pfsl8(15).tex

Lecture 8. 2.11.2015

We form the *probability generating function (PGF)*

$$P(s), \text{ or } P_X(s), := E[s^X] = \sum_{n=0}^{\infty} s^n P(X=n) = \sum_{n=0}^{\infty} p_n s^n.$$

This is a power series in s, and since $\sum p_n = 1$, it converges for $s = 1$. So the radius of convergence R is at least 1.

If $R > 1$, $P(s)$ is analytic (= holomorphic) at $s = 1$, so we may differentiate termwise:

$$P'(s) = \sum_{n=1}^{\infty} n s^{n-1} p_n; \qquad P''(s) = \sum_{n=2}^{\infty} n(n-1) s^{n-2} p_n.$$

Taking $s = 1$:

$$P'(1) = \sum n p_n = \sum n P(X=n) = E[X];$$

$$P''(1) = \sum n(n-1) p_n = \sum n(n-1) P(X=n) = E[X(X-1)],$$

etc. (the right-hand sides are called the *factorial moments* of X; they determine the moments, and vice versa). Thus

$$E[X] = P'(1)];$$

$$var(X) = E[X^2] - (E[X])^2 = E[X(X-1)] + E[X] - (E[X])^2$$
$$= P''(1) + P'(1) - [P'(1)]^2.$$

This gives the mean and variance in terms of the first two derivatives of the PGF, in the case $R = 1$. We quote that these formulae still hold even if $R = 1$. This depends on Abel's Continuity Theorem from Analysis; we omit this.

Convolution.

Just as with MGFs and CFs: the PGF of an independent sum is the product of the PGFs.

Random sums.

If we have a random sum – a sum $X_1 + \ldots + X_N$ of a random number N of iid random variables X_i, where the X_i have PGF $P(s)$ and N is independent of the X_i with PGF $Q(s)$ – then $X_1 + \ldots + X_N$ has PGF $Q(P(s))$, the functional composition of P and Q. This result is very useful in the study of *branching processes*, which model the growth of biological populations (or chain reactions in Physics, Chemistry etc.); see Problems 7 Q4.

III. CONVERGENCE and LIMIT THEOREMS

1. Modes of convergence.

In Analysis, we deal with convergence and limits all the time, but in Probability Theory we have to modify our requirements.

Example: Coin tossing. Consider repeated (independent) tosses of a fair coin (outcomes iid Bernoulli $B(\frac{1}{2})$). What can we say about the long-run behaviour of the observed frequency to heads to date? The man/woman in the street will say, "tends to a half – Law of Averages". There is much good sense in this, and we will prove a theorem that says just this, but *subject to a qualification*, that turns out to be inevitable.

The coin can fall tails (frequency of heads 0; pr $\frac{1}{2}$). So it can fall tails 10 times (frequency of heads 0; pr 2^{-10}); 100 times (frequency 0; pr 2^{-100}), etc. Such highly exceptional behaviour is certainly very unusual (highly unlikely – and we can say exactly how unlikely). In the limit, we would expect the probability of this or any other aberrant behaviour to tend to 0, and it does. The fact remains that the limit of 0 is 0, and *not* the $\frac{1}{2}$ occurring in the Law of Averages.

Because of such examples, the best we can hope for is the following.

Definition. We say that random variables X_n, $n = 1, 2, \ldots$, *converge to X almost surely (a.s.)*, or *with probability 1 (wp1)*, if

$$P(X_n \to X \ (n \to \infty)) = 1,$$

and write

$$X_n \to X \qquad a.s.$$

This is one of our two *strong* modes of convergence. For the other:

Definition. For $p \geq 1$, X_n *converges to X in pth mean*, or *in L_p*, if

$$E[|X_n - X|^p] \to 0 \qquad (n \to \infty)$$

(L for Lebesgue, p for pth power). The two most important cases for us are $p = 1$ – convergence *in mean*, or *in L_1*, and $p = 2$ – convergence *in mean square*, or *in L_2*.

We quote (see e.g. [GS], 7.2): if $1 \leq p \leq r$,

(a) $L_r \subset L_p$ [true in any finite measure space, but not in general];
(b) convergence in rth mean implies convergence in pth mean
[as expected: the higher the moment, the more restrictive the condition].

Neither of these two strong modes of convergence implies the other.

Definition. We say that $X_n \to X$ *in probability (in pr)* if for all $\epsilon > 0$,

$$P(|X_n - X| > \epsilon) \to 0 \qquad (n \to \infty).$$

This is a mode of convergence of intermediate strength. Each of the two strong modes above implies it, but not conversely.

There is a partial converse, due to F. Riesz: if $X_n \to X$ in probability, there is a subsequence along which $X_n \to X$ a.s.

Finally, we have a (very useful but) weak mode of convergence.

Definition. $X_n \to X$ *in distribution* if

$$E[f(X_n)] \to E[f(X)]$$

for all bounded continuous functions f.

The content of Lévy's convergence theorem for CFs (II.3 above) is that such behaviour on the particular functions

$$f(x) := e^{itx}$$

for t real suffices here.

As the names suggest, the intermediate mode convergence in probability implies the weak mode convergence in distribution, but not conversely.

Metrics and completeness.

Recall that a *metric* $d = d(.,.)$ is a distance function, generalising that in Euclidean space. Some but not all metrics are generated by *norms* $\|.\|$, again as in Euclidean space:

$$d(x, y) = \|x - y\|.$$

Recall a sequence $\{s_n\}$ is called *Cauchy* if it satisfies the *Cauchy condition*

$$\forall \epsilon > 0 \qquad \exists N \text{ s.t. } \forall m, n \geq N, \ \ |s_m - s_n| < \epsilon,$$

and *convergent to s* if

$$\forall \epsilon > 0 \qquad \exists N \text{ s.t. } \forall n \geq N, \ \ |s_n - s| < \epsilon.$$

Convergence implies the Cauchy condition (by the triangle inequality). The converse is *true* for the reals $\mathbb{R}$ (Cauchy's General Principle of Convergence), and similarly for the complex plane $\mathbb{C}$, but *false* for the rationals $\mathbb{Q}$. We call a metric space *complete* if every Cauchy sequence is convergent; thus $\mathbb{R}$, $\mathbb{C}$ are complete, but $\mathbb{Q}$ not.

Convergence in pth mean (or in L_p) is metric, and generated by the L_p-*norm*:

$$\|X\|_p := (E[|X|^p])^{1/p}.$$

By the *Riesz-Fischer theorem*, the L_p-spaces are complete.

Convergence in probability is also given by a metric:

$$d(X,Y) := E\Big(\frac{|X-Y|}{1+|X-Y|}\Big).$$

This metric is also complete.

Convergence in distribution is also generated by a metric, the *Lévy metric*:

$$d(F,G) := \inf\{\epsilon > 0 : F(x-\epsilon) - \epsilon \le G(x) \le F(x+\epsilon) + \epsilon \text{ for all } x\}$$

(the French probabilist Paul LÉVY (1886-1971) in 1937) (it is not obvious, but it is true, that d is a metric): if F_n, F are distribution functions,

$$F_n \to F \quad \text{in distribution} \quad \Leftrightarrow \quad d(F_n, F) \to 0.$$

This is also equivalent to

$$F_n(x) \to F(x) \qquad (n \to \infty) \qquad \text{at all continuity points } x \text{ of } F.$$

The restriction to continuity points x of F here is vital: take X_n, X as constants c_n, c with $c_n \to c$. We should clearly have $c_n \to c$ in distribution regarded as random variables; the distribution function F of c is 0 to the left of c and 1 at c and to the right; pointwise convergence takes place everywhere *except* c (the only interesting point here).

We quote that the Lévy metric is complete.

Egorov's theorem; almost uniform convergence. We quote (D. F. EGOROV (1869-1931) in 1911)

Egorov's theorem. If $X_n \to Z$ a.s., then for all $\epsilon > 0$ there exists a set of probability $< \epsilon$ off which $X_n \to X$ uniformly (in ω). This property is called *almost uniform convergence.* So Egorov's theorem states that almost sure and almost uniform convergence are equivalent.

It follows from this that almost sure convergence ('strong') implies convergence in probability ('intermediate'), as above.