SECTION 6

Convex Hulls

Sometimes interesting random processes are expressible as convex combinations of more basic processes. For example, if $0 \le f_i \le 1$ for each i then the study of f_i reduces to the study of the random sets $\{\omega : s \le f_i(\omega, t)\}$, for $0 \le s \le 1$ and $t \in T$, by means of the representation

$$f_i(\omega, t) = \int_0^1 \{s \le f_i(\omega, t)\}\, ds.$$

More generally, starting from $f_i(\omega, t)$ indexed by T, we can construct new processes by averaging out over the parameter with respect to a probability measure Q on T:

$$f_i(\omega, Q) = \int f_i(\omega, t) Q(dt).$$

[This causes no measure-theoretic difficulties if there is a σ-field $\mathcal{T}$ on T such that f_i is jointly measurable in ω and t and Q is defined on $\mathcal{T}$.] Let us denote the corresponding process of sums by $S_n(\omega, Q)$, and its expectation by $M_n(Q)$. Because

$$|S_n(\omega, Q) - M_n(Q)| \le \int \sup_t |S_n(\omega, t) - M_n(t)| Q(dt),$$

it is easy to verify that

$$\sup_Q |S_n(\omega, Q) - M_n(Q)| = \sup_t |S_n(\omega, t) - M_n(t)|. \tag{6.1}$$

Some uniformity results for the processes indexed by probability measures on T follow trivially from uniformity results for processes indexed by T.

The operation of averaging out over t corresponds to the formation of convex combinations in $\mathbb{R}^n$. The vectors with coordinates $f_1(\omega, Q), \ldots, f_n(\omega, Q)$ all lie within the closed convex hull $\overline{co}(\mathcal{F}_\omega)$ of the set $\mathcal{F}_\omega$. The symmetrization analogue of the equality (6.1) is

$$\sup_{\overline{co}(\mathcal{F}_\omega)} |\boldsymbol{\sigma} \cdot \mathbf{f}| = \sup_{\mathcal{F}_\omega} |\boldsymbol{\sigma} \cdot \mathbf{f}|,$$

which suggests that there might be a connection between the packing numbers for $\mathcal{F}_\omega$ and the packing numbers for $\overline{co}(\mathcal{F}_\omega)$. A result of Dudley (1987) establishes such a connection for the ubiquitous case of sets whose packing numbers grow like a power of $1/\epsilon$. Even though inequality (6.1) makes the result slightly superfluous for the purposes of these lecture notes, it is worth study as a beautiful example of a probabilistic method for proving existence theorems.

The result could be stated in great generality—for Hilbert spaces, or even for "spaces of type 2"—but the important ideas all appear for the simple case of a bounded subset of Euclidean space.

(6.2) THEOREM. *Let $\mathcal{F}$ be a subset of the unit ball in a Euclidean space. If there exist constants A and W such that*

$$D_2(\epsilon, \mathcal{F}) \leq A(1/\epsilon)^W \qquad \text{for } 0 < \epsilon \leq 1,$$

then for each τ with $2 > \tau > \frac{2W}{2+W}$,

$$D_2(\epsilon, \overline{co}(\mathcal{F})) \leq \exp(C(1/\epsilon)^\tau) \qquad \text{for } 0 < \epsilon \leq 1.$$

for some constant C that depends only on A, W and τ. □

Note that the inequality $2 > \tau$ ensures

$$\int_0^1 \sqrt{\log D_2(x, \overline{co}(\mathcal{F}))}\, dx < \infty.$$

Indeed $\tau = 2$ represents the critical value at which the integral would diverge. For these notes the theorem has one major application, which deserves some attention before we get into the details of the proof for Theorem 6.2.

(6.3) EXAMPLE. Let $\mathcal{F}$ be a bounded subset of $\mathbb{R}^n$ with envelope $\mathbf{F}$. The convex cone generated by $\mathcal{F}$ is the set $\mathcal{G} = \{r\mathbf{f} : r > 0, \mathbf{f} \in \mathcal{F}\}$. Suppose $\mathcal{G}$ has the property: *for some integer V, no $(V+1)$-dimensional coordinate projection of $\mathcal{G}$ can surround the corresponding projections of $\mathbf{F}$ or $-\mathbf{F}$.* Then Theorem 6.2 and the results from Section 4 will imply that $D_2(\epsilon|\boldsymbol{\alpha} \odot \mathbf{F}|_2, \boldsymbol{\alpha} \odot \mathcal{F}) \leq \exp[C(1/\epsilon)^\tau]$ for $0 < \epsilon \leq 1$ and all nonnegative $\boldsymbol{\alpha}$, with constants C and $\tau < 2$ depending only on V.

Without loss of generality, suppose $\boldsymbol{\alpha}$ has all components equal to one, and $\mathcal{F}$ is a subset of the positive orthant with $F_i > 0$ for each i. [The projection property of $\mathcal{G}$ still holds if we replace each $\mathbf{f}$ by the vector with coordinates f_i^+ or the vector with coordinates f_i^-.] By a trivial rescaling, replacing $\mathbf{f}$ by $\mathbf{f}/|\mathbf{F}|_2$, we may also assume that $|\mathbf{F}|_2 = 1$, so that $\mathcal{F}$ is a subset of the unit ball.

Define a new set $\mathcal{H}$ of all vectors with coordinates of the form

$$h(r, \mathbf{f})_i = F_i\{rf_i \geq F_i\},$$

where r ranges over positive real numbers and $\mathbf{f}$ ranges over $\mathcal{F}$. Certainly $\mathcal{H}$ is a subset of the unit ball. Its closed convex hull contains $\mathcal{F}$, because

$$f_i = \int_0^1 F_i\{f_i > sF_i\}\, ds$$

for every nonnegative f_i. We have only to check that

$$D_2(\epsilon, \mathcal{H}) \leq A(1/\epsilon)^W \qquad \text{for } 0 < \epsilon \leq 1$$

then appeal to Theorem 6.2.

The geometric bound for packing numbers of $\mathcal{H}$ will follow from the results in Section 4 if we show that $(V+1)$-dimensional proper coordinate projections of $\mathcal{H}$ cannot surround any $\mathbf{t}$ in $\mathbb{R}^{V+1}$. Let I be the set of $V+1$ coordinates that defines the projection. Let J be the orthant of the I-projections of $\mathbf{F}$ that the I-projection of $\mathcal{G}$ cannot occupy. Suppose, however, that the I-projection of $\mathcal{H}$ does surround some point $\mathbf{t}$. This could happen only if $0 < t_i < F_i$ for each i. For the projection of the vector $\mathbf{h}(r, \mathbf{f})$ to occupy orthant J of $\mathbf{t}$ we would need to have

$$\begin{aligned} rf_i &\geq F_i \qquad \text{for } i \in J, \\ rf_i &< F_i \qquad \text{for } i \in I \backslash J. \end{aligned}$$

Increasing r slightly to make these inequalities strict, we would then have found a projection of a vector in $\mathcal{G}$ occuping the orthant J. The contradiction establishes the desired projection property for $\mathcal{H}$, and hence leads to the asserted rate of growth for the packing numbers of $\mathcal{F}$. □

PROOF OF THEOREM 6.2. We may as well assume that $\mathcal{F}$ is compact, because packing numbers for a set always agree with packing numbers for its closure. This makes $\overline{co}(\mathcal{F})$ the same as the convex hull $co(\mathcal{F})$, which will slightly simplify the argument.

By a succession of approximations, we will be able to construct a set with cardinality at most $\exp(C(1/\epsilon)^\tau)$ that approximates each vector of $\mathcal{F}$ within an ℓ_2 distance less than 4ϵ. With some adjustment of the constant C after replacement of ϵ by $\epsilon/8$, this would give the asserted bound for the packing numbers.

In what follows the ℓ_2 norm will be denoted by $|\cdot|$, without the subscript 2.

Let $\alpha = 2/(2+W)$. Choose a maximal subset $\mathcal{F}_\epsilon$ of points from $\mathcal{F}$ at least ϵ apart, then let $\{\phi_1, \ldots, \phi_m\}$ be a maximal subset of $\mathcal{F}_\epsilon$ with points at least ϵ^α apart. By assumption,

$$\begin{aligned} m &\leq D_2(\epsilon^\alpha, \mathcal{F}) \leq A(1/\epsilon^\alpha)^W, \\ \#\mathcal{F}_\epsilon &\leq D_2(\epsilon, \mathcal{F}) \leq A(1/\epsilon)^W. \end{aligned}$$

Notice that m is smaller than $A(1/\epsilon)^\tau$; the exponent of $1/\epsilon$ is $2W/(2+W)$, which is less than τ. Each $\mathbf{f}$ in $\mathcal{F}$ lies within ϵ of some $\mathbf{f}^*$ in $\mathcal{F}_\epsilon$. Each finite convex combination $\sum_{\mathcal{F}} \theta(\mathbf{f})\mathbf{f}$ lies within ϵ of the corresponding $\sum_{\mathcal{F}} \theta(\mathbf{f})\mathbf{f}^*$. (Here the $\theta(\mathbf{f})$ multipliers denote nonnegative numbers that sum to one, with $\theta(\mathbf{f}) \neq 0$ for only finitely many $\mathbf{f}$.) It therefore suffices to construct approximations within 3ϵ to the vectors in $co(\mathcal{F}_\epsilon)$.

Because each vector in $\mathcal{F}_\epsilon$ lies within ϵ^α of some ϕ_i, there exists a partition of $\mathcal{F}_\epsilon$ into subsets $\mathcal{E}_1, \ldots, \mathcal{E}_m$ for which

$$|\mathbf{f} - \phi_i| \leq \epsilon^\alpha \qquad \text{if } \mathbf{f} \in \mathcal{E}_i. \tag{6.4}$$

Each convex combination $\Sigma\theta(\mathbf{f})\mathbf{f}$ from $co(\mathcal{F}_\epsilon)$ can then be reexpressed as a convex combination of vectors from the convex hulls $co(\mathcal{E}_i)$:

$$\sum_{\mathbf{f} \in \mathcal{F}_\epsilon} \theta(\mathbf{f})\mathbf{f} = \sum_{i \leq m} \lambda_i \mathbf{e}_i,$$

where

$$\lambda_i = \sum_{\mathbf{f} \in \mathcal{E}_i} \theta(\mathbf{f})$$

and

$$\mathbf{e}_i = \sum_{\mathbf{f} \in \mathcal{E}_i} \frac{\theta(\mathbf{f})}{\lambda_i} \mathbf{f}.$$

Here the vector $\boldsymbol{\lambda}$ of convex weights ranges over the m-dimensional simplex

$$\Lambda = \{\boldsymbol{\lambda} \in \mathbb{R}^m : \lambda_i \geq 0 \quad \text{for all } i, \text{ and } \sum_i \lambda_i = 1\}.$$

Because $\mathcal{E}_i$ lies inside the unit ball,

$$\Big|\sum_{i \leq m} \lambda_i \mathbf{e}_i - \sum_{i \leq m} \mu_i \mathbf{e}_i\Big| \leq \sum_{i \leq m} |\lambda_i - \mu_i|.$$

We can therefore approximate each point in $co(\mathcal{F}_\epsilon)$ within ϵ by means of a convex combination with weights $\boldsymbol{\lambda}$ chosen from a maximal subset Λ_ϵ of points from Λ at least ϵ apart in ℓ_1 distance. Notice that

$$\#\Lambda_\epsilon \leq (4/\epsilon)^m,$$

because the ℓ_1 balls of radius $\epsilon/2$ about each point in Λ_ϵ are pairwise disjoint, and their union lies within an ℓ_1 ball of radius 2.

Fix a $\boldsymbol{\lambda}$ in Λ_ϵ. Define positive integers $n(1), \ldots, n(m)$ by

$$\lambda_i(1/\epsilon)^{2-2\alpha} < n(i) \leq 1 + \lambda_i(1/\epsilon)^{2-2\alpha}.$$

Let $\Phi(\boldsymbol{\lambda})$ denote the set of all convex combinations

$$\sum_{i \leq m} \lambda_i \overline{\mathbf{y}}_i$$

with $\overline{\mathbf{y}}_i$ a simple average of $n(i)$ vectors from $\mathcal{E}_i$. Its cardinality is bounded by the number of ways to choose all the averages,

$$\#\Phi(\boldsymbol{\lambda}) \leq \prod_{i \leq m} (\#\mathcal{F}_\epsilon)^{n(i)}.$$

The upper bound has logarithm less than

$$\sum_{i \leq m} n(i) \log[A(1/\epsilon)^W] \leq (m + (1/\epsilon)^{2-2\alpha}) \log[A(1/\epsilon)^W].$$

The nicest part of the argument will show, for each $\boldsymbol{\lambda}$, that each convex combination $\sum_i \lambda_i \mathbf{e}_i$ from $co(\mathcal{F}_\epsilon)$ can be approximated within 2ϵ by a vector in $\Phi(\boldsymbol{\lambda})$. Hence the union of the $\Phi(\boldsymbol{\lambda})$ as $\boldsymbol{\lambda}$ ranges over Λ_ϵ will approximate to the whole of $co(\mathcal{F}_\epsilon)$ within 3ϵ. The cardinality of this union is at most

$$(\#\Lambda_\epsilon) \max_{\boldsymbol{\lambda} \in \Lambda_\epsilon} \#\Phi(\boldsymbol{\lambda}),$$

which has logarithm less than

$$m \log(4/\epsilon) + [m + (1/\epsilon)^{2-2\alpha}] \log[A(1/\epsilon)^W].$$

The small interval between τ and

$$\frac{2W}{2+W} = \alpha W = 2 - 2\alpha$$

absorbs the factors of $\log(1/\epsilon)$, leading to the desired bound, $C(1/\epsilon)^\tau$, for an appropriately large constant C.

It remains only to prove the assertion about the approximation properties of $\Phi(\boldsymbol{\lambda})$, for a fixed $\boldsymbol{\lambda}$ in Λ_ϵ. Given $\mathbf{e}_i$ from $co(\mathcal{E}_i)$, we need to find simple averages $\overline{\mathbf{y}}_i$ of $n(i)$ vectors from $\mathcal{E}_i$ such that

$$\Big|\sum_{i\le m} \lambda_i \mathbf{e}_i - \sum_{i\le m} \lambda_i \overline{\mathbf{y}}_i\Big| \le 2\epsilon.$$

Existence of such $\overline{\mathbf{y}}_i$ will be established probabilistically, by means of randomly generated vectors $\overline{\mathbf{Y}}_i$ for which

$$\mathbb{P}\Big|\sum_{i\le m} \lambda_i \mathbf{e}_i - \sum_{i\le m} \lambda_i \overline{\mathbf{Y}}_i\Big|^2 \le 4\epsilon^2.$$

Some realization of the $\overline{\mathbf{Y}}_i$ must satisfy the desired inequality.

Each $\mathbf{e}_i$, as a vector in $co(\mathcal{E}_i)$, has a representation as a convex combination

$$\mathbf{e}_i = \sum_{\mathbf{f}\in\mathcal{E}_i} p_i(\mathbf{f})\mathbf{f}.$$

Interpret $p_i(\cdot)$ as a probability distribution on $\mathcal{E}_i$. Generate independent random vectors $\mathbf{Y}_{ij}$, for $j = 1, \ldots, n(i)$ and $i = 1, \ldots, m$, with

$$\mathbb{P}\{\mathbf{Y}_{ij} = \mathbf{f}\} = p_i(\mathbf{f}) \qquad \text{for } \mathbf{f} \in \mathcal{E}_i.$$

By this construction and inequality (6.4),

$$\begin{aligned} &\mathbb{P}\mathbf{Y}_{ij} = \mathbf{e}_i, \\ &\mathbb{P}|\mathbf{Y}_{ij} - \mathbf{e}_i|^2 \le (\operatorname{diam} \mathcal{E}_i)^2 \le 4\epsilon^{2\alpha}. \end{aligned}$$

Define $\overline{\mathbf{Y}}_i$ to be the average of the $\mathbf{Y}_{ij}$ for $j = 1, \ldots, n(i)$. With independence accounting for the disappearance of the crossproduct terms we get

$$\begin{aligned} \mathbb{P}\Big|\sum_{i\le m} \lambda_i(\mathbf{e}_i - \overline{\mathbf{Y}}_i)\Big|^2 &= \sum_{i\le m} \lambda_i^2 \mathbb{P}|\mathbf{e}_i - \overline{\mathbf{Y}}_i|^2 \\ &\le \sum_{i\le m} \lambda_i^2 4\epsilon^{2\alpha}/n(i). \end{aligned}$$

Our choice of $n(i)$ lets us bound $\lambda_i/n(i)$ by $\epsilon^{2-2\alpha}$, then sum over the remaining λ_i to end up with the desired $4\epsilon^2$. □

To generalize the result to subsets $\mathcal{F}$ of more general normed linear spaces, we would need only to rejustify the last few assertions in the proof regarding the $\overline{\mathbf{Y}}_i$. Certainly the necessary cancellations are still valid for any Hilbert space. Type 2 spaces (Araujo and Giné 1980, page 158) enjoy a similar bound for $\mathcal{L}^2$ norms of sums of independent random elements, essentially by definition of the type 2 property.

REMARKS. The property of $\mathcal{F}$ introduced in Example 6.3 corresponds to the *VC major* property for classes of functions, studied by Dudley (1987). My example merely translates his result for empirical processes indexed by classes of functions to the more general setting, relaxing his assumption of bounded envelopes.

Dudley (1985) has shown that the Donsker-class property is preserved under the formation of (sequential closures of) convex hulls of classes of functions. (See the notes to Section 10 for more about Donsker classes.) This gives yet another way of handling processes representable as convex combinations of simpler processes. The same stability property is also implied by the first theorem of Talagrand (1987).