SECTION 2

Symmetrization and Conditioning

In this section we begin the task of bounding $\mathbb{P}\,\Phi(\sup_t |S_n(\cdot,t) - M_n(t)|)$ for a general convex, increasing function Φ on $\mathbb{R}^+$. The idea is to introduce more randomness into the problem and then work conditionally on the particular realization of the $\{f_i\}$. This is somewhat akin to the use of randomization in experimental design, where one artificially creates an extra source of randomness to ensure that test statistics have desirable behavior conditional on the experimental data.

As a convenience for describing the various sources of randomness, suppose that the underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is a product space,

$$\Omega = \Omega_1 \otimes \cdots \otimes \Omega_n \otimes \Omega_1' \otimes \cdots \otimes \Omega_n' \otimes \mathcal{S},$$

equipped with a product measure

$$\mathbb{P} = \mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_n \otimes \mathbb{P}_1' \otimes \cdots \otimes \mathbb{P}_n' \otimes \mathbb{P}_\sigma.$$

Here $\Omega_i' = \Omega_i$ and $\mathbb{P}_i' = \mathbb{P}_i$. The set $\mathcal{S}$ consists of all n-tuples $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$ with each σ_i either $+1$ or -1, and $\mathbb{P}_\sigma$ is the uniform distribution, which puts mass 2^{-n} on each n-tuple.

Let the process $f_i(\cdot,t)$ depend only on the coordinate ω_i in Ω_i; with a slight abuse of notation write $f_i(\omega_i,t)$. The Ω_i' and $\mathbb{P}_i'$ are included in order to generate an independent copy $f_i(\omega_i',t)$ of the process. Under $\mathbb{P}_\sigma$, the σ_i are independent sign variables. They provide the randomization for the symmetrized process

$$S_n^\circ(\omega,t) = \sum_{i\le n} \sigma_i f_i(\omega_i,t).$$

We will find that this process is more variable than S_n, in the sense that

$$\mathbb{P}\,\Phi(\sup_t |S_n(\cdot,t) - M_n(t)|) \le \mathbb{P}\,\Phi(2 \sup_t |S_n^\circ(\cdot,t)|) \tag{2.1}$$

The proof will involve little more than an application of Jensen's inequality.

To take advantage of the product structure, rewrite the lefthand side of (2.1) as

$$\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_n \, \Phi\left(\sup_t \Big| \sum_{i \le n} [f_i(\omega_i, t) - \mathbb{P}_i' f_i(\omega_i', t)] \Big| \right).$$

We can replace the $\mathbb{P}_i'$ by $\mathbb{P}_1' \otimes \cdots \otimes \mathbb{P}_n'$, then pull that product measure outside the sum, without changing the value of this expression. The argument of Φ, and hence the whole expression, is increased if we change

$$\sup_t |\mathbb{P}_1' \otimes \cdots \otimes \mathbb{P}_n' \cdots | \qquad \text{to} \qquad \mathbb{P}_1' \otimes \cdots \otimes \mathbb{P}_n' \sup_t | \cdots |.$$

Jensen's inequality then gives the upper bound

$$\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_n \otimes \mathbb{P}_1' \otimes \cdots \otimes \mathbb{P}_n' \, \Phi\left(\sup_t \Big| \sum_{i \le n} f_i(\omega_i, t) - f_i(\omega_i', t) \Big| \right).$$

The last expression would be unaffected if we interchanged any ω_i with its ω_i', because $\mathbb{P}_i = \mathbb{P}_i'$. More formally, the $2n$-fold product measure is invariant under all permutations of the coordinates generated by interchanges of an ω_i with its ω_i'. For each $\boldsymbol{\sigma}$ in $\mathcal{S}$, the $2n$-fold expectation would be unchanged if the integrand were replaced by

$$\Phi\left(\sup_t \Big| \sum_{i \le n} \sigma_i [f_i(\omega_i, t) - f_i(\omega_i', t)] \Big| \right),$$

which, because Φ is convex and increasing, is less than

$$\frac{1}{2}\Phi\left(2 \sup_t \Big| \sum_{i \le n} \sigma_i f_i(\omega_i, t) \Big| \right) + \frac{1}{2}\Phi\left(2 \sup_t \Big| \sum_{i \le n} \sigma_i f_i(\omega_i', t) \Big| \right).$$

These two terms have the same $2n$-fold expectation. Averaging over all choices of $\boldsymbol{\sigma}$, ω_i, and ω_i', we arrive at a $(2n+1)$-fold expectation that is equal to the righthand side of the symmetrization inequality (2.1). Notice that the auxiliary ω_i' randomization has disappeared from the final bound, which involves only $\omega = (\omega_1, ..., \omega_n)$ and $\boldsymbol{\sigma}$.

For each ω, the sample paths of the processes trace out a subset

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

of $\mathbb{R}^n$. Consolidating the product $\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_n$ into a single $\mathbb{P}_\omega$, and reexpressing inequality (2.1) in terms of the usual inner product on $\mathbb{R}^n$, we get a neater looking bound.

(2.2) THEOREM. *For each convex, increasing Φ,*

$$\mathbb{P}\,\Phi(\sup_t |S_n(\cdot, t) - M_n(t)|) \le \mathbb{P}_\omega \mathbb{P}_\sigma \, \Phi\left(2 \sup_{\mathbf{f} \in \mathcal{F}_\omega} |\boldsymbol{\sigma} \cdot \mathbf{f}| \right). \qquad \square$$

The inner expectation, with respect to $\mathbb{P}_\sigma$, involves a very simple process indexed by a (random) subset $\mathcal{F}_\omega$ of $\mathbb{R}^n$. The fact that T indexes the points of the sets $\mathcal{F}_{n\omega}$ now becomes irrelevant. The sets themselves summarize all we need to know about the $\{f_i(\omega, t)\}$ processes. If we absorb the factor 2 into the function Φ, the problem has now become: find bounds for $\mathbb{P}_\sigma \, \Phi \, (\sup_{\mathcal{F}} |\boldsymbol{\sigma} \cdot \mathbf{f}|)$ for various convex Φ and various subsets $\mathcal{F}$ of $\mathbb{R}^n$.

REMARKS. There are many variations on the symmetrization method in the empirical process literature. In the original paper of Vapnik and Červonenkis (1971) the symmetrized process was used to bound tail probabilities. I learned about the simplifications arising from the substitution of moments for tail probabilities by reading the papers of Pisier (1983) and Giné and Zinn (1984). Symmetrization via moments also works with more complicated processes, for which tail probabilities are intractable, as in the papers of Nolan and Pollard (1987, 1988) on U-processes. In their comments on Pollard (1989), Giné and Zinn have traced some of the earlier history of symmetrization, with particular reference to the theory of probability in Banach spaces.