Project 3

Empirical processes

3.1 Progress so far

The main lemma from Project 2 can be restated in a slightly different form.

<1> **Lemma.** *Let $\{X_n : n \in \mathbb{N}\}$ be a sequence of maps from Ω into $\mathcal{X}$ and let X_∞ be a $\mathcal{B}(\mathcal{X})$-measurable random element of $\mathcal{X}$. Suppose $\{\epsilon_k : k \in \mathbb{N}\}$ and $\{\delta_k : k \in \mathbb{N}\}$ are both sequences of positive numbers that converge to zero as $k \to \infty$. For each k suppose there exist maps $X_{n,k}$ and a $\mathcal{B}(\mathcal{X})$-measurable $X_{\infty,k}$ for which*

(i) $X_{\infty,k} \rightsquigarrow X_\infty$ *as* $k \to \infty$

(ii) $X_{n,k} \rightsquigarrow X_{\infty,k}$ *as* $n \to \infty$, *for each fixed* k

(iii) $\limsup_{n\to\infty} \mathbb{P}^*\{d(X_n, X_{n,k}) > \delta_k\} < \epsilon_k$ *for each fixed* k

Then $X_n \rightsquigarrow X_\infty$.

In applying this Lemma to establish convergence in distribution of the partial-sum process X_n to a Brownian motion X_∞ we used $X_{n,k} = L_{J(k)}X_n$ and $X_{\infty,k} = L_{J(k)}X_\infty$, with $J(k) = \{j/k := 0, 1, \dots, k\}$ and $L_{J(k)}$ a continuous map from $D[0,1]$ into $D[0,1]$ defined as follows.

- For each finite $J \subseteq [0,1]$ we wrote π_J for the map that projects $D[0,1]$ into $\mathbb{R}^J$ and we let A_J denote the map from $\mathbb{R}^J$ into $D[0,1]$ defined by

$$A_J(z) := z(t_{k+1})\{t = 1\} + \sum\nolimits_{i=0}^{k} z(t_i)\{t_i \le t < t_{i+1}\} \qquad \text{for } 0 \le t \le 1$$
$$\text{if } z \in \mathbb{R}^J \text{ and } J = \{t_0, \dots, t_{k+1}\} \text{ with } 0 = t_0 < \dots < t_{k+1} = 1.$$

  We defined $L_J = A_J \circ \pi_J$, a map from $D[0,1]$ into $D[0,1]$.

- For $J = \{t_0, \dots, t_{k+1}\}$ with $0 = t_0 < \dots < t_{k+1} = 1$ we defined $I_j := [t_j, t_{j+1}]$ then showed that

$$d(x, L_J x) \le \Delta_J(x) := \max\nolimits_{j=0}^{k} \sup\nolimits_{t\in I_j} |x(t) - x(t_j)| \qquad \text{for each } x \text{ in } D[0,1].$$

The application of the Lemma then amounted to showing that

(i) $\pi_J X_n \rightsquigarrow \pi_J X_\infty$ as $n \to \infty$, for each finite J

(ii) $d(L_{J(k)}x, x) \to 0$ for each x in $C[0,1]$, a subset of $D[0,1]$ in which the sample paths of X_∞ concentrate.

(iii) For each $\delta > 0$ and $\epsilon > 0$ there exists a k for which

$$\limsup_{n\to\infty} \mathbb{P}^*\{\Delta_{J(k)}(X_n) > \delta\} < \epsilon$$

> In fact this argument works for more general sequences of processes with paths in $D[0,1]$. Your first task: Turn the argument into a theorem that handles not only the partial sum process but also the empirical process described in Section 3.

3.2 A maximal inequality

You will find the following inequality useful in Section 3.

Suppose $\{Z_t : 0 \le t \le b\}$ is a stochastic process with sample paths in $D[0,b]$. Suppose also that Z is adapted to a filtration $\{\mathcal{F}_t : 0 \le t \le b\}$, that is, Z_t is $\mathcal{F}_t$-measurable for each t. Write $\mathbb{P}_t(\dots)$ instead of $\mathbb{P}(\cdots \mid \mathcal{F}_t)$.

<2> **Lemma.** *For each $\delta > 0$ suppose there exists a constant β (depending on δ) for which $\mathbb{P}_s\{|Z_b - Z_s| \le \frac{1}{2}|Z_s|\} \ge 1/\beta$ almost surely on the set $\{|Z_s| > 2\delta\}$, for each s. Then*

$$\mathbb{P}\{\sup_{0\le t\le b} |Z_t| > 2\delta\} \le \mathbb{P}\{|Z_b| > \delta\}.$$

Proof (outline) Reduce to the case of a maximium taken over a finite subset S consisting of points $0 = s_0 < s_1 < \cdots < s_k = b$. Define (a stopping time?) $\tau = \inf\{s \in S : |Z_s| > 2\delta\}$. Show that

$$\mathbb{P}\{|Z_b| > \delta\}\{\tau = s\} \ge \beta\mathbb{P}\{\tau = s\} \qquad \text{for each } s \in S.$$

Then sum over s in S.

□

3.3 Uniform empirical process

Let $\xi_1, \xi_2, \dots$ be independent Uniform$[0,1]$ distributed random variables. Define the ***uniform empirical distribution function*** to be

$$U_n(t,\omega) := n^{-1} \sum_{i\le n} \{\xi_i(\omega) \le t\} \qquad \text{for } 0 \le t \le 1$$

and the uniform empirical process to be $\nu_n(t,\omega) := \sqrt{n}\,(U_n(t,\omega) - t)$ for $0 \le t \le 1$.

See Pollard (1984, Chapter V) for ideas on how to show that ν_n converges in distribution to a Brownian bridge. (Be careful: There are some subtle errors in that Chapter. Also, in 1984, I was using a slightly different definition of convergence in distribution. You need to understand ideas, not just copy out proofs.) A scanned copy of the book is available (for free) at http://www.stat.yale.edu/~pollard/Books/1984book/ .

(i) Note that $\nu_n(0) = \nu_n(1) = 0$ Check that $\mathbb{P}\nu_n(t) = 0$ for each t and $\text{cov}(\nu_n(s), \nu_n(t)) = s \wedge t - st$ for $s, t \in [0,1]$.

(ii) Define the ***Brownian bridge*** process to be the random element of $C[0,1]$ obtained by "tying down" a Brownian motion B with continuous sample paths,

$$G_t = B_t - tB_1 \qquad \text{for } 0 \le t \le 1.$$

Note that $G_0 = G_1 = 0$. Show that $\mathbb{P}G_t = 0$ for each t and $\text{cov}(G_s, G_t) = s \wedge t - st$.

(iii) Invoke a multivariate CLT to show that $\pi_J \nu_n \rightsquigarrow \pi_J G$ as $n \to \infty$, for each finite subset J of $[0,1]$.

(iv) Use Lemma 2 to show that: For each $\delta > 0$ and $\epsilon > 0$ there exists a k for which

$$\limsup_{n\to\infty} \mathbb{P}^*\{\Delta_{J(k)}(\nu_n) > \delta\} < \epsilon.$$

(v) Complete the proof that $\nu_n \rightsquigarrow G$.

Exercise Show that $\int_0^1 \nu_n(t)^2 dt \rightsquigarrow \int_0^1 G_t^2 dt$

□

References

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer.