SECTION 3

Chaining

The main aim of the section is to derive a maximal inequality for the processes $\boldsymbol{\sigma} \cdot \mathbf{f}$, indexed by subsets of $\mathbb{R}^n$, in the form of an upper bound on the Ψ norm of $\sup_{\mathcal{F}} |\boldsymbol{\sigma} \cdot \mathbf{f}|$. [Remember that $\Psi(x) = {}^1\!/\!_5 \exp(x^2)$.] First we need a bound for the individual variables.

(3.1) LEMMA. *For each $\mathbf{f}$ in $\mathbb{R}^n$, the random variable $\boldsymbol{\sigma} \cdot \mathbf{f}$ has subgaussian tails, with Orlicz norm $\|\boldsymbol{\sigma} \cdot \mathbf{f}\|_\Psi$ less than $2|\mathbf{f}|$.*

PROOF. The argument has similarities to the randomization argument used in Section 2. Assume the probability space is a product space supporting independent $N(0,1)$ distributed random variables $g_1, \ldots, g_n$, all of which are independent of the sign variables $\sigma_1, \ldots, \sigma_n$. The absolute value of each g_i has expected value

$$\gamma = \mathbb{P}|N(0,1)| = \sqrt{2/\pi}.$$

By Jensen's inequality,

$$\begin{aligned}\mathbb{P}_\sigma \exp\Big(\sum_{i \le n} \sigma_i f_i / C\Big)^2 &= \mathbb{P}_\sigma \exp\Big(\sum_{i \le n} \sigma_i f_i \mathbb{P}_g |g_i| / \gamma C\Big)^2 \\ &\le \mathbb{P}_\sigma \mathbb{P}_g \exp\Big(\sum_{i \le n} \sigma_i |g_i| f_i / \gamma C\Big)^2.\end{aligned}$$

The absolute value of any symmetric random variable is independent of its sign. In particular, under $\mathbb{P}_\sigma \otimes \mathbb{P}_g$ the products $\sigma_1|g_1|, \ldots, \sigma_n|g_n|$ are independent $N(0,1)$ random variables. The last expected value has the form $\mathbb{P}\exp\big(N(0,\tau^2)^2\big)$, where the variance is given by

$$\tau^2 = \sum_{i \le n} (f_i/\gamma C)^2 = |\mathbf{f}|^2/\gamma^2 C^2.$$

Provided $\tau^2 < {}^1\!/\!_2$, the expected value is finite and equals $(1 - 2|\mathbf{f}|^2/\gamma^2 C^2)^{-1}$. If we choose $C = 2|\mathbf{f}|$ this gives $\mathbb{P}\,\Psi(\boldsymbol{\sigma} \cdot \mathbf{f}/C) \le 1$, as required. □

The next step towards the maximal inequality is to bound the Ψ norm of the maximum for a finite number of random variables.

(3.2) LEMMA. *For any random variables $Z_1, \ldots, Z_m$,*

$$\big\|\max_{i\le m}|Z_i|\big\|_\Psi \le \sqrt{2+\log m}\,\max_{i\le m}\|Z_i\|_\Psi.$$

PROOF. The inequality is trivially satisfied if the right-hand side is infinite. So let us assume that each Z_i belongs to $\mathcal{L}^\Psi$. For all positive constants K and C,

$$\begin{aligned}\Psi(\max|Z_i|/C) &\le \Psi(1) + \int_1^\infty \{K\max|Z_i|/C > Kx\}\Psi(dx)\\ &\le \Psi(1) + \int_1^\infty \frac{\Psi(K\max|Z_i|/C)}{\Psi(Kx)}\Psi(dx)\\ &\le \Psi(1) + \sum_{i\le m}\int_1^\infty \frac{\Psi(KZ_i/C)}{\Psi(Kx)}\Psi(dx).\end{aligned}$$

If we choose $C = K\max\|Z_i\|_\Psi$ then take expectations we get

$$\begin{aligned}\mathbb{P}\,\Psi(\max|Z_i|/C) &\le \Psi(1) + m\int_1^\infty \frac{1}{\Psi(Kx)}\Psi(dx)\\ &= \frac{e}{5} + m(K^2-1)^{-1}\exp(-K^2+1).\end{aligned}$$

The right-hand side will be less than 1 if $K = \sqrt{2+\log m}$. (Now you should be able to figure out why the $1/5$ appears in the definition of Ψ.) □

Clearly the last lemma cannot be applied directly to bound $\big\|\sup_{\mathcal{F}}|\boldsymbol{\sigma}\cdot\mathbf{f}|\big\|_\Psi$ if $\mathcal{F}$ is infinite. Instead it can be used to tie together a sequence of approximations to $\sup_{\mathcal{F}}|\boldsymbol{\sigma}\cdot\mathbf{f}|$ based on an increasing sequence of finite subsets $\mathcal{F}$. The argument, which is usually referred to as *chaining*, depends on the geometry of $\mathcal{F}$ only through the size of its *packing numbers.* To begin with, let us consider a more general—more natural—setting: a stochastic process $\{Z(t) : t \in T\}$ whose index set T is equipped with a metric d. [Actually, d need only be a pseudometric; the argument would not be affected if some distinct pairs of points were a distance zero apart.]

(3.3) DEFINITION. The packing number $D(\epsilon, T_0)$ for a subset T_0 of a metric space is defined as the largest m for which there exist points $t_1, \ldots, t_m$ in T_0 with $d(t_i, t_j) > \epsilon$ for $i \ne j$. The covering number $N(\epsilon, T_0)$ is defined as the smallest number of closed balls with radius ϵ whose union covers T_0.

The two concepts are closely related, because

$$N(\epsilon, T_0) \le D(\epsilon, T_0) \le N(\epsilon/2, T_0).$$

Both provide approximating points $t_1, \ldots, t_m$ for which $\min_i d(t, t_i) \le \epsilon$ for every t in T_0. Sometimes the $\{t_i\}$ provided by D are slightly more convenient to work with, because they lie in T_0; the centers of the balls provided by N need not lie in T_0. The definition of D depends only upon the behavior of the metric d on the set T_0; the value of N can depend upon the particular T into which T_0 is embedded. If $T = T_0$ the ambiguity disappears. However, it is largely a matter of taste, or habit, whether one works with covering numbers or packing numbers. Notice that finiteness of all the packing or covering numbers is equivalent to total boundedness of T_0.

For the general maximal inequality let us suppose that some point t_0 has been singled out from T. Also, let us assume that the process $Z(t) = Z(\omega, t)$ has continuous sample paths, in the sense that $Z(\omega, \cdot)$ defines a continuous function on T for each ω. For the intended application, this causes no loss of generality: clearly $\boldsymbol{\sigma} \cdot \mathbf{f}$ is a continuous function of $\mathbf{f}$ for each fixed $\boldsymbol{\sigma}$. [Without the continuity assumption the statement of the next lemma would have to be modified to assert existence of a version of the process Z having continuous sample paths and satisfying the stated inequality.]

(3.4) LEMMA. *If the process Z has continuous sample paths and its increments satisfy the inequality*

$$\big\|Z(s) - Z(t)\big\|_{\Psi} \leq d(s,t) \qquad \textit{for all } s,t \textit{ in } T,$$

and if $\delta = \sup_t d(t, t_0)$, then

$$\big\|\sup_T |Z(t)|\big\|_{\Psi} \leq \big\|Z(t_0)\big\|_{\Psi} + \sum_{i=0}^{\infty} \frac{\delta}{2^i}\sqrt{2 + \log D(\delta/2^{i+1}, T)}.$$

PROOF. The inequality holds trivially if the right-hand side is infinite. So let us assume that δ and all the packing numbers are finite.

Abbreviate $\delta/2^i$ to δ_i. Construct a succession of approximations to the supremum based on a sequence of finite subsets $\{t_0\} = T_0 \subseteq T_1 \subseteq \cdots$ with the property that

$$\min_{t^* \in T_i} d(t, t^*) \leq \delta_i \qquad \text{for every } t \text{ in } T.$$

Such sets can be obtained inductively by choosing T_i as a maximal superset of T_{i-1} with all points of T_i greater than δ_i apart. [Notice that the definition of δ ensures that $\{t_0\}$ has the desired property for δ_0.] The definition of packing number gives us a bound on the cardinality of T_i, namely, $\#T_i \leq D(\delta_i, T)$. Let us write m_i for this bound.

Fix, for the moment, a non-negative integer k. Relate the maximum of $|Z(t)|$ over T_{k+1} to the maximum over T_k. For each t in T_{k+1} let t^* denote a point in T_k such that $d(t, t^*) \leq \delta_k$. By the triangle inequality,

$$\max_{t \in T_{k+1}} |Z(t)| \leq \max_{t \in T_{k+1}} |Z(t^*)| + \max_{t \in T_{k+1}} |Z(t) - Z(t^*)|.$$

On the right-hand side, the first term takes a maximum over a subset of T_k. The second term takes a maximum over at most m_{k+1} increments, each of which has Ψ norm at most δ_k. Taking Ψ norms of both sides of the inequality, then applying Lemma 3.2 to the contribution from the increments, we get

$$\big\|\max_{T_{k+1}} |Z(t)|\big\|_{\Psi} \leq \big\|\max_{T_k} |Z(t)|\big\|_{\Psi} + \delta_k\sqrt{2 + \log m_{k+1}}.$$

Repeated application of this recursive bound increases the right-hand side to the contribution from T_0, which reduces to $\|Z(t_0)\|_\Psi$, plus a sum of terms contributed by the increments.

As k tends to infinity, the set T_{k+1} expands up to a countable dense subset T_∞ of T. A monotone convergence argument shows that

$$\big\|\max_{T_{k+1}} |Z(t)|\big\|_\Psi \nearrow \big\|\sup_{T_\infty} |Z(t)|\big\|_\Psi.$$

Continuity of the sample paths of Z lets us replace T_∞ by T, since

$$\sup_{T_\infty} |Z(\omega,t)| = \sup_{T} |Z(\omega,t)| \qquad \text{for every } \omega.$$

This gives the asserted inequality. □

Now we have only to specialize the argument to the process $\boldsymbol{\sigma}\cdot\mathbf{f}$ indexed by a subset $\mathcal{F}$ of $\mathbb{R}^n$. The packing numbers for $\mathcal{F}$ should be calculated using the usual Euclidean distance on $\mathbb{R}^n$. By Lemma 3.1 the increments of the process satisfy

$$\big\|\boldsymbol{\sigma}\cdot(\mathbf{f}-\mathbf{g})\big\|_\Psi \le 2|\mathbf{f}-\mathbf{g}|,$$

which differs from the inequality required by Lemma 3.4 only through the presence of the factor 2. We could eliminate the factor by working with the process $1/2\,\boldsymbol{\sigma}\cdot\mathbf{f}$.

To get a neater bound, let us take the origin of $\mathbb{R}^n$ as the point corresponding to t_0. At worst, this increases the packing numbers for $\mathcal{F}$ by one. We can tidy up the integrand by noting that $D(x,\mathcal{F}) \ge 2$ for $x < \delta$, and then using the inequality

$$\sqrt{2+\log(1+D)}/\sqrt{\log D} < 2.2 \qquad \text{for } D \ge 2.$$

It has also become traditional to replace the infinite series in Lemma 3.4 by the corresponding integral, a simplification made possible by the geometric rate of decrease in the $\{\delta_i\}$:

$$\delta_i\sqrt{\log D(\delta_{i+1},\mathcal{F})} \le 4\int \{\delta_{i+2} < x \le \delta_{i+1}\}\sqrt{\log D(x,\mathcal{F})}\,dx.$$

With these cosmetic changes the final maximal inequality has a nice form.

(3.5) THEOREM. *For every subset $\mathcal{F}$ of $\mathbb{R}^n$,*

$$\big\|\sup_{\mathcal{F}} |\boldsymbol{\sigma}\cdot\mathbf{f}|\big\|_\Psi \le 9\int_0^\delta \sqrt{\log D(x,\mathcal{F})}\,dx \qquad \textit{where } \delta = \sup_{\mathcal{F}} |\mathbf{f}|. \qquad □$$

The theorem has several interesting consequences and reformulations. For example, suppose the integral on the right-hand side is finite. Then there exist positive constants κ_1 and κ_2 such that

$$\mathbb{P}_\sigma\{\sup_{\mathcal{F}} |\boldsymbol{\sigma}\cdot\mathbf{f}| \ge \epsilon\} \le \kappa_1\exp(-\kappa_2\epsilon^2) \qquad \text{for all } \epsilon > 0.$$

It will also give bounds for less stringent norms than the Ψ norm. For example, for each p with $\infty > p \ge 1$ there exists a constant C_p such that $|x|^p \le \Psi(C_p x)$ for all x.

This implies that $\|Z\|_p \le C_p \|Z\|_\Psi$ for every random variable Z, and, in particular,

$$\left\|\sup_{\mathcal{F}} |\boldsymbol{\sigma} \cdot \mathbf{f}|\right\|_p \le 9C_p \int_0^\delta \sqrt{\log D(x, \mathcal{F})}\, dx, \qquad \text{where } \delta = \sup_{\mathcal{F}} |\mathbf{f}|. \tag{3.6}$$

Such bounds will prove convenient in later sections.

REMARKS. The literature contains many different maximal inequalities derived by chaining arguments. The method presented in this section could be refined to produce more general inequalities, but Theorem 3.5 will suffice for the limited purposes of these notes.

I learnt the method for Lemma 3.1 from Gilles Pisier. The whole section is based on ideas exposited by Pisier (1983), who proved an inequality equivalent to

$$\mathbb{P} \sup_{s,t} |Z(s) - Z(t)| \le K \int_0^\delta \Phi^{-1}(D(x, T))\, dx$$

for general convex, increasing Φ with $\Phi(0) = 0$. This result is weaker than the corresponding inequality with the left-hand side increased to

$$\left\|\sup_{s,t} |Z(s) - Z(t)|\right\|_\Phi .$$

For the special case where $\Phi(x) = {}^1\!/_5 \exp(x^2)$ the improvement is made possible by the substitution of my Lemma 3.2 for Pisier's Lemma 1.6 in the chaining argument. Ledoux and Talagrand (1990, Chapter 11) have shown how the stronger form of the inequality can also be deduced directly from a slight modification of Pisier's inequality.

Both Gaenssler and Schlumprecht (1988) and Pollard (1989) have established analogues of Theorem 3.5 for $\|\cdot\|_p$ norms instead of the $\|\cdot\|_\Psi$.