Statistics 607 **Due: Wednesday 9 February**
2005: Sheet 2

(2.1) Let $(\mathcal{X}, \mathcal{A}, P)$ be a probability space for which the corresponding $\mathcal{L}^2(P)$ has an orthonormal basis $\{\phi_i : i \in \mathbb{N}\}$. Define $\langle f, g\rangle := P(fg)$ and $\|f\| := \sqrt{Pf^2}$. You may assume these facts (cf. Pollard (2001, Appendix B)):

(a) The series $\sum_{i\in\mathbb{N}} \phi_i \langle f, \phi_i\rangle$ converges in $\mathcal{L}^2(P)$ norm, for each f in $\mathcal{L}^2(P)$

(b) for each pair f, g in $\mathcal{L}^2(P)$,

$$\langle f, g\rangle = \sum\nolimits_{i\in\mathbb{N}} \langle f, \phi_i\rangle\langle g, \phi_i\rangle$$

Let $\{\eta_i : i \in \mathbb{N}\}$ be a sequence of independent $N(0, 1)$ random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

(i) For each $f \in \mathcal{L}^2(P)$ show that the series $\sum_i \eta_i \langle f, \phi_i\rangle$ converges in $\mathcal{L}^2(\mathbb{P})$. Write Z_f for the limit. (More properly, choose Z_f from the equivalence class of possible limits.)

(ii) Show that $Z_f : f \in \mathcal{L}^2(P)\}$ is a centered Gaussian process with $\operatorname{cov}\left(Z_f, Z_g\right) = \langle f, g\rangle$.

(iii) Let $\mathcal{C}$ be a subset of $\mathcal{L}^2(P)$ with covering numbers $N(\cdot)$ for the metric d on $\mathcal{C}$ defined by the $\mathcal{L}^2(P)$ distance. Suppose

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

Show that there is a version of $\{Z_f : f \in \mathcal{C}\}$ with sample paths that are uniformly continuous (with respect to the d-metric). Compare with Dudley (1999, Section 2.5) and Dudley (1973).

(2.2) Let $F_n(t) := n^{-1}\sum_{i\le n}\{\xi_i \le t\}$ for $0 \le t \le 1$ be the empirical distribution function based on independent observation $\xi_1, \xi_2, \ldots$ from the Unif(0, 1) distribution. Define the empirical process $\nu_n(t) = \sqrt{n}(F_n(t) - t)$ for $0 \le t \le 1$.

(i) Show that $\operatorname{cov}\left(\nu_n(s), \nu_n(t)\right) = \min(s, t) - st$.

(ii) Show that there exists a centered Gaussian process $\{\nu(t) : 0 \le t \le 1\}$ with continuous sample paths and $\operatorname{cov}\left(\nu(s), \nu(t)\right) = \min(s, t) - st$. (The process ν is usually called a Brownian bridge.)

(iii) Show that

$$\mathbb{P}|\nu_n(s) - \nu_n(t)|^4 \le n^{-2}\left(np + 3(np)^2\right) \qquad \text{where } p := |s - t|.$$

(iv) Define $d(s, t) := \sqrt{|s - t|}$. Use facts about $\mathcal{L}^2$ norms to show that d is a metric on $[0, 1]$.

(v) For pairs s_i, t_i with $n^{-1} \le |s_i - t_i| \le \delta^2$, show that

$$\|\max_{i\le N} |\nu_n(s_i) - \nu_n(t_i)|\,\|_4 \le 2N^{1/4}\delta$$

(vi) Let $k(n)$ be the integer for which $n^{-1} \le 2^{-k(n)} \le 2n^{-1}$. Let $T_{k(n)} := \{j/2^{k(n)} : j = 1, 2, \ldots, 2^{k(n)}\}$. Given $\epsilon > 0$, show that there exists an η, ***which does not depend on*** n, for which

$$\mathbb{P}\{\sup\{|\nu_n(s) - \nu_n(t)| : d(s, t) < \eta \text{ and } s, t \in T_{k(n)}\} > \epsilon\} < \epsilon$$

(vii) Suppose $t, t' \in T_{k(n)}$ with $t' - t = 2^{-k(n)}$ and $t \le s < t'$. Show that

$$|\nu_n(s) - \nu_n(t)| \le |\nu_n(t') - \nu_n(t)| + 2/\sqrt{n}$$

(viii) Show that, for all n large enough, the inequality in (vi) can be extended to all pairs s, t in $[0, 1]$ for which $d(s, t) < \eta$, perhaps with a slight increase in ϵ.

(ix) For each real-valued function $\{x(t) : 0 \le t \le 1\}$ and $m \in \mathbb{N}$, define $t_i = i/m$ for $i = 0, 1, \ldots, m + 1$ and

$$(A_m x)(t) = x\left(t_i\right) \qquad \text{for } t_i \le t < t_{i+1}.$$

Show that $\|x - A_m x\| \le \sup\{|x(s) - x(t) : |s - t| \le m^{-1}\}$, where $\|z\| := \sup_{0\le t\le 1} |z(t)|$.

(x) For each $\epsilon > 0$ show that there exists an m, which depends on ϵ but ***not on*** n, for which

$$\mathbb{P}\{\|\nu_n - A_m\nu_n\| > \epsilon\} \leq \epsilon \qquad \text{for all } n \text{ large enough}$$

and

$$\mathbb{P}\{\|\nu - A_m\nu\| > \epsilon\} \leq \epsilon.$$

(xi) Suppose f is a real-valued functional defined at least for piecewise continuous functions on $[0, 1]$ such that

$$|f(x) - f(y) \leq \|x - y\|.$$

For each fixed m, show that $f(A_m\nu_n) \rightsquigarrow f(A_m\nu)$.

(xii) Deduce that $f(\nu_n) \rightsquigarrow f(\nu)$. (Do we need to worry about measurability?)

Congratulations. You have just proved the classical form of Donsker's theorem. See Doob (1949) and Donsker (1952) for the glorious beginnings. See also Le Cam (1986, Section) who pointed out that perhaps Kolmogorov deserved more of the credit. See Billingsley (1968, Section 10) for two elegant treatments (one based on Kolmogorov's method) and more about the history.

References

Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley, New York.

Donsker, M. D. (1952), 'Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems', *Annals of Mathematical Statistics* **23**, 277–281.

Doob, J. L. (1949), 'Heuristic approach to the Kolmogorov-Smirirnov theorems', *Annals of Mathematical Statistics* **20**, 393–403.

Dudley, R. M. (1973), 'Sample functions of the Gaussian process', *Annals of Probability* **1**, 66–103.

Dudley, R. M. (1999), *Uniform Central Limit Theorems*, Cambridge University Press.

Le Cam, L. (1986), 'The central limit theorem around 1935', *Statistical Science* **1**, 78–96.

Pollard, D. (2001), *A User's Guide to Measure Theoretic Probability*, Cambridge University Press.