Chapter 9

Chaining methods

Much of this Chapter is unedited or incomplete.

1. The chaining strategy

The previous Chapter gave several ways to bound the tails of $\max_{i\le N}|X_i|$. The chaining method applies such bounds recursively, taking advantage of extra structure on the index set.

Roughly speaking, the chaining method will work for any stochastic process $\{Z_t : t \in T\}$ for which we have some probabilistic control over the maxima of finite sets of increments $Z(s) - Z(t)$. For the basic arguments we may assume T is finite. It will be important that the bounds do not depend explicitly on the size of T if we wish to get inequalities for infinite T by passing to the limit in inequalities for finite subsets of T.

The chaining method works by linking together approximations to Z based on the values it takes on different finite subsets of T. Typically T is equipped with a metric $d(\cdot,\cdot)$ (or pseudometric) and the subsets are chosen as ***δ-nets*** for various δ.

<1> **Definition.** *A finite subset T_δ of T is a δ-net if $\min_{s\in T_\delta} d(s,t) \le \delta$ for each t in T. Equivalently, T is the union of the closed balls of radius δ with centers in T_δ. The smallest value of $\#T_\delta$ for all possible δ-nets is called the covering number, which denoted by $N(\delta)$, or $N(\delta, d, T)$ if there is any ambiguity over the choice of metric.*

REMARKS.

(i) In practice we do not need to know $N(\delta)$ exactly; an upper bound will suffice. In particular, we can often avoid messy details by choosing an upper bound that is continuous and strictly decreasing in δ. To avoid tedious qualifications, I will sometime call a subset T_δ a δ-net if $\#T_\delta$ is no larger than the upper bound on $N(\delta)$.

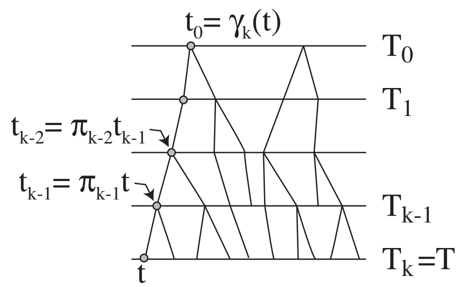
(ii) Often T itself will be a subset of a larger metric space S. As stated, the definition of a δ-net for T does not allow centers to lie in $S\backslash T$. As shown by Section 3, the restriction has only a minor effect on applications.

The typical chaining argument starts by choosing δ_i-nets T_i for numbers $\delta_0 > \delta_1 > \dots \delta_k > 0$, with $T_k = T$. We then define π_i as the map that takes each t to its closest point in T_i, with some arbitrary rule for breaking ties. That is, we construct maps $\pi_i : T \to T_i$ for which $d(t, \pi_i t) \le \delta_i$. For the basic approximation argument we do not have to know that the maps π_i have been chosen is such a way, and we do not even have to require the T_i to be δ_i-nets.

<2> **Lemma.** *Suppose $T_0, \dots, T_k = T$ are subsets of a finite set T. Suppose there exist maps $\pi_{i-1} : T_i \to T_{i-1}$. Define $\gamma_i : T_i \to T_0$ as the composition $\pi_{i-1} \circ \pi_{i-2} \circ \dots \circ \pi_0$. Then*

$$\max_{t \in T} |Z(t) - Z(\gamma_k(t))| \le M_1 + \dots + M_k$$

where $M_i := \max_{s \in T_i} |Z(s) - Z(\pi_{i-1}(s))|$.



Proof. Write D_i for $\max_{s \in T_i} |Z(s) - Z(\gamma_i(s))|$. Note that $D_1 = M_1$. For t in T write t_{k-1} for $\pi_{k-1} t$. Then

$$\begin{aligned} D_k &= \max_{t \in T_k} |Z(t) - Z(t_{k-1}) + Z(t_{k-1}) - Z(\gamma_{k-1}(t_{k-1}))| \\ &\le \max_{t \in T_k} |Z(t) - Z(t_{k-1})| + \max_{t \in T_k} |Z(t_{k-1}) - Z(\gamma_{k-1}(t_{k-1})| \\ &\le M_k + D_{k-1} \end{aligned}$$

□ Argue similarly to bound D_{k-1}, and so on.

We could use the Lemma to bound the maximum of the process Z. For each t in T,

$$Z_t \le Z(\gamma_k(t)) + |Z_t - Z(\gamma_k(t))| \le \max_{s \in T_0} Z_s + \max_{s \in T} |Z(s) - Z(\gamma_k(s))|.$$

Taking the maximum over t on the left-hand side we then get

<3> $$\max_{t \in T} Z_t \le \max_{s \in T_0} Z_s + \sum_{i=1}^{k} M_i$$

A very similar argument would show establish the analogous two-sided bound,

<4> $$\max_{t \in T} |Z_t| \le \max_{s \in T_0} |Z_s| + \sum_{i=1}^{k} M_i$$

Write N_i for $\#T_i$. Note that both the left-hand side of <3> and M_k involve a maximium over N_k variables. We can hope to get an improvement if the variables involved in M_k are "smaller than those involved in the left-hand side. It is here that control of the increments by a metric becomes important.

<5> **Example.** Let $\{Z_t : 0 \le t \le 1\}$ be a standard Brownian motion. We know that

$$\mathbb{P}\{\sup_t Z_t \ge x\} = 2\mathbb{P}\{Z_1 \ge x\} = \bar{\Phi}(x) \le \tfrac{1}{2}\exp(-x^2/2)$$

Use Orlicz norm bound for $\psi(t) = \frac{1}{2}\exp(-t^2)$ to get comparable maximal inequality. Point out that the chaining method also works in higher dimensions.

□ More details needed.

2. Chaining inequalities for norms

In Chapter ??? we found several inequalities for maxima of finitely many random variables expressible in terms of norms. For example, if pth moments are finite then

$$\| \max_{i \le N} |X_i| \,\|_p \le N^{1/p} \max_{i \le N} \|X_i\|_p$$

For an Orlicz norm defined by a convex increasing function ψ,

$$\mathbb{P} \max_{i \le N} |X_i| \le \psi^{-1}(N) \max_{i \le N} \|X_i\|_\psi$$

and

<6> $$\mathbb{P}_A \max_{i \le N} |X_i| \le \psi^{-1}(N/\mathbb{P}A) \max_{i \le N} \|X_i\|_\psi,$$

where $\mathbb{P}_A$ denotes expectation conditional on an event A with $\mathbb{P}A > 0$. If the ψ function satisfies a ***moderate growth condition***,

<7> $$\psi(\alpha)\psi(\beta) \le \psi(C_0\alpha\beta) \qquad \text{for } \psi(\alpha) \wedge \psi(\beta) \ge 1,$$

where C_0 is a finite constant, then

<8> $$\left\| \max_{i \le N} |X_i| \right\|_\psi \le C\psi^{-1}(N) \max_{i \le N} \|X_i\|_\psi \qquad \text{where } C := \frac{2 - \psi(0)}{1 - \psi(0)} C_0$$

For example, if $\psi(t) = \frac{1}{2}\exp(t^2)$ then condition <7> holds with $C_0 = 3/(\log 2)$, in which case $C = 9/(\log 2) \approx 13$.

The chaining method works well with any norm $\rho(\cdot)$ for random variables (such as an $\mathcal{L}^p$ or Orlicz norm) for which there exists a (slowly) increasing function $H(\cdot)$ such that

<9> $$\rho\left(\max_{i \le N} |Z(s_i) - Z(t_i)| \right) \le H(N) \max_{i \le N} d(s_i, t_i)$$

REMARK. We do not need ρ to be a norm. It would suffice if it were a seminorm for which $\rho(X) = 0$ implies that $X = 0$ almost surely and $\rho(X) < \infty$ implies $|X| < \infty$ almost surely. If we work with equivalence classes of random variables for which $\rho(X) < \infty$ then we get a true norm. It is traditional to abuse notation and call a seminorm a norm.

We also need to assume that $\rho(X) \le \rho(Y)$ whenever $|X| \le |Y|$. Applying Lemma <2> with the T_i as δ_i-nets and the π_i as the maps to the nearest point of T_i, we then get

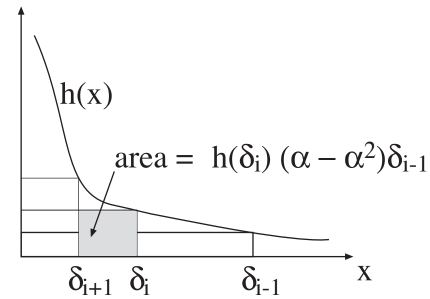
$$\rho\left(\max_{t \in T} |Z(t) - Z(\gamma_k(t))| \right) \le \rho(M_1) + \ldots + \rho(M_k)$$

<10> $$\le \sum_{i=1}^{k} H(N_i)\delta_{i-1}$$

It is traditional to bound sums by integrals to make the inequalities look cleaner.

<11> **Lemma.** *Let h be a nonnegative, decreasing function defined on an interval $(0, \delta]$. For a fixed α in $(0, 1)$, define $\delta_i := \delta\alpha^i$ for $i = 0, 1, 2, \ldots$. Then*

$$\sum_{i=1}^{k} \delta_{i-1} h(\delta_i) \le \frac{1}{\alpha - \alpha^2} \int_{\delta_{k+1}}^{\delta_1} h(x)\, dx.$$



Proof. By monotonicity of h we have $\left(\delta_i - \delta_{i+1}\right) h(\delta_i) \le \int_{\delta_{i+1}}^{\delta_i} h(x)\, dx$. Sum over i, noting that $\delta_i - \delta_{i+1} = \delta_{i-1}(\alpha - \alpha^2)$. □

When we are not worried about precise values of constants it is often convenient to choose $\alpha = 1/2$ and expand the range of integration slightly, keaving a bound $2 \int_0^\delta h(x)\, dx$. Of course, if the integral is divergent we should not be so cavalier about a contribution from $(0, \delta_{k+1})$.

REMARK. A similar trick works if we partition the vertical axis in such a way that $h(\delta_i)$ increases geometrically fast. Some of the early papers in the empirical process literature used this variation of the method to bound sums by integrals.

To summarize, let me choose $\alpha = 2$ in Lemma <11> and also disguise the evidence of the chaining construction from Lemma <2> to get a neater result.

<12> **Theorem.** *Let $\{Z_t : t \in T\}$ be a process indexed by a finite metric space T, with covering numbers $N(\cdot)$. Suppose $\rho(\cdot)$ is a norm on random variables for which $\rho(X) \le \rho(Y)$ whenever $|X| \le |Y|$. Suppose $H(\cdot)$ is an increasing function for which inequality <9> holds. Let T_δ be a δ-net for T. Then there is a map $\gamma : T \to T_\delta$ for which*

(i) $d(t, \gamma(t)) \le 2\delta$ for every t in T

(ii) $\rho\left(\max_{t\in T} |Z(t) - Z(\gamma(t))|\right) \le 4 \int_{\delta/2^{k+1}}^{\delta/2} H(N(x))\, dx$, where k is the smallest integer for which $\min\{d(s, t) : s \ne t\} \ge \delta/2^k$.

Is this the right k?

Proof. Invoke <10> for δ_i-nets T_i with $\delta_i = \delta/2^i$. Write γ instead of γ_k. Let the π_i's map to the nearest point of T_i. For a given t in T, let $t = t_k \to t_{k-1} \to \ldots \to t_1 \to t_0$ be the chain from t to $\gamma(t)$. Then

$$d(t, \gamma(t) \le d(t_k, t_{k-1}) + d(t_{k-1}, t_{k-2}) + \ldots + d(t_1, t_0) \le \delta_{k-1} + \delta_{k-2} + \ldots + \delta_0 \le 2\delta.$$

Invoke Lemma <11> to bound the sum from <10>. □

<13> **Example.** Suppose the process $\{Z(t) : t \in T\}$ satisfies the bound

$$\|Z(s) - Z(t)\|_\psi \le d(s, t) \qquad \text{for all } s, t \in T,$$

with T a finite metric space, where the convex function ψ has the moderate growth property <7>.

Let $T_0 = \{t_0\}$ and $\delta := \max_{t\in T} d(t, t_0)$.

Apply Theorem <12> with ρ as the conditional $\mathcal{L}^1$ norm for $\mathbb{P}_A$ and $H(N) = \psi^{-1}(N/\mathbb{P}A)$ to get

<14> $$\mathbb{P}_A X \le 4 \int_0^\delta \psi^{-1}\left(\frac{N(x)}{\mathbb{P}A}\right) dx \qquad \text{where } X := \sup_t |Z(t) - Z(t_0)|.$$

The $N(x)$ can be disentangled from the $\mathbb{P}A$ using the moderate growth property for ψ. Invoke <7> with $x = \psi(\alpha)$ and $y = \psi(\beta)$ to deduce that

<15> $$\psi^{-1}(xy) \le C_0 \psi^{-1}(x)\psi^{-1}(y) \qquad \text{for } x \wedge y \ge 1.$$

In particular,

$$\psi^{-1}\left(\frac{N(x)}{\mathbb{P}A}\right) \le C_0 \psi^{-1}(N(x))\psi^{-1}\left(\frac{1}{\mathbb{P}A}\right),$$

which, together with inequality <14>, implies

$$\mathbb{P}_A X \le 4C_0 \psi^{-1}\left(\frac{1}{\mathbb{P}A}\right)\int_0^\delta \psi^{-1}(N(x))\,dx.$$

Define $J := 4C_0 \int_0^\delta \psi^{-1}(N(x))\,dx$. If we choose $A = \{X \ge \epsilon\}$, for a positive ϵ, then

$$\epsilon \le \mathbb{P}_A X \le \psi^{-1}\left(\frac{1}{\mathbb{P}A}\right) J,$$

from which it follows that

$$\mathbb{P}A \le 1/\psi(\epsilon/J).$$

Compare with the tail bound we would get via a bound such as $\|X\|_\psi \le J_0$.

You might find it enlightening to consult the book of Ledoux & Talagrand (1991), who have shown that the conditional $\mathcal{L}^1$ norm is ideally suited to another, more powerful, method for deriving maximal inequalities. □

3. Covering and packing numbers

Section not yet edited.

Suppose T is a set equipped with a pseudometric d. That is, d has all the properties of a metric except that distinct points might lie at zero distance. The slight increase in generality will allow us to equip function spaces with various $\mathcal{L}^p$ norms (seminorms really) without too much fussing over almost sure equivalences.

For a subset A of T write $N_T(\delta, A, d)$ for the ***δ-covering number***, the smallest number of closed δ-balls needed to cover A. That is, the covering number is the smallest N for which there exist points $t_1, \dots, t_N$ in T with

$$\min_{i \le N} d(t, t_i) \le \delta \qquad \text{for each } t \text{ in } A.$$

The set of centers $\{t_i\}$ is called a ***δ-net*** for A. Finiteness of all covering numbers is equivalent to total boundedness of A. Covering numbers are also called ***metric entropies***.

Notice a small subtlety related to the subscript T in the definition. If we regard A as a pseudometric space in its own right, not just as a subset of T, then the covering numbers might be larger because the centers t_i would be forced to lie in A. It is an easy exercise (select a point of A from each covering ball that actually intersects A) to show that

$$N_A(2\delta, A, d) \le N_T(\delta, A, d).$$

The extra factor of 2 will be of little consequence for the bounds derived in this Chapter. When in doubt, you should interpret covering numbers to refer to N_A.

On occasion it will prove slightly more convenient to work with the ***packing number*** $D(\delta, A, d)$, defined as the largest N for which there exist points $t_1, \dots, t_N$ in A for which $d(t_i, t_j) > \delta$ if $i \neq j$. Notice the lack of a subscript T; the packing numbers are an intrinsic property of A, and do not depend on T except through the pseudometric it defines on A. The $\delta/2$-balls with centers at the t_i are disjoint; the balls are packed into A like oranges in a bag (perhaps protruding out into the larger space T).

<16> **Lemma.** *For each* $\delta > 0$,

$$N_A(\delta, A, d) \le D(\delta, A, d) \le N_T(\delta/2, A, d) \le N_A(\delta/2, A, d).$$

Proof. For the middle inequality, observe that no closed ball of radius $\delta/2$ can contain points more than δ apart. Each of the centers for $D(\delta, A, d)$ must lie in a distinct $\delta/2$ covering ball. The other inequalities have similarly simple proofs. □

I will refer to any calculation based on covering numbers or packing numbers as an ***entropy method***, to avoid unfruitful distinctions.

<17> **Example.** Let T be the real line equipped with its usual metric d, and let $A = [0, 1]$. For $\delta < 1/2$, the $N + 1$ intervals of length 2δ and centers $\delta, 3\delta, \dots, (2N-1)\delta, 1$ cover A if N is the largest integer such that $(2N-1)\delta < 1-\delta$. Thus $N_A(\delta, A, d) \le \lceil (2\delta)^{-1} \rceil$. For a lower bound, note that the Lebesgue measure of the union of covering intervals of length 2δ must be no smaller than the Lebesgue measure of A. Thus $2\delta N_T(\delta, A, d) \ge 1$. The covering numbers increase like δ^{-1} as $\delta \to 0$. Actually, only the $O(\delta^{-1})$ upper bound will matter; the lower bound merely assures us that we have found the best rate. □

<18> **Example.** Let $\|\cdot\|$ denote any norm on $\mathbb{R}^k$. For example, it might be ordinary Euclidean distance (the ℓ_2 norm), or the ℓ_1 norm, $\|x\|_1 = \sum_{i \le k} |x_i|$. The covering numbers for any such norm share a common geometric bound.

Write B_R for the ball of radius R centered at the origin. For a fixed ϵ, with $0 < \epsilon \le 1$, how many balls of radius ϵR does it take to cover B_R? Equivalently, what are the packing numbers for B_R?

Let $x_1, \dots, x_N$ be a maximal set of points in B_R with $\|x_i - x_j\| > \epsilon R$ for $i \neq j$. The closed balls of radius $\epsilon R/2$ centered at the x_i are disjoint, and their union lies within $B_{R+\epsilon R/2}$. If we write Γ for the Lebesgue measure of the unit ball B_1 then

$$N(\epsilon R/2)^k \Gamma \le (R + \epsilon R/2)^k \Gamma,$$

from which we deduce $N \le ((2+\epsilon)/\epsilon)^k \le (3/\epsilon)^k$, for $0 < \epsilon \le 1$. □

4. Infinite index sets

Suppose the norm ρ from Theorem <12> also has the property

<19> $$\text{if } 0 \le X_1 \le X_2 \le \dots \uparrow X \text{ then } \rho(X_n) \uparrow \rho(X).$$

Then we can pass to the limit in the inequality asserted by that Theorem to get bounds involving points from a countable dense subset of T. There are a few small subtleties in the construction, which I will illustrate by establishing a very useful equicontinuity bound.

<20> **Theorem.** *Let $\{Z_t : t \in T\}$ be a process indexed by a metric space T, with covering numbers $N(\cdot)$. Suppose $\rho(\cdot)$ is a norm on random variables for which $\rho(X) \le \rho(Y)$ whenever $|X| \le |Y|$ and for which properties <9> and <19> hold. Suppose $H(\cdot)$ is an increasing function for which inequality <9> holds and for which*

<21> $$\int_0^1 H(N(x))\,dx < \infty$$

Then:

(i) There exists a countable dense subset T_∞ of T for which: to each $\epsilon > 0$ there exists an $\eta > 0$ such that

$$\rho\left(\sup |Z_s - Z_t| : s, t \in T_\infty \text{ and } d(s,t) < \eta\right) \le \epsilon$$

(ii) Almost all sample paths of $\{Z_t : t \in T_\infty\}$ are uniformly continuous.

(iii) There exists a process $\{\widetilde{Z}_t : t \in T\}$ with uniformly continuous sample paths such that $\mathbb{P}\{\widetilde{Z}_t = Z_t\} = 1$ for each t in T and for which

$$\rho\left(\sup |\widetilde{Z}_s - \widetilde{Z}_t| : s, t \in T \text{ and } d(s,t) < \eta\right) \le \epsilon$$

Proof. It will be easier to work with packing numbers $D(\cdot)$ rather than covering numbers. The finiteness condition <21> still holds if we replace $N(x)$ by $D(x)$ because (Lemma <16>) $D(x) \le N(x/2)$. Choose a $\delta > 0$ for which

$$\int_0^\delta H(D(x))\,dx \le \epsilon$$

Define $\delta_i := \delta/2^i$ for $i = 0, 1, 2, \ldots$. Construct sets $T_0 \subseteq T_1 \subseteq T_2 \subseteq \ldots$ by choosing T_0 as a maximal set of points for which $d(s,t) > \delta_0$ if $s \ne t$, for $s, t \in T_0$. Then add extra points to T_0 to create a maximal set of points $T_1 \supseteq T_0$ for which $d(s,t) > \delta_1$ if $s \ne t$, for $s, t \in T_1$. And so on. Thus $\#T_k \le D(\delta_k)$ for each k and T_k is a δ_k-net for T. Moreover,

$$T_k \uparrow T_\infty := \cup_i T_i \qquad \text{as } k \uparrow \infty.$$

Construct chains and maps $\gamma_k : T_k \to T_0$ as in Section 1.

Temporarily hold k fixed. Invoke Theorem <12> to show that

$$\rho\left(G_k\right) \le 4\int_0^\delta H(N(x))\,dx \le 4\epsilon \qquad \text{where } G_k := \max_{t \in T_k} |Z(t) - Z(\gamma_k(t))|.$$

Now we come to a subtle part of the argument, making use of a clever construction from Ledoux & Talagrand (1991, Section 11.1).

The map γ_k partitions T_k into $N \le \#T_0 \le D(\delta)$ equivalence classes $E_1, \ldots, E_N$, by means of the relation $s \sim t$ if $\gamma_k s = \gamma_k t$. If $s \sim t$ then

<22> $$|Z(s) - Z(t)| \le |Z(s) - Z(\gamma_k s)| + |Z(\gamma_k t) - Z(t)| \le 2G_k.$$

For an as yet unspecified $\eta > 0$, write $E_i \approx E_j$ if there exist points $t_{ij} \in E_i$ and $t_{ji} \in E_j$ such that $d(t_{ij}, t_{ji}) < \eta$. Define

$$G := \max_{E_i \approx E_j} |Z(t_{ij}) - Z(t_{ji})|.$$

The maximum runs over at most N^2 pairs (t_{ij}, t_{ji}). By inequality <9>

$$\rho(G) \le H(N^2)\eta \le H(D(\delta)^2)\eta,$$

which is less than ϵ if η is chosen small enough.

If $S \subseteq T$ define

$$M(S,\eta) := \sup\{|Z_s - Z_t| : d(s,t) < \eta \text{ and } s, t \in S\}.$$

Of course, if S is finite then the sup could be replaced by a max.

A multi-step approximation will let us reduce comparison of pairs in T_k to comparison between pairs in T_0. Suppose s and t are points of T_k such that $d(s,t) < \eta$. If $s \in E_i$ and $t \in E_j$ then $E_i \approx E_j$, and $s \sim t_{ij}$ and $t \sim t_{ji}$. It follows that

Any difficulty if $i = j$?

$$\begin{aligned} |Z(s) - Z(t)| &\le |Z(s) - Z(t_{ij})| + |Z(t_{ij}) - Z(t_{ji})| + |Z(t_{ji}) - Z(t)| \\ &\le 2G_k + G + 2G_k. \end{aligned}$$

Take the maximum over all such (s,t) pairs then take norms of both sides.

<23> $$\rho\left(M(T_k,\eta)\right) \le 4\rho(G_k) + \rho(G) \le 5\epsilon.$$

You should repeat the argument with ϵ replaced by $\epsilon/5$ if you want the final inequality to exactly as stated.

Now let k tend to infinity. Each pair (s,t) that contributes to the supremum in $M(T_\infty,\eta)$ must appear in some T_k. It follows that

$$M(T_k,\eta) \uparrow M(T_\infty,\eta)$$

Invoke property <19> of the norm ρ to deduce assertion (i) of the Theorem.

To show that almost all sample paths of $\{Z_t : t \in T_\infty\}$ are uniformly continuous, invoke (i) to find a sequence $\{\eta_m\}$ for which $M(T_\infty,\eta_m) \le 2^{-m}$. Then, by the continuity property <19>,

$$\rho\left(\sum_{m\in\mathbb{N}} M(T_\infty,\eta_k)\right) \le \sum_{m\in\mathbb{N}} \rho\left(M(T_\infty,\eta_k)\right) < \infty.$$

The sum $\sum_{m\in\mathbb{N}} M(T_\infty,\eta_k)$ is finitely almost surely and, consequently there exists a negligible set $\mathcal{N}$ such that

$$M(T_\infty,\eta_k) \to 0 \qquad \text{for } \omega \in \mathcal{N}^c.$$

The sample paths for $\omega \in \mathcal{N}^c$ are uniformly continuous (as a function on T_∞). For those ω, the path extends to a unique uniformly continuous function $\widetilde{Z}_t(\omega)$ on T. Define $\widetilde{Z}_t(\omega) \equiv 0$ for $\omega \in \mathcal{N}$.

Finish the argument

For pairs s,t in T for which $d(s,t) < \eta$, find sequences $\{t_k\}$ and $\{s_k\}$ in T_∞ for which $d(s_k,s) \to 0$ and $d(t_k,t) \to 0$. Then what?

□

5. Chaining with random distances

Section not yet edited. Please ignore.

<24> **Theorem.** *Let $\{\Delta(s,t) : (s,t) \in T \times T\}$ be a random distance with $\Delta(s,t) \in \mathcal{M}^+_\rho$ for all (s,t). Suppose there exists an an increasing function H for which*

$$\rho\left(\max\nolimits_{i=1}^N \Delta(s_i,t_i)\right) \le H(N)\max\nolimits_{i=1}^N d(s_i,t_i)$$

for all finite sets of pairs $(s_1,t_1),\dots,(s_N,t_N)$. Then for each δ-net T_δ and each finite subset S of T ...

$$\rho\Big(\max_{t\in S} \Delta(t,\gamma_t)\Big) \le 4\int_0^\delta H(N(x))\,dx.$$

Proof.

$$\rho\big(\max_{t\in S}\Delta(t,\gamma_t)\big) \le \rho\big(\max_{t\in S}\big(\Delta(t,t_k)\big) + \sum\nolimits_{j=1}^{k}\rho\big(M_j\big)$$

$$\le H(\#S)\left(2\sum_{j\ge k}\delta_j\right) + \sum\nolimits_{j=1}^{k} H(N(\delta_j))\delta_{j-1}$$

□

<25> **Example.** Suppose $\{Z(t) : t\in[0,1]\}$ is a process for which there exists a finite measure μ on $[0,1]$ and constants $\gamma>0$ and $\alpha>1$ for which

$$\left(\mathbb{P}|Z(s)-Z(t)|^{\gamma}\right)^{1/\alpha} \le \mu(s,t],$$

for all $0\le s<t\le 1$. Show that Z has a version with cts paths. Argue first with $\gamma\ge\alpha$, then the other case, for $0<\gamma<\alpha$, as in L&T. [Comment on the usefulness of the two-parameter process in the chaining argument. □

cf. Ledoux & Talagrand (1991, page 308)

6. Maximal inequalities for tail probabilities

Section not yet edited. Please ignore.

Let $\{\Delta(s,t) : (s,t)\in T\times T\}$ be a random distance indexed by a pseudometric space (T,d) for which we have a bound $N(\cdot)$ on the covering numbers. Suppose the tail probabilities for $\Delta(s,s')$ are controlled by the pseudometric by means of a nonnegative function $\beta(\cdot,\cdot)$, which is decreasing in its first argument and increasing in its second argument, such that

<26> $$\mathbb{P}\{\Delta(s,s')\ge\eta\}\le\beta(\eta,d(s,s')) \qquad \text{for } s,s'\in T \text{ and } \eta\ge 0.$$

For N pairs (s_i,s_i') each with $d(s_i,s_i')\le\delta$ we then have a bound,

<27> $$\mathbb{P}\{\max\nolimits_{i\le N}\Delta(s_i,s_i')\ge\eta\}\le N\beta(\eta,\delta).$$

For nonnegative numbers $\eta,\eta_1,\eta_2,\ldots$, Lemma <GENERAL.CHAIN> gives, for each finite subset S of T, a maximal inequality:

$$\mathbb{P}\{\max\nolimits_{t\in S}\Delta(t,\gamma_t)\ge\eta+\eta_1+\ldots+\eta_k\}$$

$$\le\mathbb{P}\{\max\nolimits_{t\in S}\Delta(t,t_k)\ge\eta\}+\sum\nolimits_{i=1}^{k}\mathbb{P}\{M_i\ge\eta_i\}$$

<28> $$\le\big(\#S\big)\ \beta(\eta,\delta_k+\sum\nolimits_{i\ge k}\delta_i)+\sum\nolimits_{i=1}^{k}N(\delta_i)\beta(\eta_i,\delta_{i-1}).$$

Provided $\beta(\eta,\delta)\to 0$ as $\delta\to 0$ for each fixed η, we also have a limiting form of the maximal inequality:

<29> $$\mathbb{P}\{\max\nolimits_{t\in S}\Delta(t,\gamma_t)>\sum\nolimits_{i=1}^{\infty}\eta_i\}\le\sum\nolimits_{i=1}^{\infty}N(\delta_i)\beta(\eta_i,\delta_{i-1}).$$

Notice the strict inequality on the left-hand side, to accommodate a small positive η.

If the covering bound $N(x)$ increases slowly enough as x tends to zero, and if $\beta(\eta,\delta)$ tends to zero rapidly enough when $\eta\to\infty$ and $\delta\to 0$ at appropriate rates, the maximal inequalities can be expressed in slightly more explicit forms. It is traditional to bound sums by integrals to make the inequalities look even simpler.

<30> **Example.** Let $\{Z_t : t \in T\}$ be a stochastic process whose increments satisfy a subgaussian inequality controlled by the pseudometric on T:

<31> $$\mathbb{P}\{|Z_s - Z_t| \ge \eta d(s,t)\} \le c_0 \exp\left(-c_1^2\eta^2\right),$$

for some positive constants c_0 and c_1. That is, the tail bound <26> holds with $\beta(\eta,\delta) = c_0\exp(-c_1^2\eta^2/\delta^2)$. Once again write $\Delta(s,t)$ for $|Z_s - Z_t|$. Inequality <29> becomes

$$\mathbb{P}\{\max_{t\in S}\Delta(t,\gamma_t) > \sum\nolimits_{i=1}^{\infty}\eta_i\} \le c_0\sum\nolimits_{i=1}^{\infty} N(\delta_i)\exp(-c_1^2\eta_i^2/\delta_{i-1}^2).$$

We need to choose the $\{\eta_i\}$ to make the sum on the right-hand side converge. A geoemtric rate of decrease woulld ensure that the sum behaves like its first term. For a fixed, positive x define η_i so that

$$\exp(-c_1^2\eta_i^2/\delta_{i-1}^2) = e^{-x}2^{-i}/N(\delta_i),$$

that is,

$$\begin{aligned}\eta_i &:= c_1^{-1}\delta_{i-1}\sqrt{\log N(\delta_i) + i\log 2 + x}\\ &\le c_1^{-1}2\delta_i\left(h(\delta_i) + \sqrt{i\log 2} + \sqrt{x}\right) \qquad \text{where } h(y) := \sqrt{\log N(y)}.\end{aligned}$$

With the help of Lemma <11> we then get

$$\sum\nolimits_{i=1}^{\infty}\eta_i \le c_2\int_0^{\delta/2} h(y)\,dy + c_3\delta + c_4\delta\sqrt{x},$$

where $c_2 := 4c_1^{-1}$ and $c_3 := 2c_1^{-1}\sqrt{\log 2}\sum_{i=1}^{\infty}\sqrt{i}/2^i$ and $c_4 := 2c_1^{-1}$. Assume that the covering bounds increase slowly enough that the integral

$$J_z := \int_0^z \sqrt{\log N(y)}\,dy$$

is convergent for each $z > 0$. Then, for each finite subset S of T,

<32> $$\mathbb{P}\{\max_{t\in S}|Z_t - Z_{\gamma_t}| > c_2J_\delta + c_3\delta + c_4\delta\sqrt{x}\} \le c_0e^{-x},$$

where the constants c_3, c_4, and c_5 depend only on the c_1 from <31>.

Now suppose that T has radius at most R, in the sense that there is some point τ in T for which $\sup_{t\in T} d(t,\tau) = R < \infty$, and that we wish to determine how large a value w is needed to make the tail probability $\mathbb{P}\{\max_{t\in S}|Z_t - Z_\tau| > w\}$ smaller than a prescribed quantity, which for convenience I write as $2c_0e^{-x}$.

As t ranges over S, the value γ_t ranges over a subset of the δ-net T_δ, a set with at most $N(\delta)$ points each at a distance at most R from τ. The inequality

$$\max_{t\in S}|Z_t - Z_\tau| \le \max_{t\in T_\delta}|Z_t - Z_\tau| + \max_{t\in S}|Z_t - Z_{\gamma_t}|$$

then leads us to a bound

$$\mathbb{P}\{\max_{t\in S}|Z_t - Z_\tau| > w + c_2J_\delta + c_3\delta + c_4\delta\sqrt{x}\} \le N(\delta)c_0\exp(-c_1^2w^2/R^2) + c_0e^{-x}.$$

Remark. Notice that I have built in the assumption that a reasonable way to make a sum of two terms small is to put each of them equal to half the desired sum. Perhaps a significantly better bound could be obtained by discarding the assumption and trying to optimize over the alocation of how much of the final tail bound comes from each term.

For a given x we are left with the task of choosing δ and w to make

$$w + c_2J_\delta + c_3\delta + c_4\delta\sqrt{x} \quad \text{small subject to } N(\delta)\exp(-c_1^2w^2/R^2) \le e^{-x}.$$

Of course there is no point in making δ larger than R, because we may assume $N(y) = 1$ for $y > R$. Also, we may suppose x is bounded away from zero (say $x > c_5$), because there is no point in trying to optimize when $c_0 e^{-x}$ is not a lot smaller than 1.

The smallest w satisfying the constraint is $(R/c_1)\sqrt{x + h(\delta)^2}$, where once again $h(y) = \sqrt{\log N(y)}$. We have the formidable task of finding $\delta \in (0, R]$ to minimize

$$\frac{R}{c_1}\sqrt{x + h(\delta)^2} + c_2 \int_0^\delta h(y)\,dy + \delta\bigl(c_3 + c_4\sqrt{x}\bigr).$$

□

7. Chaining via majorizing measures

Talagrand (1996)

8. Problems

More problems to come

[1] Show that the convex function $\psi(x) = (1 + x)\log(1 + x) - x$ does not satisfy the growth condition <7>. Hint: Consider the limit of $\psi(C_0 x^2)/\psi(x)^2$ as $x \to \infty$.

9. Notes (inaccurate and incomplete)

Acknowledge Ledoux & Talagrand (1991) for several of the ideas used in this Chapter: Example <CONDIT.MEAN>; the introduction of the two-parameter process in the proof of Theorem <FINITE.MAXIMAL> (and its usefulness in the analog of Example <25> for $0 < \gamma < \alpha$); Example <ORLICZ2>, or maybe cite Pisier; the subtle equivalence class idea in Example <ORLICZ2>; and the method used in Example <13>.

Give some history of earlier work: Dudley, Pisier?

van der Vaart & Wellner (1996)

References

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

Talagrand, M. (1996), ‘Majorizing measures: The generic chaining (in special invited paper)’, *Annals of Probability* **24**, 1049–1103.

van der Vaart, A. W. & Wellner, J. A. (1996), *Weak Convergence and Empirical Process: With Applications to Statistics*, Springer-Verlag.