Printed: 19 April 2024 at 11:50

Majorizing::Majorizing

Chapter 12

Majorizing measures for gaussian processes

SECTION 12.1 *introduces the functional that underlies some of the remarkable properties of centered gaussian processes with bounded sample paths.*
SECTION 12.2 *identifies the special property of gaussian processes—a strengthened version of the* SUDAKOV *minoration—that plays a key role in the development of lower bounds that complement the chaining upper bounds.*
SECTION 12.3 *constructs nested partitions with weights for* DOOB-SEPARABLE *gaussian processes with bounded sample paths.*
SECTION 12.4 *describes the effect on the majorizing measure lower bound if a centered gaussian processes has uniformly continuous sample paths.*
SECTION 12.5 *constructs admissible partitions for* DOOB-SEPARABLE *gaussian processes with bounded sample paths.*
SECTION 12.6 *briefly describes why a thorough understanding of the gaussian case is a good starting point for further exploration of Talagrand's amazing discoveries.*

12.1 Introduction

Majorizing::S:intro

Broadly speaking, Chapter 11 established the equivalence of three methods for deriving upper bound and oscillation control for stochastic processes whose increments are controlled by an ORLICZ norm for a Ψ in $\mathfrak{Y}_{\exp}$. It was also noted that there are companion lower bounds for gaussian processes, a fact that the current Chapter focuses on.

Throughout the Chapter, $X := \{X_t : t \in T\}$ will always be a centered, DOOB-SEPARABLE, gaussian process with bounded sample paths. As shown by Problem [1], the assumption about the sample paths implies that $\mathbb{P}\sup_{t\in T} X_t$ is finite. Consequently, the functional

\E@ F.def <1> $$\mathfrak{F}(A) := \mathbb{P}\sup\nolimits_{t\in A} X_t \qquad \text{for } A \subset T$$

is well-defined and finite.

The index set T will always by equipped with the natural metric, $d(s,t) := \|X_s - X_t\|_2$. The diameter of T, the supremum of $d(s,t)$ over all pairs of points s and t in T, will always be denoted by $\mathrm{diam}(T)$ or abbreviated to D. As noted in Section 11.3, there is no loss of generality in assuming that T is countably infinite.

Remark. In the greatest generality, d would be a semi-metric if we allowed the possibility that $X_s = X_t$ almost surely, for some pair with $s \neq t$. Such minor complications could always be avoided by partitioning X into equivalence classes then selecting only one representative from each class.

The functional $\mathfrak{F}$ is the key to the whole analysis. It works its magic through a strengthened form of the SUDAKOV minoration (Lemma <2>) that captures everything we need to know about gaussianity. With just that Lemma in hand we can recursively construct two slightly different nested sequences of finite partitions $\mathbb{A} = \{\mathcal{A}_i : i \in \mathbb{N}_0\}$ of T. The easier construction, in Section 12.3, gives weighted partitions like those in Section 11.5; the other, in Section 12.5, involves some extra subtleties in building admissible sequences like those in Section 11.7.

Of course there is some redundancy in proving two separate equivalences with the existence of a majorizing measure (MM). However, I think it is worthwhile to see both constructions, as a way of comparing the advantages of (and the technical difficulties involved in) both approaches.

12.2 A gaussian growth property

Majorizing::S:superSud

Recall the following two results from Section 6.1.

GG1 (SUDAKOV minoration) If $(W_1, W_2, \dots, W_k)$ has a centered normal distribution with $\mathbb{P}|W_i - W_j|^2 \geq \delta^2$ for all distinct i and j, then $\mathbb{P}\max_{j \leq k} W_j \geq C_{\text{sud}} \delta \mathcal{L}(k)$ with C_{sud} a universal (positive) constant and $\mathcal{L}(k) := (\log k)^{1/2}$.

GG2 (SUBGAUSSIAN max) Suppose $M := \sup_{t \in B} Y_t$ where $\{Y_t : t \in B\}$ is a DOOB-SEPARABLE gaussian process. If both $\mathbb{P}M$ and $\sigma^2 := \sup_{t \in B} \mathrm{var}(Y_t)$ are finite then $M - \mathbb{P}M$ has a subgaussian distribution:

$$\mathbb{P}e^{\lambda(M - \mathbb{P}M)} \leq \exp(\sigma^2 \lambda^2 / 2) \qquad \text{for all real } \lambda.$$

Remark. GG1 also proves that T is totally bounded if $\mathfrak{F}(T)$ is finite.

Majorizing::superSud <2> **Lemma.** *If $\{\xi_1, \dots, \xi_k\}$ is a δ-separated subset of T and $B_j \subseteq B[\xi_j, \theta\delta]$ for each j, then*

\E@ superSud.lower <3>
$$\mathfrak{F}\left(\bigcup\nolimits_{j \leq k} B_j\right) \geq C_\theta \delta \mathcal{L}(k) + \min\nolimits_{j \leq k} \mathfrak{F}(B_j) \qquad \text{where } \mathcal{L}(k) := \sqrt{\log k},$$

□ *with $C_\theta := C_{sud} - 4\theta$, which is > 0 if θ is small enough.*

Remark. I write ξ_j, instead of the more natural t_j, to avoid subsequent confusion with the points in a chaining framework.

Proof. The assertion of the Lemma is trivially true when k equals 1, so assume $k \ge 2$.

Define $W_j := X_{\xi_j}$ and $Y_t := X_t - W_j$ for $t \in B_j$ and

$$M_j := \sup_{t\in B_j} Y_t = \sup_{t\in B_j} X_t - W_j.$$

Notice that $\mathbb{P}M_j = \mathbb{P}\sup_{t\in B_j} X_t - \mathbb{P}W_j = \mathfrak{F}(B_j)$ and

$$\operatorname{var}(Y_t) = \|X_t - W_j\|_2^2 = d(t,\xi_j)^2 \le (\theta\delta)^2 \qquad \text{for all } t \text{ in } B_j.$$

Similarly, from $\mathbb{P}|W_i - W_j|^2 = d(\xi_i,\xi_j)^2 > \delta^2$ for $i \ne j$ we get

$$\mathbb{P}\max_{i\le k} W_j \ge C_{\text{sud}}\delta\mathcal{L}(k) \qquad \text{by GG1.}$$

The idea behind inequality <3> is that $\sup_{t\in B_j} X_t = W_j + M_j$ should be close to the corresponding $W_j + \mathfrak{F}(B_j)$ for each j, by virtue of GG2. The supremum of X_t over $\cup_j B_j$ should then be close to the maximum of $W_j + F(B_j)$. More precisely,

$$\begin{aligned}\sup_{t\in\cup_j B_j} X_t &= \max_j \sup_{t\in B_j} X_t\\ &= \max_j\big(W_j + \mathfrak{F}(B_j) + M_j - \mathfrak{F}(B_j)\big)\\ &\ge \max_j\big(W_j + \min_j \mathfrak{F}(B_j) - |M_j - \mathfrak{F}(B_j)|\big)\\ &\ge \max_j W_j + \min_j \mathfrak{F}(B_j) - \max_j |M_j - \mathfrak{F}(B_j)|.\end{aligned}$$

Taking expected values then invoking GG1 we get <3> if $\mathbb{P}\max_j |M_j - \mathfrak{F}(B_j)|$ is small enough. To control that term the usual JENSEN inequality trick suffices: by GG2, for $\lambda > 0$,

$$\begin{aligned}\exp\left(\lambda\mathbb{P}\max_j |M_j - \mathfrak{F}(B_j)|\right) &\le \sum\nolimits_j \mathbb{P}\exp\left(\lambda|M_j - \mathfrak{F}(B_j)|\right)\\ &\le \sum\nolimits_j \mathbb{P}\exp\left(\lambda(M_j - \mathbb{P}M_j)\right) + \mathbb{P}\exp\left(-\lambda(M_j - \mathbb{P}M_j)\right)\\ &\le 2k\exp\left(\lambda^2(\theta\delta)^2/2\right).\end{aligned}$$

Take logs, divide through by λ, then minimize by choosing $\lambda = \sqrt{2}\mathcal{L}(2k)/(\theta\delta)$ to get, for $k \ge 2$,

$$\mathbb{P}\max_j |M_j - \mathfrak{F}(B_j)| \le 2\theta\delta\sqrt{2\log(2k)} \le 4\theta\delta\mathcal{L}(k)$$

and hence

$$F(\cup_j B_j) \ge C_{\text{sud}}\delta\mathcal{L}(k) + \min_{j\le k} F(B_j) - 4\theta\delta\mathcal{L}(k).$$

□ We just need to make sure that $4\theta < C_{\text{sud}}$.

12.3 From functional to nested partitions with weights

ing::S:fnal-to-Wpartition

The main result (Theorem <5>) in this Section uses the functional $\mathfrak{F}$ from <1> to construct suitable weighted partitions by an argument adapted from Talagrand (1992) via the exposition of Ledoux (1996, Chapter 6). It complements the main Theorem from Section 11.5, which I wrote using a weighted chaining framework, but which could also have been written (as in Theorem <4>) using nested partitions:

Majorizing::Ledoux

<4> **Theorem.** *(adapted from Chapter 11) Let $\{X_t : t \in T\}$ be a* DOOB-SEPARABLE *process indexed by a totally bounded metric space (T, d) for which:*

(i) $\|X_s - X_t\|_\Psi \le K_0 d(s,t)$ *where* $\Psi = e^g - 1$ *is an* ORLICZ *function in* $\mathcal{Y}_{\exp}$.

(ii) There is a nested sequence $\mathbb{A} = \{\mathcal{A}_i : i \in \mathbb{N}_0\}$ *of finite partitions with* $\mathcal{A}_0 = \{T\}$ *and a sequence* $\{\delta_i\}$ *of positive constants, for which and* $\max_{A\in\mathcal{A}_i} \operatorname{diam}(A) \le 2\delta_i$.

(iii) For each i there exists a weight function $w_i : \mathcal{A}_i \to (0,1]$.

Let $\{(T_i, \ell_i) : i \in \mathbb{N}_0\}$ be the corresponding chaining framework, with the finite set T_i consisting of a single point from each A in $\mathcal{A}_i$. Then, for $m > k$, the expected value $\mathbb{P}\max_{t\in T_m} |X(t) - X(L_k t)|$ *is bounded by*

$$C \max\nolimits_{t\in T_m} \sum\nolimits_{i=k+1}^{m} \delta_i \Psi^{-1}\left(1/w_i(A_i(t))\right) + \sum\nolimits_{i=k+1}^{m} \delta_i w_i(\mathcal{A}_i),$$

where $t \in A_i(t) \in \mathcal{A}_i$ *and* $w_i(\mathcal{A}_i) := \sum_{A\in\mathcal{A}_i} w_i(A)$ *and* $C = C(K_0, \Psi)$ *is a constant.* □

I also noted that if $w_i(\mathcal{A}_i) = O(R^i)$ for some R and $\delta_i = O(\theta^i)$ for some θ with $R\theta < 1$ then $\sum_i \delta_i w_i(\mathcal{A}_i)$ converges. For the converse the $\{\delta_i\}$ sequence has that form.

ajorizing::fnal.to.Ledoux

<5> **Theorem.** *Suppose $\{X_t : t \in T\}$ is a* DOOB-SEPARABLE *centered gaussian process for which* $\mathfrak{F}(T) := \mathbb{P}\sup_{t\in T} X_t$ *is finite. Define* $\theta := \min(C_{sud}/5, 1/3)$ *and* $\delta_i := D\theta^i$. *Then there exists a nested sequence of finite partitions* $\mathbb{A} = \{\mathcal{A}_i : i \in \mathbb{N}_0\}$ *with* $\mathcal{A}_0 = \{T\}$ *and* $\max_{A\in\mathcal{A}_i} \operatorname{diam}(A) \le 2\delta_i$, *and weight functions* $w_i : \mathcal{A}_i \to (0,1]$ *with* $w_i(\mathcal{A}_i) := \sum_{A\in\mathcal{A}_i} w_i(A) \le 2^i$, *for which*

$$C\mathfrak{F}(T) \ge \sum\nolimits_{i\in\mathbb{N}}^{\infty} \delta_i \sqrt{\log(1/w_i(A_i(t))} \qquad \text{for each } t \in T,$$

where C is a constant that depends on θ. □

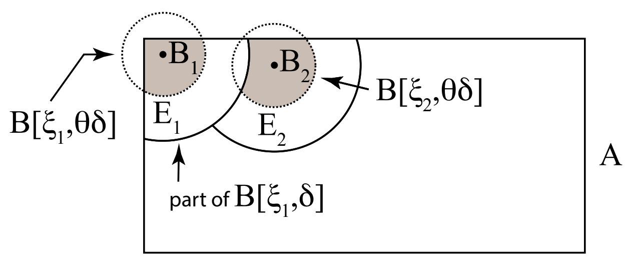
Remark. As noted at the start of Section 11.6, the existence of such a weighting scheme implies existence of a MM μ on T for which $C_1\mathfrak{F}(T) \ge \sup_{t\in T} \int_0^{\operatorname{diam}(T)} \Psi_2^{-1}\left(1/B[t,r]\right) dr$ with C_1 a universal constant.

The proof of Theorem <5> consists of recursive appeals to a greedy procedure (WPARTITION, with the leading W to suggest weights) that partitions a set into smaller pieces (of diameter at most 2δ) using a localized version of the functional,

\E@ localF <6>
$$f(S,r) := \sup_{s\in S} \mathfrak{F}\left(S\cap B[s,r]\right) \qquad \text{for each subset } S \text{ of } T.$$

The starting point is a set A with weight W. The η provides some wiggle room when the supremum is not achieved at some s in S.

procedure WPARTITION(A, δ, W, η)
  Initialize: $j = 1$, $S_1 = A$.
  loop
    Find a ξ_j in S_j for which the set $B_j := S_j \cap B[\xi_j, \theta\delta]$
      satisfies $\mathfrak{F}(B_j) \geq f(S_j, \theta\delta) - \eta$.
    Define $E_j := S_j \cap B[\xi_j, \delta]$.   % The jth child of A.
    Attach weight $w(E_j) := W/j^2$ to the set E_j.
    Define $S_{j+1} := S_j \backslash E_j$.
    If S_{j+1} is empty **then**
      Exit the loop.
    else
      Increase j by 1 (and continue with the next iteration).
    end If
  end loop
end procedure



After the loop exits we are left with a δ-separated set of points $\xi_1, \dots, \xi_J$ from A (so that $J \leq \text{PACK}(\delta, A) < \infty$), a weighted partition $E_1, \dots, E_J$ of A, and sets $S_{j+1} = S_j \backslash E_j$ with $A = S_1 \supset S_2 \supset \cdots \supset S_J = E_J$ for which

$$\xi_j \in B_j := S_j \cap B[\xi_j, \theta\delta] \subset E_j := S_j \cap B[\xi_j, \delta].$$

The greed ensures that

$$\eta + \mathfrak{F}(B_j) \geq f(S_j, \theta\delta) \geq \mathfrak{F}(S_j \cap B[s, \theta\delta]) \qquad \text{for each } s \text{ in } S_j.$$

Most importantly, for $1 \leq \alpha \leq J$, Lemma <2> gives

\E@ superSud2 <7>
$$\begin{aligned} \mathfrak{F}\left(\cup_{j\leq\alpha} B_j\right) &\geq C_\theta \delta \sqrt{\log \alpha} + \min\nolimits_{j\leq\alpha} \mathfrak{F}(B_j) \\ &\geq 2^{-1/2} C_\theta \delta \sqrt{\log \alpha^2} + \min\nolimits_{j\leq\alpha} f(S_j, \theta\delta) - \eta \\ &= C_1 \delta \sqrt{\log(W/w(E_\alpha))} + \sup\nolimits_{\xi\in S_\alpha} \mathfrak{F}\left(S_\alpha \cap B[\xi, \theta\delta]\right) - \eta \end{aligned}$$

where $C_1 := 2^{-1/2}C_\theta$. This inequality will provide an upper bound for $\mathfrak{F}(E'')$ for a partitioning set E'' generated in a later iteration of WPARTITION.

> **Remark.** For <7> to hold when α equals 1 then we must have $w(E_1) = W$. The choice $w(E_j) = W/j^2$ then gives a bound $\sum_j w(E_j) \le 2W$ (Compare with $\sum_{j\in\mathbb{N}} j^{-2} \approx 1.65$.) That was my reason for allowing $w_i(T_i) > 1$ in Section 11.5 and for inserting the $1/3$ into the definition of θ. Talagrand (1992, p. 122)'s argument seems to corresponds to weights $w(E_j) = W/(j+1)^2$ and Ledoux (1996, p. 243)'s to $w(E_j) = W/(2j^2)$, presumably to ensure $\sum_j w(E_j) \le W$.

Proof (of Theorem <5>, based on Ledoux 1996, pp. 242–244). For reasons that will become clearer towards the end of the proof, define $\eta_i := \mathfrak{F}(T)/2^i$ for $i \in \mathbb{N}$.

The construction consists of recursive appeals to WPARTITION, starting from $\mathcal{A}_0 = \{T\}$ and $w_0(T) = 1$.

- Let $\mathcal{A}_1$ consist of the sets $E_1, E_2, \dots, E_J$, with weights $w_1(E_j) = j^{-2}$, generated by WPARTITION$(T, \delta_1, w_0(T), \eta_1)$.
- Then partition each A in $\mathcal{A}_1$ by WPARTITION$(A, \delta_2, w_1(A), \eta_2)$.
- And so on. In general, WPARTITION$(A, \delta_{i+1}, w_i(A), \eta_{i+1})$ applied to each A in $\mathcal{A}_i$ generates $\mathcal{A}_{i+1}$.

By construction, $w_{i+1}(\mathcal{A}_{i+1}) \le 2w_i(\mathcal{A}_i)$ and hence $w_i(\mathcal{A}_i) \le 2^i$ for each i.

Consider a fixed t in T. To simplify notation write A_j instead of $A_j(t)$ for $j = 0, 1, \dots$. Suppose $A_{i+1} = E'_\alpha$ (the αth child of A_i) and $A_{i+2} = E''_\beta$ (the βth child of A_{i+1}) as generated by

$$
\begin{aligned}
&\text{WPARTITION}(A_i, \delta_{i+1}, w_i(A_i), \eta_{i+1}) && \text{with } \delta_{i+1} = D\theta^{i+1} = \theta\delta_i, \\
&\text{WPARTITION}(A_{i+1}, \delta_{i+2}, w_{i+1}(A_{i+1}), \eta_{i+2}) && \text{with } \delta_{i+2} = \theta^2\delta_i.
\end{aligned}
$$

The primes identify quantities involved in the construction of $\mathcal{A}_{i+1}$ and the double primes identify quantities involved in the construction of $\mathcal{A}_{i+2}$. (Some such notation is needed, for example, to distinguish between the E_1 that is a member of $\mathcal{A}_{i+1}$ and the E_1 that is a member of $\mathcal{A}_{i+1}$.)

The construction gives the following pattern of inclusions and inequalities:

$$
\begin{array}{llllll}
A_i \supset S'_\alpha \supset & A_{i+1} = & E'_\alpha & \supset & B'_\alpha = S'_\alpha \cap B[\xi'_\alpha, \theta^2\delta_i] & \\
 & & \cup & & & \\
 & & S''_\beta & \supset & A_{i+2} = E''_\beta & = S''_\beta \cap B[\xi''_\beta, \theta^2\delta_i];
\end{array}
$$

and

$$
\begin{aligned}
&\mathfrak{F}A_i \ge C_1\delta_{i+1}\sqrt{\log\left(w_i(A_i)/w_{i+1}(E'_\alpha)\right)} + \mathfrak{F}B'_\alpha && \text{by } \langle 7\rangle; \\
&\eta_{i+1} + \mathfrak{F}(B'_\alpha) \ge \sup_{\xi\in S'_\alpha} \mathfrak{F}(S'_\alpha \cap B[\xi, \theta^2\delta_i]) \ge \mathfrak{F}E''_\beta && \text{because } \xi''_\beta \in S''_\beta \subset S'_\alpha.
\end{aligned}
$$

It follows that

\E@ grandchild

$$
\begin{aligned}
&\eta_{i+1} + \mathfrak{F}\left(A_i\right) \ge C_1\delta_{i+1}\sqrt{\rho_{i+1}} + \mathfrak{F}\left(A_{i+2}\right) \\
&\qquad\qquad \text{where } \rho_{i+1} := \log\left(w_i(A_i)/w_{i+1}(A_{i+1})\right).
\end{aligned}
\tag{8}
$$

If you write out <8> for $i = 0, 2, \ldots, 2m$, add, cancel a bunch of terms, then let m tend to infinity (and remember that $A_0(t) = T$) you should get

$$\mathfrak{F}(T) + \sum\nolimits_{j \text{ odd}} \eta_j \ge C_1 \sum\nolimits_{j \text{ odd}} \delta_j \sqrt{\rho_j}.$$

A similar summation over $i = 1, 3, 5, \ldots$ gives

$$\mathfrak{F}(A_1) + \sum\nolimits_{j \text{ even}} \eta_j \ge C_1 \sum\nolimits_{j \text{ even}} \delta_j \sqrt{\rho_j}.$$

Replace $\mathfrak{F}(A_1)$ by the larger $\mathfrak{F}(T)$ then add the sums for even and odd j to deduce (remembering that $\eta_i = \mathfrak{F}(T)/2^i$) that

$$3\mathfrak{F}(T) = 2\mathfrak{F}(T) + \sum\nolimits_{i\in\mathbb{N}} \eta_j \ge C_1 \sum\nolimits_{i\in\mathbb{N}} \delta_i \sqrt{\rho_i}.$$

This inequality does not quite match the inequality asserted by Theorem <5>, which has $\sqrt{\log(1/w_i(A_i(t))}$ in place of $\sqrt{\rho_i(t)}$. That problem is easily fixed because $w_0(A_0) = w_0 T = 1$ and

$$\sum\nolimits_{j=1}^{i} \rho_j(t) = \log\left(\frac{w_0(T)}{w_1(A_1)}\frac{w_1(A_1)}{w_2(A_2)}\cdots\frac{w_{i-1}(A_{i-1})}{w_i(A_i)}\right) = \log\left(\frac{1}{w_i(A_i)}\right).$$

Using the inequality $\left(\sum_i a_i\right)^{1/2} \le \sum_i a_i^{1/2}$ for $a_i \in \mathbb{R}^+$ and making an interchange in the order of summation we then get

$$\begin{aligned}\sum\nolimits_{i\in\mathbb{N}} \delta_i \sqrt{\log(1/w_i(A_i(t))} &= \sum\nolimits_{i\in\mathbb{N}} \delta_i \sqrt{\sum\nolimits_{j\in\mathbb{N}} \{j \le i\}\rho_j(t)} \\ &\le \sum\nolimits_{j\in\mathbb{N}} \sqrt{\rho_j(t)} \sum\nolimits_{i\in\mathbb{N}_0} \delta_i\{j \le i\} \\ &= \sum\nolimits_{j\in\mathbb{N}} \sqrt{\rho_j(t)}\, \delta_j/(1-\theta).\end{aligned}$$

□ It is now just a matter of juggling some constants.

What a beautiful argument! Fernique was correct: existence of a MM is both necessary and sufficient for a DOOB-SEPARABLE, centered gaussian process to have bounded sample paths.

*12.4 Uniformly continuous gaussian sample paths

Majorizing::S:cts

As you saw in Section 10.6, if the increments of a DOOB-SEPARABLE process are controlled by an ORLICZ norm $\|\cdot\|_\Psi$ then a sufficient condition for the process to have uniformly continuous sample paths is the existence of a MM μ for which

$$\sup_{t\in T} \int_0^\delta \Psi^{-1}(1/\mu B[t,r])\, dr \to 0 \qquad \text{as } \delta \to 0.$$

For gaussian processes this condition, with the ORLICZ function $\Psi(x)$ specialized to $\Psi_2(x) := \exp(x^2) - 1$, is also necessary.

Majorizing::MM.cts <*9*> **Theorem.** *Let $X = \{X_t : t \in T\}$ be a centered* GAUSSIAN *process. Suppose that T is totally bounded under the metric $d(s,t) := \|X_s - X_t\|_2$ and that (almost all) the sample paths of X are d-uniformly continuous:*

$$\text{OSC}(\delta, X, T) := \sup\{|X_s - X_t| : d(s,t) < \delta\} \to 0 \qquad \text{almost surely as } \delta \to 0.$$

Then there exists a MM μ on T for which

$$\sup_{t\in T} \int_0^\delta \Psi_2^{-1}(1/\mu B[t,r])\,dr \to 0 \qquad \text{as } \delta \to 0.$$

Proof. The proof will combine the virtues of packing numbers and majorizing measures.

A uniformly continuous functions on a totally bounded space must be bounded. Thus $\mathfrak{F}(T) := \mathbb{P}\sup_{t\in T}|X_t|$ is finite. Define $\delta_k := 2^{-k}$. By dominated convergence (with $2\sup_{t\in T}|X_t|$ as the dominating function) we have $\beta_k := \mathbb{P}\text{OSC}(\delta_k, X, T) \to 0$ as $k \to \infty$.

Total boundedness implies that $\text{PACK}(\delta, T)$ is finite for each positive δ. As in Section 10.4, there is then a nested sequence of finite sets $\{T_k : k \in \mathbb{N}\}$ with T_k a δ_k-packing set, $|T_k| = N_k \le \text{PACK}(\delta_k, T)$. Also there exists a map $\tau_k : T \to T_k$ with $\sup_{s\in T} d(s, \tau_k(s)) \le \delta_k$. The sets $\{s \in T : \tau_k(s) = t\}$ for $t \in T_k$ form a partition $\mathcal{E}_k$ of T with each E in $\mathcal{E}_k$ containing exactly one point, $\tau_k(E)$, from T_k.

From the fact that $\sup_{s\in E} d(s, \tau_k(E)) \le \delta_k$ we get

$$\mathfrak{F}(E) := \mathbb{P}\sup_{t\in E} X_t = \mathbb{P}\sup_{t\in E}(X_t - X_{\tau_k(E)}) \le \beta_k \qquad \text{for } E \in \mathcal{E}_k.$$

Theorem <5> then provides a probability measure μ_E on E for which

\E@ muE <10>
$$\sup_{t\in E} \int_0^{\text{diam}(E)} \Psi_2^{-1}(1/\mu_E B[t,r])\,dr \le C\beta_k,$$

where C is a constant that doesn't depend on E or k.

I claim that the probability measure

$$\mu := \sum\nolimits_{k\ge1} \frac{1}{2^k N_k} \sum\nolimits_{E\in\mathcal{E}_k} \mu_E,$$

is a MM with the property asserted by the Theorem. The proof is easy.

For any given positive ϵ first choose $k = k(\epsilon)$ so that $\beta_k < \epsilon$ then choose $\delta := \epsilon/\Psi_2^{-1}(2^k N_k)$. Consider any point t in T. To simplify notation write E_k for the member of $\mathcal{E}_k$ that contains t and abbreviate μ_{E_k} to μ_k. For $r \le \delta$ use the fact that $\mu B[t,r] \ge (2^k N_k)^{-1}\mu_k B[t,r]$ and the inequality $\Psi_2^{-1}(uv) \le \Psi_2^{-1}(u) + \Psi_2^{-1}(v)$ for $u, v \ge 0$ to deduce that

$$\begin{aligned}\int_0^\delta \Psi_2^{-1}(1/\mu B[t,r])\,dr &\le \int_0^\delta \Psi_2^{-1}\left(\frac{2^k N_k}{\mu_k B[t,r]}\right)dr \\ &\le \int_0^\delta \Psi_2^{-1}(2^k N_k)\,dr + \int_0^\delta \Psi_2^{-1}(1/\mu_k B[t,r])\,dr. \\ &\le \delta\Psi_2^{-1}(2^k N_k) + C\beta_k + \delta\Psi_2^{-1}(1) \le (2+C)\epsilon.\end{aligned}$$

For the final bound it helps to consider two cases: if $\delta \le \operatorname{diam}(E)$ then <10> bounds the second integral by $C\beta_k$; if $\delta > \operatorname{diam}(E)$ then $\int_0^{\operatorname{diam}(E)} \dots dr \le C\beta_k$ and the integrand for $\int_{\operatorname{diam}(E)}^{\delta} \dots dr$ is less than $\Psi_2^{-1}(1)$ because over that range we have $B[t,r] \supset E$, implying $\mu_k B[t,r] = 1$.

To check that μ is actually a MM just note that $\int_0^D \dots dr$ is smaller than D/δ times $\int_0^\delta \dots dr$ when the integrand is a decreasing function of r. □

*12.5 From functional to admissible partitions

zing::S:fnal-to-Apartition

This Section shows how to construct suitable admissible partitions from the functional $\mathfrak{F}$. I have cobbled together a proof from several different arguments used by Talagrand (see Notes), rearranged to emphasize the similarities with the proof in my Section 12.3. My treatment corresponds roughly to Talagrand (2021, pp 47–58).

> **Remark.** The Section has gone through many versions, as each fix led to new, more subtle errors. I would be very pleasantly surprised if it were now error-free, even though I am confident that the underlying idea is sound.

rizing::fnal.to.admissible

<11> **Theorem.** *Let $\{X_t : t \in T\}$ be a* DOOB-SEPARABLE, *centered gaussian process for which $\mathfrak{F}(T) := \mathbb{P}\sup_{t\in T} X_t$ is finite. Then there exists a nested sequence $\mathbb{A} = \{\mathcal{A}_i : i \in \mathbb{N}_0\}$ of finite partitions, with $\mathcal{A}_0 = \{T\}$ and $|\mathcal{A}_i| \le n_i := 2^{2^i}$ such that, for some universal constant C and $\lambda := \sqrt{2}$,*

$$C\mathfrak{F}(T) \ge \sum\nolimits_{i\in\mathbb{N}_0}^{\infty} \operatorname{diam}\left(A_i(t)\right)\lambda^i \qquad \text{for each } t \in T,$$

□ *where $A_i(t)$ denotes the member of $\mathcal{A}_i$ that contains t.*

The construction in Section 12.3 gave a very tight control over the diameters of the individual members of $\mathcal{A}_i$, with the weights providing only an indirect control for the $|\mathcal{A}_i|$, the number of sets in $\mathcal{A}_i$. By contrast, Talagrand's generic chaining involves the strict bound $n_i = 2^{2^i}$ on the size $|\mathcal{A}_i|$ but only indirect control over the diameters of the members of the partitions.

Despite these differences, the two partitioning methods have much in common. Brutally speaking, the new procedure could be implemented by:

(i) Generate the sets $E_1, \dots, E_J$ using WPARTITION$(A, \dots)$.

(ii) Discard all the weights.

(iii) Replace E_m by $\cup_{j\ge m} E_j$ if J is larger than some magic number m. and label the new E_m as BIG. (The sets that are not BIG are called SMALL, of course.) Then discard the individual E_j sets for $j > m$.

Needless to say, it makes more sense to start from a clean definition, although the brutal description does explain why some of the arguments in this Section share similarities with the arguments in Section 12.3. The main differences comes in:

(a) The key role that is played by the BIG sets.

(b) The choice $\theta := \min\left(C_{\text{sud}}/5, (2\lambda)^{-1}\right)$. The $C_{\text{sud}}/5$ term will let us invoke Lemma <2> and the $(2\lambda)^{-1}$ will ensure that $\sum_i (\lambda\theta)^i$ converges.

(c) The use of a quantity $\rho(A) \geq \operatorname{diam}(A)/2$ for the each set in each $\mathcal{A}_i$, so that the desired inequality becomes $C\mathfrak{F}(T) \geq \sup_{t\in T}\sum_i \rho\left(A_i(t)\right)\lambda^i$.

Again the localized functional, $f(S, r) := \sup_{s\in S}\mathfrak{F}\left(S \cap B[s, r]\right)$ will guide the greedy construction of the partitions using the following procedure, which assumes that we have already assigned the value $\rho(A)$ for the the set that is to be partitioned. The parameter m plays the role of a generic n_i. Again the small η provides wiggle room below a supremum.

```
procedure APARTITION(A, m, η)
    Initialize: j = 1, S_1 = A, and define ρ := ρ(A).
    loop
        Find ξ_j in S_j for which the set B_j := S_j ∩ B[ξ_j, ρθ^2]
                satisfies 𝔉(B_j) ≥ f(S_j, ρθ^2) − η.
        If j is equal to m then
            Define E_m := S_m and ρ(E_m) := ρ and label E_m as BIG.
            Exit the loop.
        end If
        Define E_j := S_j ∩ B[ξ_j, ρθ] and ρ(E_j) := ρθ.
        Define S_{j+1} := S_j \ E_j.
        If S_{j+1} is empty then
            Exit the loop.
        end If
        Increase j by 1 (and continue with the next iteration).
    end loop
end procedure
```

For future reference, here are the most important facts about the output from APARTITION summarized in the form of a Lemma.

Majorizing::APART <12> **Lemma.** *The subset A of T, with $\rho := \rho(A)$, can be partitioned into subsets $E_1, \dots, E_J$, for $J \leq m$, such that:*

(i) If $j < m$ then E_j is SMALL with $\rho(E_j) = \rho\theta$ and $E_j \subset A \cap B[\xi_j, \rho\theta]$ for some ξ_j in A.

(ii) If $J = m$ then E_m is BIG with $\rho(E_m) = \rho$ and

\E@ BIG.Sud2 <13>
$$\mathfrak{F}(A) \geq C_\theta \rho\theta\sqrt{\log m},$$

\E@ BIG.superSud <14>
$$\eta + \mathfrak{F}(A) \geq C_\theta \rho\theta\sqrt{\log m} + f(E_m, \rho\theta^2),$$

□ *where $f(E_m, r) := \sup\{\mathfrak{F}(E_m \cap B[\xi, r]) : \xi \in E_m\}$.*

Proof. The facts about the the $\rho\theta$ separated ξ_j's and the greedily chosen B_j's are hidden inside the two inequalities for the BIG sets. The first is effectively just the SUDAKOV minoration (GC1). The second comes from Lemma <2>:

$$\begin{aligned}\mathfrak{F}(A) &\ge C_\theta(\rho\theta)\sqrt{\log m} + \min\nolimits_{j\le m}\mathfrak{F}(B_j) \qquad \text{which is } \ge C_\theta\rho\theta\sqrt{\log m}\\ &\ge C_\theta\rho\theta\sqrt{\log m} + \min\nolimits_{j\le m}\left(f(S_j,\rho\theta^2)-\eta\right)\\ &= C_\theta\rho\theta\sqrt{\log m} + f(E_m,\rho\theta^2)-\eta.\end{aligned}$$

□

Remark. Lemma <2> is invoked only when E_m is BIG. Notice that the $\mathfrak{F}(B_j)$'s need not decrease monotonely but the the $f(S_j,\rho\theta^2)$'s do. I have separated out the role for GC1 in order to emphasize the places where the full force of Lemma <2> is needed.

Proof (of Theorem <11>). Again define $\eta_i := \mathfrak{F}(T)/2^{i+1}$ for $i \in \mathbb{N}_0$ and choose $\theta := \min\left(C_{\text{sud}}/5, (2\lambda)^{-1}\right)$ as in (b) on the previous page.

Make recursive appeals to procedure APARTITION, starting from $\mathcal{A}_0 = \{T\}$ and $\rho(T) := D$. For the sake of a uniform notation, define $n_0 := n_1$. (The first step doesn't fit neatly into the general pattern because $|\mathcal{A}_0| = 1$.)

- Let $\mathcal{A}_1$ consist of the sets $E_1, E_2, \dots, E_J$, with $J \le n_0 = n_1$, generated by APARTITION(T, n_0, η_0).
- Then partition each A in $\mathcal{A}_1$ into the new sets $E_1, E_2, \dots$ generated by APARTITION(A, n_1, η_1). Each of the (at most) n_1 sets in $\mathcal{A}_1$ is partitioned into (at most) n_1 subsets to create an $\mathcal{A}_2$ consisting of at most $n_1^2 = n_2$ subsets, as required by admissibility.
- And so on: in general, APARTITION(A, n_i, η_i) applied to each A in $\mathcal{A}_i$ generates the sets for $\mathcal{A}_{i+1}$ and $|\mathcal{A}_{i+1}| \le n_i|\mathcal{A}_i| \le n_i^2 = n_{i+1}$.

As before, consider a fixed t in T and the sets $A_i := A_i(t) \in \mathcal{A}_i$ for which $t \in A_i(t)$. The sequence $\{A_i : i \in \mathbb{N}_0\}$ can consist of both SMALL sets and BIG sets. To simplify notation I'll treat $A_0 = T$ as small. Lemma <12> gives:

- If A_{i+1} is SMALL then $\rho(A_{i+1}) = \rho(A_i)\theta$ and $A_{i+1} \subset A_i \cap B[\xi, \rho(A_i)\theta]$ for some ξ in A_i.
- If A_{i+1} is BIG then $\rho(A_{i+1}) = \rho(A_i)$ and

\E@ Sud.BIG <15>
$$\mathfrak{F}(A_i) \ge C_1\rho(A_{i+1})\lambda^{i+1},$$
\E@ Ai+1.BIG <16>
$$\eta_i + \mathfrak{F}(A_i) \ge C_1\rho(A_{i+1})\lambda^{i+1} + f(A_{i+1}, \rho(A_{i+1})\theta^2),$$

where $C_1 := \theta\lambda^{-1}C_\theta\sqrt{\log 2}$.

If every A_i is SMALL then $\rho(A_i) \le D\theta^i$ for every i so that

$$\sum\nolimits_{i\in\mathbb{N}_0}^{\infty} \lambda^i\rho(A_i) \le \sum\nolimits_{i\in\mathbb{N}_0} D(\theta\lambda)^i \le 2D \le 2\sqrt{2\pi}\mathfrak{F}(T),$$

the final inequality coming from Problem [2].

Things become more interesting if some of the sets are BIG. The argument then involves a close cousin of the odd/even subsequencing idea used in the Proof of Theorem <5>. To make it easier to spot any errors I had to break the argument into a short sequence of claims followed by proofs and consequences.

Claim (a):
There can be no infinite subsequence of consecutive BIG A_j's.

Proof of (a):
Suppose A_j is big for $\alpha \le j \le \omega$, so that $\rho := \rho(A_{\alpha-1}) = \rho(A_\alpha) = \cdots = \rho(A_\omega)$. Inequality <15> with $i = \omega - 1$ and monotonicity of $\mathfrak{F}$ then imply

\E@ one.step <17>

$$\mathfrak{F}(T) \ge \mathfrak{F}(A_{\omega-1}) \ge C_1 \lambda^\omega \rho\,(A_\omega) = C_1 \lambda^\omega \rho,$$

□ an inequality that prevents ω from being too large.

Remark. Note that argument would not work if we replaced $\rho(A_i)$ by the smaller $\operatorname{diam}(A_i)/2$. It was important that $\rho(A_i) = \rho(A_\omega)$ for $\alpha - 1 \le i \le \omega$.

Consequence (a):
It follows that each stretch of consecutive BIG A_j's must be finite and separated from the next stretch of BIG's by at least one SMALL set. Moreover, one stretch of SMALL A_j's might be infinite, in which case there are no more BIG sets. The pattern could look like:

$$\begin{array}{ccccccccc}
\text{small}_0 & \text{big}_1 & \text{small}_1 & \text{big}_2 & \text{small}_2 & \text{big}_3 & \text{small}_3 & \text{big}_4 & \\
sss & bbb & sssss & bb & sss & b & sss & bb & \dots \\
 & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
 & \omega[1] & & \omega[2] & & \omega[3] & & \omega[4] &
\end{array}$$

The index $\omega[\ell]$ gives the position of the last BIG A_j in big_ℓ, the ℓth block of
□ BIG sets.

The contribution from small_0 can be handled in the same way as the case where all A_j's are SMALL.

Claim (b):
If A_j is BIG for $\alpha \le j \le \omega$ and SMALL for $\omega < j \le \beta$ then $\sum_{j=\alpha}^{\beta} \lambda^j \rho(A_j) \le C_0 \lambda^\omega \rho(A_\omega)$ for some constant C_0.

Proof of (b):

$$\sum\nolimits_{\alpha \le j \le \omega} \lambda^j \rho(A_j) = \rho(A_\omega) \sum\nolimits_{\alpha \le j \le \omega} \lambda^j \le \rho(A_\omega) \lambda^{\omega+1}/(\lambda - 1),$$
$$\sum\nolimits_{\omega < j \le \beta} \lambda^j \rho(A_j) \le \sum\nolimits_{\omega < j \le \beta} \lambda^\omega \lambda^{j-\omega} \rho(A_\omega) \theta^{j-\omega}.$$

□ The last sum is $\le \lambda^\omega \rho(A_\omega)$ because we chose θ to make $\theta\lambda \le 1/2$.

Consequence (b):
The contributions from any big_ℓ plus the following small_ℓ are bounded above by $C_0\lambda^\omega\rho(A_\omega)$ with $\omega = \omega[\ell]$. □

Thus it remains only to control the sum

\E@ sum.big.ends <18>
$$\sum\nolimits_{\ell\in\mathbb{L}} \lambda^{\omega[\ell]}\rho(A_\omega[\ell]) \qquad \text{where } \mathbb{L} := \{\ell : \text{big}_\ell \neq \emptyset\}.$$

If $|\mathbb{L}| \leq 3$ then three appeals to inequality <17> would suffice. If $|\mathbb{L}| \geq 4$ we can use a trick that is similar to the odd/even decomposition used to prove Theorem <5>.

Claim (c):
If $\ell, \ell+1, \ell+2, \ell+3 \in \mathbb{L}$ then

\E@ four.step <19>
$$\eta_{\omega[\ell]} + \mathfrak{F}(A_{\omega[\ell]}) \geq C_1\lambda^{\omega[\ell+1]}\rho(A_{\omega[\ell+1]}) + \mathfrak{F}(A_{\omega[\ell+3]}).$$

Proof of (c):
To slightly simplify notation I'll assume $\ell = 1$. The proof for general ℓ is similar.

It is just a matter of pulling together the various inequalities derived in the previous few pages. Define $i := \omega[2] - 1$ and $i' := \omega[3]$.

$$\begin{array}{ccccccccc} s\ldots & bb\underline{b} & s\ldots & & b\underline{b} & s\ldots & b\underline{b} & s\ldots & b\underline{b} \quad \ldots \\ & \omega[1] & & \omega[2] = i+1 & & \omega[3] = i' & & & \omega[4] \end{array}$$

Because there is at least one SMALL between each big block we must have $\omega[1] + 2 \leq \omega[2] = i + 1 \leq \omega[3] - 2$ and $\omega[3] = i' \leq \omega[4] - 2$. Most importantly, $A_{i'+1}$ is a SMALL set contained within a ball of radius at most $\rho(A_{i'})\theta \leq \rho(A_{i+1})\theta^2$ with center in $A_{i'}$, a subset of A_{i+1}. Thus

$$\begin{aligned} &\eta_{\omega[1]} + \mathfrak{F}(A_{\omega[1]}) \\ &\qquad \geq \eta_i + \mathfrak{F}(A_i) \qquad \text{by monotonicity of } \mathfrak{F} \text{ and } \eta_{\omega[1]} > \eta_i \\ &\qquad \geq C_1\rho(A_{i+1})\lambda^{i+1} + \sup\{\mathfrak{F}(A_{i+1} \cap B[\xi, \rho(A_{i+1})\theta^2] : \xi \in A_{i+1}\} \\ &\qquad \geq C_1\rho(A_{\omega[2]})\lambda^{\omega[2]} + \mathfrak{F}\left(A_{i'+1}\right). \end{aligned}$$

The second inequality comes from <16> with $f(A_{i+1}, \rho(A_{i+1})\theta^2)$ replaced by its definition. Monotonicity gives $\mathfrak{F}\left(A_{i'+1}\right) \geq \mathfrak{F}\left(A_{\omega[4]}\right)$, completing the argument □

Consequence (c):
The case where $\mathbb{L}$ is infinite is the most interesting. The argument for finite $\mathbb{L}$ is similar.

Successive appeals to inequality <19> gives

$$\begin{aligned}
\eta_{\omega[1]} + \mathfrak{F}(A_{\omega[1]}) - \mathfrak{F}(A_{\omega[4]}) &\geq C_1 \lambda^{\omega[2]} \rho(A_{\omega[2]}), \\
\eta_{\omega[4]} + \mathfrak{F}(A_{\omega[4]}) - \mathfrak{F}(A_{\omega[7]}) &\geq C_1 \lambda^{\omega[5]} \rho(A_{\omega[5]}), \\
\eta_{\omega[7]} + \mathfrak{F}(A_{\omega[7]}) - \mathfrak{F}(A_{\omega[10]}) &\geq C_1 \lambda^{\omega[8]} \rho(A_{\omega[8]}), \\
&\dots
\end{aligned}$$

and so on. If we add these inequalities the sum on the left-hand side telescopes, leaving

$$\mathfrak{F}(A_{\omega[1]}) + \left(\eta_{\omega[1]} + \eta_{\omega[4]} + \dots\right) \geq C_1 \left(\lambda^{\omega[2]} \rho(A_{\omega[2]}) + \lambda^{\omega[5]} \rho(A_{\omega[5]}) + \dots\right)$$

It follows that, for some constant C_2,

$$C_2 \mathfrak{F}(T) \geq \sum\nolimits_{\ell \in \mathbb{L}_2} \lambda^{\omega[\ell]} \rho(A_{\omega[\ell]}) \qquad \text{where } \mathbb{L}_2 := \{\ell \in \mathbb{L} : \ell \equiv 2 (\text{mod } 3)\ \}.$$

Similar arguments bound the sums over the similarly defined subsequences $\mathbb{L}_0$ and $\mathbb{L}_1$. After combining the three bounds then invoking inequality <15> to take care of $\rho(A_{\omega[1]})\lambda^{\omega[1]}$ we arrive at the inequality asserted by
□ Theorem <11>.

12.6 Beyond gaussian

Majorizing::S:next

Even though this Chapter has established lower bounds only in the gaussian case, I believe it does point the way to other applications of Talagrand's general method. As explained by Talagrand (2021, pp. 47–49), the key requirement of his approach is existence of an analog of inequality <3> for analogs of the functional $\mathfrak{F}$. I haven't read enough of his book to provide any more details. I am looking forward to further exploration.

12.7 Problems

Majorizing::S:Problems

Majorizing::P:bdd.paths

[1] (repeated from Chapter 6) Suppose $\{X_i : i \in \mathbb{N}\}$ is a gaussian process with $X_i \sim N(0, \sigma_i^2)$. Define $M_n := \max_{i \leq n} X_i$ and $M_\infty := \sup_{i \in \mathbb{N}} X_i$. Show that $M_\infty < \infty$ almost surely if and only if $\mathbb{P}M_\infty < \infty$. Argue as follows for the nontrivial implication. Suppose $\mathbb{P}\{M_\infty < \infty\} = 1$.

(i) Show that there exists an $R \in \mathbb{R}$ for which $\mathbb{P}\{M_n > R\} < 1/4$ for all large enough n.

(ii) Show that $1/4 > \mathbb{P}\{X_n > R\} = \bar{\Phi}(R/\sigma_n)$ if n is large enough. Deduce that $\sigma^2 = \sup_{i \in \mathbb{N}} \sigma_i^2$ is finite.

(iii) Write m_n for $\mathbb{P}M_n$. Use the concentration inequality from Section 6.1 to show that $\mathbb{P}\{|M_n - m_n| \ge \sigma r\} \le 2\exp(-r^2/2)$ for each n and each $r \ge 0$. Deduce that there exists an r for which $\mathbb{P}\{M_n > m_n - \sigma r\} \ge 3/4$ for each n.

(iv) From (i) and (iii) deduce that $m_n \le R + \sigma r$ for all n large enough.

(v) Show that $\mathbb{P}M_\infty = \lim_{n\to\infty} m_n \le R + \sigma r$. Hint: $0 \le M_n + |X_1| \uparrow M_\infty + |X_1|$.

(vi) Extend the argument to a two-sided equivalence: $\mathbb{P}\{\sup_i |X_i| < \infty\} = 1$ iff $\mathbb{P}\sup_i |X_i| < \infty$. Hint: $\sup_i |X_i| = \max(\sup_i X_i, \sup_i(-X_i))$.

Majorizing::P:pairwise.max

[2] Suppose (X_1, X_2) has a centered bivariate normal distribution for which $\|X_1 - X_2\|_2 = \delta > 0$. Show that $\mathbb{P}\max(X_1, X_2) = \delta/\sqrt{2\pi}$. Hint: First show that $2\max(X_1, X_2) - X_1 - X_2 = |X_1 - X_2|$. Also note that $\mathbb{P}|Z| = 2/\sqrt{2\pi}$ if $Z \sim N(0,1)$.

Majorizing::P:bdd.paths2

[3] Suppose $\{X_i : i \in \mathbb{N}\}$ is a centered gaussian process with $F := \mathbb{P}\sup_{i\in\mathbb{N}} X_i$ finite.

(i) Show that $\mathbb{P}\sup_{i,j} |X_i - X_j| = \mathbb{P}\sup_{i,j}(X_i - X_j) = 2F$. Hint: The first equality comes from the fact that both $X_i - X_j$ and $X_j - X_i$ appear in $\sup_{i,j}(X_i - X_j)$. Then note that $\sup_{i,j} |X_i - X_j|$ is an increasing limit of

$$\max_{(i,j)\in[\![n]\!]^2} |X_i - X_j| = \max_{(i,j)\in[\![n]\!]^2} (X_i - X_j) = \max_{i\in[\![n]\!]} X_i + \max_{j\in[\![n]\!]}(-X_j).$$

Use the fact that $\{(-X_j) : j \in \mathbb{N}\}$ has the same distribution as X.

(ii) Define $D := \sup_{i,j} \|X_i - X_j\|_2$. Show that $F \ge D/\sqrt{2\pi}$. Hint: Problem [2].

Majorizing::P:Sud.small

[4] Under the conditions of Theorem <9>, prove that $\delta\Psi_2^{-1}(\text{PACK}(\delta, T, d)) \to 0$ as $\delta \to 0$.

12.8 Notes

Majorizing::S:Notes

My exposition is based largely on papers by Talagrand (1992, 1996, 2001) and the three versions of his book on generic chaining: Talagrand (2005, Sections 1.3 and 2.1), Talagrand (2014, Chapter 2), and Talagrand (2021, Chapter 2). In particular, Lemma <2> comes from Talagrand (1992) via Talagrand (2021, Proposition 2.10.8).

The highly informative book of Ledoux and Talagrand (1991) provided welcome backup, particularly for the insights it gave into earlier ways of handling chaining ideas. And again I also benefitted greatly from reading the expositions by Ledoux (1996, Chap 6) and van Handel (2016). See Chapter 11 for further references.

The proof for Theorem <9> is based on Ledoux and Talagrand (1991, Thm 12.9), a version of a more involved result presented by Talagrand (1987, page 122-124). My rearrangement of their argument avoids the need for an appeal to Ledoux and Talagrand (1991, Cor 3.19).

References

Ledoux96StFlour

Ledoux, M. (1996). Isoperimetry and Gaussian analysis. *Springer Lecture Notes in Mathematics 1648*, 165–294. École d'Été de Probabilités de St-Flour XXIV , 1994.

LedouxTalagrand91book

Ledoux, M. and M. Talagrand (1991). *Probability in Banach Spaces: Isoperimetry and Processes*. New York: Springer.

Talagrand1987ActaMath

Talagrand, M. (1987). Regularity of Gaussian processes. *Acta Mathematica 159*, 99–149.

Talagrand1992GAFA

Talagrand, M. (1992). A simple proof of the majorizing measure theorem. *Geometric and Functional Analysis 2*(1), 118–125.

Talagrand96MM

Talagrand, M. (1996). Majorizing measures: The generic chaining. *Annals of Probability 24*, 1049–1103.

Talagrand2001AnnProb-MMwM

Talagrand, M. (2001). Majorizing measures without measures. *The Annals of Probability 29*(1), 411–417.

Talagrand2005MMbook

Talagrand, M. (2005). *The Generic Chaining: Upper and lower bounds of stochastic processes*. Springer-Verlag.

Talagrand2014MMbook

Talagrand, M. (2014). *Upper and Lower Bounds for Stochastic Processes: Modern methods and classical problems*. Springer-Verlag.

Talagrand2021MMbook

Talagrand, M. (2021). *Upper and Lower Bounds for Stochastic Processes: Decomposition Theorems* (Second ed.), Volume 60 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer-Verlag.

vanHandel2016Pihd

van Handel, R. (2016, December). Probability in high dimension. APC 550 Lecture Notes, Princeton University.

