\varrho)}(1/(n+1))\leqslant(n+1)\log 2\leqslant 2+2\varepsilon^{-1}$. In particular,

$$\sqrt{\log(N_{(T,\varrho)}(\varepsilon)\vee \mathrm{e})}\leqslant\frac{2}{\sqrt{\varepsilon}}\vee 1\qquad[0<\varepsilon<1].$$

The preceding defines an integrable function of $\varepsilon\in(0\,,1)$. Since $\triangle(T\,,\varrho)=1$ and $[d(f\,,g)]^2=\mathrm{E}(|X(f)-X(g)|^2)=\int_0^1|f(x)-g(x)|^2\,\mathrm{d}x\leqslant[\varrho(f\,,g)]^2$ for all $f,g\in T$, it follows from Theorem 2.6 that X has a continuous version Y which satisfies

$$\mathrm{E}\left(\sup_{\substack{f,g\in T\\ \varrho(f,g)\leqslant\delta}}|Y(f)-Y(g)|\right)\leqslant 36\int_0^{\delta/2}\left(\frac{2}{\sqrt{\varepsilon}}\vee 1\right)\mathrm{d}\varepsilon\leqslant\text{const}\cdot\sqrt{\delta},$$

for all $\delta\in(0\,,1)$. It is possible to adapt the proof of Theorem 2.3 in order to deduce from the above fact that

$$\mathrm{E}\left[\sup_{\substack{f,g\in T:\\ f\neq g}}\frac{|Y(f)-Y(g)|}{\sqrt{\varrho(f\,,g)}}\right]<\infty,$$

and hence $f\mapsto X(f)$ is almost surely “Hölder continuous with index 1/2,” uniformly on $(T\,,\varrho)$. It is a good exercise to deduce from the above and the Borell, Sudakov–Tsirelson inequality [Theorem 2.1, page 55] that

$$\mathbb{E}\left[\sup_{\substack{f,g\in T:\\ f\neq g}}\left(\frac{|Y(f)-Y(g)|}{\sqrt{\varrho(f\,,g)}}\right)^p\right]<\infty,$$

for all real number $p \geqslant 1$. □

3. Lower Bounds

We continue using the notation of the preceding subsection. In particular, $X := \{X_t\}_{t\in T}$ denotes a mean-zero Gaussian process with canonical distance $d(s\,,t) := \sqrt{\mathbb{E}(|X_t - X_s|^2)}$, and we assume that $(T\,,d)$ is a compact metric space. In this section we discuss some useful lower bounds for $\mathbb{E}[\max_{t\in T} X_t]$, for example when T is countable.

§1. Sudakov Minorization. Recall that $Z_1,\ldots,Z_n$ are i.i.d. with a $\mathrm{N}(0\,,1)$ distribution, and define

$$\mu(n) := \mathbb{E}\left[\max_{1\leqslant i\leqslant n} Z_i\right] \qquad [n \geqslant 1].$$

lem:M **Lemma 3.1.** *There exist positive and finite constants K, L such that*

$$K\sqrt{\log n} \leqslant \mu(n) \leqslant L\sqrt{\log n} \qquad \text{for all } n \geqslant 1.$$

Proof. The Lemma holds trivially when $n = 1$; we concentrate on $n \geqslant 2$. Since $\mu(1) = 0$, $\mu(n) = (1 + o(1))\sqrt{2\log n}$ as $n \to \infty$ [Proposition 1.3, page 7], and μ is increasing, it suffices to prove that $\mu(2) > 0$. But $\mu(2)$ is equal to $\mathbb{E}[\max(Z_1\,,Z_2)] = \mathbb{E}[\max(Z_1\,,Z_2) - Z_1] = \mathbb{E}[\max(0\,,Z_2 - Z_1)]$. It follows easily from this that $\mu(2) > 0$.[4] □

rem:M **Remark 3.2.** The numerical values of K and L are not very good. For instance, the best-possible choice for K is

$$\inf_{n\geqslant 2}\frac{\mu(n)}{\sqrt{\log n}} \leqslant \frac{\mu(2)}{\sqrt{\log 2}} = \frac{1}{\sqrt{\pi\log 2}} < 0.7,$$

which is smaller than the limiting value, $\lim_{n\to\infty}\mu(n)/\sqrt{\log n} = \sqrt{2}$.

Lemma 3.1 will now be used in order to establish a useful lower bound for $\mathbb{E}[\sup_{t\in T} X_t]$.

[4] In fact, because $Z_2 - Z_1$ has a $\mathrm{N}(0\,,2)$ distribution, $Z_2 - Z_1$ is independent of $\mathrm{sign}(Z_2 - Z_1)$, and hence $\mu(2) = \frac{1}{2}\,\mathbb{E}(|Z_2 - Z_2|) = \pi^{-1/2}$.

pr:Sudakov **Proposition 3.3** (Sudakov, XXX). *Choose and fix some $\varepsilon > 0$, and let A be a subset of T with the property that $d(s\,,t) \geqslant \varepsilon$ whenever $s, t \in A$. Then,*

$$\mathbb{E}\left[\max_{t\in A} X_t\right] \geqslant \varepsilon\mu(|A|) \geqslant K\varepsilon\sqrt{\log(|A|)},$$

where K is the constant of Lemma 3.1 and $|\cdots|$ denotes cardinality.

Proof. In the case that $\mathrm{Var}(X_1) = \cdots = \mathrm{Var}(X_n)$, this theorem is just a restatement of Example 5.15 [page 76]. The general case is handled the same way, but requires an appeal to Fernique's inequality (Theorem 5.17, page 77) instead of Slepian's (Theorem 5.13, page 75). I will work out the details once more in order to gel the underlying ideas.

Without loss of generality, we can—and will—assume that $|A| \geqslant 2$; else the statement of the theorem is the tautology that $0 \geqslant 0$.

Define, for all $t \in A$,

$$Y_t := \varepsilon Z_t + \xi,$$

where ξ and the Z_t's are all independent $\mathrm{N}(0\,,1)$ random variable. Clearly, $(Y_t)_{t\in A}$ has a mean-zero multivariate normal distribution, and

$$\mathbb{E}\left(|Y_t - Y_s|^2\right) = \varepsilon^2 \geqslant [d(s\,,t)]^2 = \mathbb{E}\left(|X_s - X_t|^2\right).$$

Therefore, Fernique's inequality [Theorem 5.17, page 77] yields

$$\mathbb{E}\left[\max_{t\in A} X_t\right] \geqslant \mathbb{E}\left[\max_{t\in A} Y_t\right] = \varepsilon\,\mathbb{E}\left[\max_{t\in A} Z_t\right] = \varepsilon\mu(|A|),$$

by the definition of μ. □

Definition 3.4. Choose and fix some $r > 0$. We say that $t_1, \ldots, t_m \in T$ is an *r-packing of T* if $d(t_i\,,t_j) \geqslant r$ whenever $1 \leqslant i \neq j \leqslant m$. Let $P_T(r)$ denote the largest integer $m \geqslant 1$ for which there exists an r-packing of T. The function $r \mapsto P_T(r)$ is the *[Kolmogorov] capacity* of the pseudo-metric space $(T\,,d)$.

lem:N:C **Lemma 3.5.** *$N_T(r) \leqslant P_T(r) \leqslant N_T(r/2)$ for every $r > 0$.*

Proof. If $P_T(r) = m$, then we can find $t_1, \ldots, t_m \in T$ such that: (i) $d(t_i\,,t_j) \geqslant r$ when $i \neq j$; and (ii) $\min_{i\leqslant m} d(t_i\,,t) \leqslant r$ for all $t \in T$. This shows that $t_1, \ldots, t_m$ is an r-covering of T. Since $N_T(r)$ is the minimum size of all r-coverings of T, it follows that $N_T(r) \leqslant P_T(r)$. Conversely, suppose we can find $t_1, \ldots, t_\nu \in T$ such that $\cup_{i=1}^{\nu} B_d(t_i\,,r/2) = T$. If s_1 and s_2 are two points in T such that $d(s_1\,,s_2) \geqslant r$, then s_1 and s_2 cannot be in the same ball $B_d(t_i\,,r/2)$ for any $1 \leqslant i \leqslant \nu$. In particular, $P_T(r) \leqslant \nu$. The minimum such ν is of course $N_T(r/2)$. □

We can summarize the results of this section as follows.

pr:Sudakov:1

Proposition 3.6 (Sudakov Minorization). *We always have*

$$\sup_{\substack{S\subset T:\\ S \text{ finite}}} \mathbb{E}\left[\max_{t\in S} X_t\right] \geqslant \sup_{0<\varepsilon<\triangle(T)} \varepsilon\mu\left(P_T(\varepsilon)\right) \geqslant \sup_{0<\varepsilon<\triangle(T)} \varepsilon\mu\left(N_T(\varepsilon)\right)$$

$$\geqslant K \sup_{0<\varepsilon<\triangle(T)} \varepsilon\sqrt{\log N_T(\varepsilon)} \geqslant K \limsup_{\varepsilon\to 0} \varepsilon\sqrt{\log N_T(\varepsilon)},$$

where K is the constant of Lemma 3.1.

§2. Fernique's Theorem. Sudakov minorization tells us that if

$$\limsup_{\varepsilon\to 0} \varepsilon\sqrt{\log N_T(\varepsilon)} = \infty,$$

then X does not have continuous trajectories. Whereas Dudley's theorem [Theorem 2.6] implies that the condition

$$\int_{0+} \sqrt{\log N_T(\varepsilon)}\, \mathrm{d}\varepsilon < \infty$$

is sufficient for the continuity of X.

There is a small gap between these two conditions in the sense that there are examples where the metric entropy integral $\int_{0+} \sqrt{\log N_T(\varepsilon)}\, \mathrm{d}\varepsilon$ diverges and yet $\limsup_{\varepsilon\to 0^+} \varepsilon\sqrt{\log N_T(\varepsilon)} < \infty$ (see, for example, XXX). As it turns out, this is because neither condition is always sharp. The sharp condition is a "majorizing-measure condition," which you can find in XXX. I will not discuss that condition in this course primarily because it is difficult to verify in concrete settings. In fact, there are only a few instances where majorizing measures have been found. In addition, the Sudakov and Dudley theorems have broad utility XXX that extend beyond the particular applications that we have in mind in these lectures.

Still, it would be a pity to say nothing about the beautiful general theory. As a compromise, I will state and prove Fernique's theorem which states that, for stationary Gaussian processes, the Dudley condition is necessary as well as sufficient. Recall that if $(T\,,d)$ is a compact abelian group, then we say that $X := \{X_t\}_{t\in T}$ is *stationary* if $d(s\,,t) = \sqrt{\mathbb{E}(|X_t - X_s|^2)}$ is a function of $s-t$, equivalently, $t-s$, where I am using the additive notation for the group T in order to be be concise.

th:Fernique:1

Theorem 3.7 (Fernique, XXX). *Let $X := \{X_t\}_{t\in T}$ be a stationary, mean-zero Gaussian process, where $(T\,,d)$ is a metric abelian group; in particular, $d(s\,,t) = d(s-t\,,0)$ for all $s, t \in T$. Then, for all denumerable sets $S \subset T$,*

$$\mathbb{E}\left[\max_{s\in S} X_s\right] \geqslant \frac{K^2}{16}\int_0^{\triangle(S)} \sqrt{\log N_S(\varepsilon)}\, \mathrm{d}\varepsilon,$$

where K is the constant of Lemma 3.1.

One can prove Fernique's theorem, fairly readily, using the following improvement of Sudakov minorization [Proposition 3.6], due to Talagrand XXX.

pr:Talagrand

Proposition 3.8 (Talagrand, XXX). *Let $X := \{X_t\}_{t\in T}$ be any mean-zero Gaussian process, indexed by a general compact $(T\,,d)$, and consider $A\subset T$, a non-empty ε-packing of T for some $\varepsilon>0$. Then,*

$$\mathbb{E}\left[\max_{t\in A} X_t\right] \geqslant \tfrac12 \varepsilon\mu(|A|) + \min_{s\in A}\mathbb{E}\left(\max_{t\in B(s,K\varepsilon/8)} X_t\right)$$
$$\geqslant \frac{K}{2}\varepsilon\sqrt{\log|A|} + \min_{s\in A}\mathbb{E}\left(\max_{t\in B(s,K\varepsilon/8)} X_t\right),$$

where K was defined in Lemma 3.1.

Proof. I will present essentially the original proof of Talagrand XXX, since it is straightforward. Marcus and Rosen XXX have devised a clever argument which yields slightly better constants, but their argument is more involved.

By considering $\{X_t\}_{t\in A}$, it suffices to consider the case that $T=A$ and $d(s\,,t)\geqslant\varepsilon$ for all $s,t\in T$. If $|T|=1$, then the proposition states that $0\geqslant 0$. Therefore, we may consider only the case that $|T|\geqslant 2$.

Define, for all $t\in T$ and $r>0$,

$$Y_t(r) := \max_{s\in B(t,r)}(X_s - X_t) = \max_{s\in B(t,r)} X_s - X_t.$$

Since $\max_{s\in B(t,r)}\mathbb{E}(|X_s-X_t|^2)\leqslant r^2$, the Borel, Sudakov–Tsirelson inequality (Theorem 2.1, page 55) implies that

$$\max_{t\in T}\mathbb{P}\left\{|Y_t(r)-\mathbb{E}[Y_t(r)]|\geqslant\lambda\right\}\leqslant 2\exp\left(-\frac{\lambda^2}{2r^2}\right),$$

for all $\lambda>0$. In particular,

$$\mathbb{P}\left\{\max_{t\in T}|Y_t(r)-\mathbb{E}[Y_t(r)]|\geqslant\lambda\right\}\leqslant 2|T|\exp\left(-\frac{\lambda^2}{2r^2}\right)\wedge 1,$$

and hence,

$$\mathbb{E}\left(\max_{t\in T}|Y_t(r)-\mathbb{E}[Y_t(r)]|\right)\leqslant\int_0^\infty\left[2|T|\exp\left(-\frac{\lambda^2}{2r^2}\right)\wedge 1\right]\mathrm{d}\lambda$$
$$= r\sqrt{2\log(2|T|)} + 2|T|\int_{r\sqrt{2\log(2|T|)}}^\infty\exp\left(-\frac{\lambda^2}{2r^2}\right)\mathrm{d}\lambda$$
$$= r\sqrt{2\log(2|T|)} + 2r|T|\sqrt{2\pi}\,\mathbb{P}\left\{U>\sqrt{2\log(2|T|)}\right\},$$

where U has a $\mathrm{N}(0\,,1)$ distribution. Define

$$V := \max_{t\in T} |Y_t(r) - \mathbb{E}[Y_t(r)]|\,.$$

Since $\mathbb{P}\{U > u\} \leqslant \frac{1}{2}\exp(-u^2/2)$ [see proof of Lemma 2.4], the preceding and the triangle inequality together yield

$$\mathbb{E}(V) \leqslant r\sqrt{2\log(2|T|)} + r\sqrt{\frac{\pi}{2}} \leqslant 4r\sqrt{\log|T|},$$

thanks to the assumption that $|T| \geqslant 2$. Since $Y_t(r) \geqslant \mathbb{E}[Y_t(r)] - V$, the definition of $Y_t(r)$ yields

$$\max_{s\in B(t,r)} X_s \geqslant X_t + \mathbb{E}\left[\max_{s\in B(t,r)} X_s\right] - V \quad \text{a.s. for all } t\in T \text{ and } r>0.$$

Maximize over all $t\in T$ and take expectations to find that

$$\begin{aligned}\mathbb{E}\left[\max_{t\in T}\max_{s\in B(t,r)} X_s\right] &\geqslant \mathbb{E}\left[\max_{t\in T} X_t\right] + \min_{t\in T}\mathbb{E}\left[\max_{s\in B(t,r)} X_s\right] - 4r\sqrt{\log|T|}\\ &\geqslant \varepsilon\mu(|T|) + \min_{t\in T}\mathbb{E}\left[\max_{s\in B(t,r)} X_s\right] - \frac{4r}{K}\mu(|T|).\end{aligned}$$

We have appealed to Sudakov's inequality [Proposition 3.3] to bound $\mathbb{E}[\max_{t\in T} X_t]$ and Lemma 3.1 to bound $4r\sqrt{\log|T|}$. Set $r := K\varepsilon/8$ to deduce the lemma. □

lem:Fernique:comb

Lemma 3.9. *Suppose $(T\,,d)$ is a compact abelian group, and $d(s\,,t) = d(s-t\,,0)$ for all $s,t\in T$. Then, $N_T(\alpha) \leqslant N_T(\beta)\cdot N_{B(t_0,\beta)}(\alpha)$ for all $0<\alpha<\beta$ such that $B(t_0\,,\beta)\subset T$.*

Proof. Observe that, regardless of the respective values of $\beta>\alpha>0$, the stationarity of X implies that $N_{B(t,\beta)}(\alpha)$ does not depend on $t\in T$. Let $K := N_T(\beta)$ so that we can find $t_1,\ldots,t_K\in T$ such that the balls $B(t_i\,,\beta)$ cover T. We can cover every ball $B(t_i\,,\beta)$ by $L := N_{B(t_1,\beta)}(\alpha)$-many balls of radius α. Therefore, we can certainly cover T with KL-many balls of radius α. By the minimality property of the covering number $N_T(\alpha)$, this implies that $N_T(\alpha)\leqslant KL$, which is the lemma. □

Proof of Theorem 3.7. Throughout the proof, we may [and will] assume without loss of generality that $S=T$ is a countable set.

Let K denote the constant of Lemma 3.1 and recall that $K<1$ [Remark 3.2], and define

$$R := \frac{8}{K} > 1.$$

Choose and fix some $t_0\in T$ and observe that $T = B(t_0\,,R^m)$, where $m\in\mathbb{Z}$ is the unique integer defined so that

$$R^m \geqslant \triangle(T) > R^{m-1}.$$

With this observation in mind, we deduce from Talagrand's inequality [Proposition 3.8] that

$$\mathbb{E}\left[\max_{t\in B(t_0,R^m)} X_t\right] \geqslant \frac{K}{2}R^{m-1}\sqrt{\log P_T\left(R^{m-1}\right)} + \min_{s\in T}\mathbb{E}\left[\max_{t\in B(s,R^{m-1})} X_s\right].$$

By stationarity, $\mathbb{E}[\max_{B(s,\varepsilon)} X]$ does not depend on $s \in T$. Therefore, Lemma 3.5 implies that

$$\begin{aligned}\mathbb{E}\left[\max_{t\in B(t_0,R^m)} X_t\right] &\geqslant \frac{K}{2}R^{m-1}\sqrt{\log N_T\left(R^{m-1}\right)} + \mathbb{E}\left[\max_{t\in B(t_0,R^{m-1})} X_s\right]\\ &= \frac{K}{2}R^{m-1}\left[\sqrt{\log N_T\left(R^{m-1}\right)} - \sqrt{\log N_T\left(R^{m}\right)}\right] + \mathbb{E}\left[\max_{t\in B(t_0,R^{m-1})} X_s\right],\end{aligned}$$

since $N_T(R^m) = 1$. We are set up nicely to carry out an induction argument: Appeal to Talagrand's inequality [Proposition 3.8] once again to see that

$$\begin{aligned}\mathbb{E}\left[\max_{t\in B(t_0,R^{m-1})} X_s\right] &\geqslant \frac{K}{2}R^{m-2}\sqrt{\log N_{B(t_0,R^{m-1})}\left(R^{m-2}\right)} + \mathbb{E}\left[\max_{t\in B(t_0,R^{m-2})} X_s\right]\\ &\geqslant \frac{K}{2}R^{m-2}\sqrt{\log N_T\left(R^{m-2}\right) - \log N_T\left(R^{m-1}\right)} + \mathbb{E}\left[\max_{t\in B(t_0,R^{m-2})} X_s\right],\end{aligned}$$

by Lemma 3.9. The concavity of $f(x) := \sqrt{x}$ implies that $\sqrt{\log a - \log b} \geqslant \sqrt{\log a} - \sqrt{\log b}$ for all $a > b \geqslant 1$. Define

$$J_k := \mathbb{E}\left[\max_{s\in B(t_0,\, R^k)} X_s\right] \qquad [k \in \mathbb{Z}].$$

We apply the preceding, inductively, in order to find that

$$\begin{aligned}\mathbb{E}\left[\max_{s\in T} X_s\right] &= J_m\\ &\geqslant \frac{K}{2}\sum_{j=0}^{1} R^{m-j-1}\left[\sqrt{\log N_T\left(R^{m-j-1}\right)} - \sqrt{\log N_T\left(R^{m-j}\right)}\right] + J_{m-1}\\ &\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots\\ &\geqslant \frac{K}{2}\sum_{j=0}^{L} R^{m-j-1}\left[\sqrt{\log N_T\left(R^{m-j-1}\right)} - \sqrt{\log N_T\left(R^{m-j}\right)}\right] + J_{-m-L-1},\end{aligned}$$

for every integer $L \geqslant 1$. Because T is a finite set, we can choose the integer $L \geqslant 1$ large enough to ensure that $B(t_0\,, R^{m-L-1}) = \{t_0\}$. Since $J_{-m-L-1} = 0$ and $\log N_T(R^{m-j}) = 0$ for all $j > L$, we can deduce the following:

$$\mathbb{E}\left[\max_{s\in T} X_s\right] \geqslant \frac{K}{2}\sum_{j=0}^{\infty} R^{m-j-1}\left[\sqrt{\log N_T\left(R^{m-j-1}\right)} - \sqrt{\log N_T\left(R^{m-j}\right)}\right]$$

$$\geqslant \frac{K}{2}\sum_{j=0}^{\infty}\left[R^{m-j-1} - R^{m+j+2}\right]\sqrt{\log N_T\left(R^{m-j-1}\right)}$$

$$= \frac{K}{2R}\sum_{j=0}^{\infty}\int_{R^{m-j-1}}^{R^{m-j}}\sqrt{\log N_T\left(R^{m-j-1}\right)}\,\mathrm{d}\varepsilon \geqslant \frac{K}{2R}\int_0^{R^m}\sqrt{\log N_T(\varepsilon)}\,\mathrm{d}\varepsilon.$$

The result follows since $R^m \geqslant \triangle(T)$. □