SECTION 7

Maximal Inequalities

Let us now pull together the ideas from previous sections to establish a few useful maximal inequalities for the partial-sum process S_n. To begin with, let us consider an infinite sequence of independent processes $\{f_i(\omega, t)\}$, in order to see how the bounds depend on n. This will lead us to the useful concept of a manageable triangular array of processes.

The symmetrization bound from Section 2 was stated in terms of a general convex, increasing function Φ on $\mathbb{R}^+$. The chaining inequality of Section 3 was in terms of the specific convex function given by $\Psi(x) = {}^1\!/_5 \exp(x^2)$.

Section 2 related the maximum deviation of S_n from its expected value,

$$\Delta_n(\omega) = \sup_t |S_n(\omega, t) - M_n(t)|,$$

to the process $\boldsymbol{\sigma} \cdot \mathbf{f}$ indexed by the random set

$$\mathcal{F}_{n\omega} = \{\big(f_1(\omega, t), \dots, f_n(\omega, t)\big) : t \in T\}.$$

If we abbreviate the supremum of $|\boldsymbol{\sigma} \cdot \mathbf{f}|$ over $\mathcal{F}_{n\omega}$ to $L_n(\boldsymbol{\sigma}, \omega)$, the inequality becomes

$$(7.1) \qquad \mathbb{P}\,\Phi(\Delta_n) \le \mathbb{P}\,\Phi(2L_n).$$

We bound the right-hand side by taking iterated expectations, initially conditioning on ω and averaging over $\boldsymbol{\sigma}$ with respect to the uniform distrbution $\mathbb{P}_\sigma$.

The chaining inequality from Theorem 3.5 bounds the conditional Ψ norm of L by

$$J_n(\omega) = 9 \int_0^{\delta_n(\omega)} \sqrt{\log D(x, \mathcal{F}_{n\omega})}\, dx, \qquad \text{where } \delta_n(\omega) = \sup_{\mathcal{F}_{n\omega}} |\mathbf{f}|.$$

Here, and throughout the section, the subscript 2 is omitted from the ℓ_2 norm $|\cdot|_2$; we will make no use of the ℓ_1 norm in this section. Written out more explicitly, the inequality that defines the Ψ norm becomes

$$(7.2) \qquad \mathbb{P}_\sigma \exp\big(L_n(\boldsymbol{\sigma}, \omega)/J_n(\omega)\big)^2 \le 5.$$

Because J_n is a random variable, in general we cannot appeal directly to inequality (7.1) with $\Phi(x) = \exp(x^2/2J_n^2)$, to get some sort of bound for the Ψ norm of the partial-sum process. We can, however, combine the two inequalities to get several most useful bounds.

The simplest situation occurs when $J_n(\omega)$ is bounded by a constant K_n. As we shall see soon, this often happens when the envelopes F_i are uniformly bounded. Increasing J_n to K_n in (7.2), then taking expectations we get, via (7.1),

$$\mathbb{P}\exp(\tfrac{1}{2}\Delta_n^2/K_n^2) \le 5.$$

It follows that Δ_n has subgaussian tails:

$$\mathbb{P}\{\Delta_n \ge t\} \le 5\exp(-\tfrac{1}{2}t^2/K_n^2) \qquad \text{for all } t > 0. \tag{7.3}$$

This is not the best subgaussian upper bound; the constant K_n could be replaced by a smaller constant.

If $J_n(\omega)$ is not uniformly bounded, but instead has a finite Ψ norm, we still get an exponential bound on the tail probabilities for Δ_n, by means of the inequality

$$2L_n/C \le J_n^2/C^2 + L_n^2/J_n^2 \qquad \text{for constant } C.$$

With $C = \|J_n\|_\Psi$ this inequality implies

$$\begin{aligned}\mathbb{P}\exp(\Delta_n/C) &\le \mathbb{P}\exp(2L_n/C)\\ &\le \mathbb{P}_\omega\left[\exp(J_n^2/C^2)\mathbb{P}_\sigma\exp(L_n^2/J_n^2)\right]\\ &\le 25.\end{aligned}$$

Consequently,

$$\mathbb{P}\{\Delta_n \ge t\} \le 25\exp(-t/\|J_n\|_\Psi) \qquad \text{for all } t > 0. \tag{7.4}$$

We have traded a strong moment condition on J_n for a rapid rate of decrease of the Δ_n tail probabilities.

With weaker moment bounds on J_n we get weaker bounds on Δ_n. Remember that for each p with $1 \le p < \infty$ there is a constant C_p such that

$$\|Z\|_p \le C_p\|Z\|_\Psi$$

for every random variable Z. In particular,

$$\mathbb{P}_\sigma|L_n|^p \le (C_p.J_n(\omega))^p,$$

which gives

$$\mathbb{P}|\Delta_n|^p \le (2C_p)^p\,\mathbb{P}J_n^p. \tag{7.5}$$

This inequality will be most useful for p equal to 1 or 2.

The preceding inequalities show that the behavior of the random variable $J_n(\omega)$ largely determines the form of the maximal inequality for the partial-sum process. In one very common special case, which is strongly recommended by the results from Section 4, the behavior of J_n is controlled by the envelope $\mathbf{F}_n(\omega)$. Let us suppose that $\lambda_n(\cdot)$ is a deterministic function for which

$$D(x|\mathbf{F}_n(\omega)|, \mathcal{F}_{n\omega}) \le \lambda_n(x) \qquad \text{for } 0 < x \le 1 \text{ and all } \omega. \tag{7.6}$$

Because $\mathcal{F}_{n\omega}$ lies within a ball of radius $|\mathbf{F}_n(\omega)|$, we could always choose $\lambda_n(x)$ equal to $(3/x)^n$. [We can pack $D(x|\mathbf{F}_n(\omega)|, \mathcal{F}_{n\omega})$ many disjoint balls of radius $1/2x|\mathbf{F}_n(\omega)|$

into the ball of radius $3/2|\mathbf{F}_n(\omega)|$.] To be of any real use, however, the λ_n function should not increase so rapidly with n. For example, if there is a fixed V such that each $\mathcal{F}_{n\omega}$ has pseudodimension V we could choose $\lambda_n(x) = Ax^{-W}$, with A and W depending only on V, which would lead to quite useful bounds. In any case, we may always assume that $\sqrt{\log \lambda_n}$ is integrable, which ensures that the function defined by

$$\Lambda_n(t) = \int_0^t \sqrt{\log \lambda_n(x)}\, dx \qquad \text{for } 0 \le t \le 1$$

is well defined and finite. A simple change of variable in the integral that defines $J_n(\omega)$ now gives

$$\text{(7.7)} \quad \begin{aligned} J_n(\omega) &\le 9|\mathbf{F}_n(\omega)|\Lambda_n\big(\delta_n(\omega)/|\mathbf{F}_n(\omega)|\big) \\ &\le 9\Lambda_n(1)|\mathbf{F}_n(\omega)| \qquad \text{because } |\mathbf{f}| \le |\mathbf{F}_n(\omega)| \text{ for every } \mathbf{f} \text{ in } \mathcal{F}_{n\omega}. \end{aligned}$$

When expressed in terms of Λ_n the inequalities for Δ_n take a particularly simple form. Suppose, for example, the envelope functions $F_i(\omega)$ are uniformly bounded, say $F_i(\omega) \le 1$ for each i and each ω. Then $J_n(\omega)$ is bounded by $9\sqrt{n}\Lambda_n(1)$. If $\Lambda_n(1)$ stays bounded as $n \to \infty$, the standardized processes

$$\frac{1}{\sqrt{n}}\Delta_n(\omega) = \frac{1}{\sqrt{n}} \sup_t |S_n(\omega, t) - M_n(t)|$$

will have uniformly subgaussian tails.

If instead of being uniformly bounded the random variables F_i^2 have uniformly bounded moment generating functions in a neighborhood of the origin, and if $\Lambda_n(1)$ stays bounded as $n \to \infty$, we get another useful bound on the Ψ norms of the J_n. For suppose that

$$\mathbb{P}\exp(\epsilon F_i^2) \le K \qquad \text{for all } i.$$

Then there is a constant K', depending on K and ϵ, such that

$$\mathbb{P}\exp(sF_i^2) \le 1 + K's \qquad \text{for } 0 \le s \le \epsilon \text{ and all } i.$$

With $C = 9\sup_n \Lambda_n(1)$, independence of the F_i gives, for $C' \ge C^2/n\epsilon$,

$$\begin{aligned} \mathbb{P}\exp(J_n^2/nC') &\le \prod_{i \le n} \mathbb{P}\exp(C^2F_i^2/nC') \\ &\le (1 + K'C^2/nC')^n. \end{aligned}$$

Certainly for $C' \ge K'C^2/\log 5$ the last bound is less than 5. It follows that

$$\|J_n\|_\Psi \le K''\sqrt{n} \qquad \text{for some constant } K'',$$

which guarantees a uniform exponential bound for the tail probabilities of the partial-sum processes with the usual standardization.

Finally, even with only moment bounds for the envelopes we still get usable maximal inequalities. For $1 \le p < \infty$, inequalities (7.5) and (7.7) give

$$\text{(7.8)} \qquad \begin{aligned} \mathbb{P}\sup_t |S_n(\cdot, t) - M_n(t)|^p &\le (18C_p)^p\, \mathbb{P}\,|\mathbf{F}_n|^p \Lambda_n\big(\delta_n/|\mathbf{F}_n|\big)^p \\ &\le \big(18C_p\Lambda_n(1)\big)^p\, \mathbb{P}\,|\mathbf{F}_n|^p. \end{aligned}$$

In applications such moment bounds are often the easiest to apply, typically for p equal to 1 or 2. They show that, in some sense, the whole process is only as badly behaved as its envelope.

The special cases considered above show that maximal inequalities for Δ_n can be derived from uniform bounds on the random packing numbers $D(x|\mathbf{F}_n(\omega)|, \mathcal{F}_{n\omega})$. The concept of *manageability* formalizes this idea. To accommodate a wider range of applications, let us expand the setting to cover triangular arrays of random processes,

$$\{f_{ni}(\omega,t) : t \in T,\ 1 \le i \le k_n\} \qquad \text{for } n = 1,2,\ldots,$$

independent within each row. Now $S_n(\omega,t)$ denotes the sum across the n^{th} row.To facilitate application of the stability arguments, let us also allow for nonnegative rescaling vectors.

(7.9) DEFINITION. Call a triangular array of processes $\{f_{ni}(\omega,t)\}$ *manageable* (with respect to the envelopes $\mathbf{F}_n(\omega)$) if there exists a deterministic function λ, the *capacity bound*, for which

(i) $\int_0^1 \sqrt{\log \lambda(x)}\, dx < \infty$,

(ii) $D(x|\boldsymbol{\alpha} \odot \mathbf{F}_n(\omega)|, \boldsymbol{\alpha} \odot \mathcal{F}_{n\omega}) \le \lambda(x)$ for $0 < x \le 1$, all ω, all vectors $\boldsymbol{\alpha}$ of nonnegative weights, and all n.

Call a sequence of processes $\{f_i\}$ manageable if the array defined by $f_{ni} = f_i$ for $i \le n$ is manageable.

In the special case where $\lambda(x) = A(1/x)^W$ for constants A and W, the processes will be called *Euclidean*. Most of the the applications in the final sections of these notes will involve Euclidean processes.

The inequalities developed in this section all carry over to the more general setting. In particular, for a manageable array there is a continuous, increasing function Λ with $\Lambda(0) = 0$, for which the analogue of (7.8) holds: for $1 \le p < \infty$ there exists a constant K_p such that

$$\text{(7.10)} \qquad \mathbb{P}\sup_t |S_n(\cdot,t) - M_n(t)|^p \le K_p \mathbb{P}|\mathbf{F}_n|^p \Lambda(\delta_n/|\mathbf{F}_n|)^p \\ \le K_p \Lambda(1)^p\, \mathbb{P}\,|\mathbf{F}_n|^p.$$

REMARKS. When specialized to empirical processes, the exponential inequality (7.3) is inferior to the results of Alexander (1984) and Massart (1986). By refinement of the approach in this section my inequality could be improved. However, a reader interested in better bounds would be well advised to first consult the book of Ledoux and Talagrand (1990).