spl13.tex
Lecture 13. 8.11.2010

8. The Borel-Cantelli lemmas and the zero-one law.

The following results are due to Borel in 1909, F. P. CANTELLI (1906-1985) in 1917.

Theorem (Borel-Cantelli lemmas). If A_n are independent events, $A := \limsup A_n = \{A_n \ i.o.\}$:
(i) If $\sum P(A_n) < \infty$, then $P(A) = 0$.
(ii) If $\sum P(A_n) = \infty$ and the A_n are independent, then $P(A) = 1$.

Proof. (i) $A = \limsup A_n = \cap_n \cup_{m=n}^{\infty} A_m$, so $A \subset \cup_{m=n}^{\infty} A_m$ for each n. So

$$P(A) \leq P(\cup_{m=n}^{\infty} A_m) \leq \sum_{m=n}^{\infty} P(A_m) \to 0 \qquad (n \to \infty)$$

(tail of a convergent series): $P(A) = 0$.
(ii) By the De Morgan laws, $A^c = \cup_n \cap_{m=n}^{\infty} A_m^c$. But for each n

$$\begin{aligned} P(\cap_{m=n}^{\infty} A_m^c) &= \lim_N P(\cap_{m=n}^{N} A_m^c) \qquad (\sigma\text{-additivity}) \\ &= \prod_{m=n}^{N} (1 - P(A_m)) \qquad (\text{independence}) \\ &\leq \prod_{m=n}^{N} \exp\{-P(A_m)\} \qquad (1 - x \leq e^{-x} \text{ for } x \geq 0) \\ &= \exp\{-\sum_{m=n}^{N} P(A_m)\} \\ &\to 0 \qquad (N \to \infty), \end{aligned}$$

as $\sum P(A_n)$ diverges. So $\cap_{m=n}^{\infty} A_m^c)$ is null. So their union $A^c = \cup_n \cap_{m=n}^{\infty} A_m^c$ is null, giving the result. //

Combining: if the A_n are independent, $P(A) = 0$ or 1 according as $\sum P(A_n)$ converges or diverges. More generally, call an event A depending on events A_n a *tail event* if it is invariant under deletion of finitely many of the A_n. Then *Kolmogorov's zero-one* law states that all tail events of independent events have probability 0 or 1.

9. Infinite product measures; replication and copies.

Independence corresponds to product measures (L9, L11); the construction of the product measure of two measure (I.7, L9) extends to finite products of measures by induction. We now consider the extension to infinite products. This will model generation of a sequence of independent identically distributed (iid) random variables, called *replications* or *copies*. Think of repeatedly tossing a coin, or repeatedly sampling in Statistics.

In fact the construction is a special case of a much more general construction (in which independence is not assumed), called the *Daniell-Kolmogorov theorem*, which we shall meet later in connection with Stochastic Processes (Ch. III). But for now, consider a sequence of measure spaces $(\Omega_n, \mathcal{A}_n, \mu_n)$, $n = 1, 2, \ldots$). We form the cartesian product $\Omega := \Omega_1 \times \ldots \times \Omega_n \times \ldots$. Call a set $A \subset \Omega$ a *cylinder set* if it is of the form $A = A_1 \times \ldots \times A_n \times \ldots$, with all but finitely many of the A_n, say $A_{n_1}, \ldots, A_{n_k}$, equal to Ω_n. Define a measure μ on the class $\mathcal{C}$ of such cylinder sets by

$$\mu(A) := \mu_{n_1}(A_{n_1}) \times \ldots \times \mu_{n_k}(A_{n_k})$$

(thus $\mu(A)$ expresses independence on the cylinder sets). The measure μ extends uniquely to a measure on the σ-field $\mathcal{A} := \sigma(\mathcal{C})$ generated by the cylinder sets. The resulting probability space is called the *infinite product* of the coordinate probability spaces, written

$$(\Omega, \mathcal{A}, \mu) = \times_{n=1}^{\infty} (\Omega_n, \mathcal{A}_n, \mu_n).$$

Example: Infinite coin tossing and the uniform distribution.
Take the Lebesgue probability space $([0,1], \mathcal{L}, \mu)$ modelling the uniform distribution $U[0,1]$ on the unit interval (probability = length). For a random variable $X \sim U[0,1]$, take its dyadic expansion

$$X = \sum_1^{\infty} \epsilon_n / 2^n.$$

Thus $\epsilon_1 = 0$ iff $X \in [0, 1/2)$, 1 iff $X \in [1/2, 1)$ (or $[1/2, 1]$: we can omit 1, as it carries 0 probability). If $\epsilon_1, \ldots, \epsilon_{n-1}$ are already defined, on the dyadic intervals $[k/2^{n-1}, (k+1)/2^{n-1})$, split each interval into two halves: $\epsilon_n = 0$ on the left half, 1 on the right half. This construction shows that $\epsilon_1, \ldots, \epsilon_n$ are independent, coin-tossing random variables (Bernoulli with parameter 1/2: take values 0, 1 with probability 1/2 each), for each n. So the ϵ_1 are independent coin-tosses.

Conversely, given ϵ_n independent coin tosses, form $X := \sum_1^\infty \epsilon_n/2^n$. Then $X_n := \sum_1^n \epsilon_k/2^k \to X$ a.s. Its distribution function has jumps $1/2^n$ at $k/2^n$, $k = 0, 1, \ldots, 2^n - 1$. This 'saw-tooth jump function' converges to x on $[0,1]$, the distribution function of $U[0,1]$. So $X \sim U[0,1]$. So:

If $X = \sum_1^\infty \epsilon_n/2^n$, $X \sim U[0,1]$ iff ϵ_n are independent coin tosses.

So the Lebesgue probability space models both length on the unit interval and infinitely many independent coin tosses. Incidentally, this shows that the hard Measure Theory content of construction of Lebesgue measure (Carathéodory's Extension Theorem, which we have quoted) is the same as that of the construction of the infinite product space for repeated coin tossing (which we have sketched above, and referred forward to the Daniell-Kolmogorov theorem – which we shall also quote).

We could instead let the ϵ_n take values ± 1 with probability 1/2. As one might expect, this leads instead to the uniform distribution $U[-1,1]$ (density 1/2 on $[-1,1]$). For such ϵ_n, the CF is $(e^{it} + e^{-it})/2 = \cos t$. So the CF of X_n above is

$$E \exp\{itX_n\} = E \exp\{it \sum_1^n \epsilon_k/2^k\} = \prod_1^n E \exp it\epsilon_k/2^k = \prod_1^n \cos(t/2^k).$$

Now

$$\sin t = 2 \cos t/2 \sin t/2 = \ldots = 2^n \cos t/2 \ldots \cos t/2^n \sin t/2^n.$$

So

$$E \exp\{itX_n\} = \frac{\sin t}{2^n \sin t/2^n} \to \frac{\sin t}{t} \qquad (n \to \infty),$$

the CF of $U[-1,1]$:

$$\int_{-1}^{1} e^{itx}.\frac{1}{2} dx = \frac{e^{it} - e^{-it}}{2it} = \frac{\sin t}{t}.$$

The case $[0,1]$ maps to the case $[-1,1]$ under the affine map $x \mapsto 2x - 1$.

The mathematics above yields infinite replication from the uniform distribution $U[0,1]$. Take the ϵ_n, and rearrange them into a two-suffix array ϵ_{jk}

(as with Cantor's proof of 1873 that the rationals are countable). The ϵ_{jk} are all independent, so the $X_j := \sum \epsilon_{jk}/2^k$ are independent, and $U[0,1]$ by above.

If F is a distribution function (right-continuous; increasing from 0 at $-\infty$ to 1 at ∞), define its (left-continuous) inverse function by

$$F^{-1}(t) := \inf\{F(x) \geq t\} \qquad (0 < t < 1).$$

Then if $U \sim U[0,1]$, $X := F^{-1}(U) \sim F$. For, $\{X \leq x\} = \{F^{-1}(U) \leq x\} = \{U \leq F(x)\}$, which has probability $F(x)$ as U is uniform. By this means (called the *probability integral transformation* – see the Introductory Lectures on Statistics, IntroStat under Handouts on the course website) we can pass from generating copies from the uniform distribution (say by Monte Carlo simulation) to generating copies from the distribution F. Since by above we can use *one* uniform to generate a sequence of independent copies of uniforms, we may then generate a sequence of independent copies drawn from F. In particular, from *one* uniform we can generate an *infinite sequence* of copies of *standard normals*. We shall see in Ch. III that from this we can generate *Brownian motion*, the prototypical stochastic process. So in this sense, the Lebesgue probability space, from which we can draw a uniform, is all we need – e.g. to generate Brownian motion. So everything rests on Lebesgue measure (as it should!)

Chebyshev's inequality.

The next result is due to P. L. CHEBYSHEV (1821-1984) in 1867.

Theorem (Chebychev's inequality). If X has mean μ and variance σ^2, and $\epsilon > 0$,

$$P(|X - \mu| \geq \epsilon) \leq \sigma^2/\epsilon^2.$$

Proof.

$$\sigma^2 = \int_\Omega |X - \mu|^2 dP \geq \int_{|X-\mu|^2 \geq \epsilon} |X - \mu|^2 dP \geq \epsilon^2 P(|X - \mu| \geq e). \qquad //$$