spl12.tex

Lecture 12. 5.11.2010

Then the joint distribution function is given by

$$F(x_1, \dots, x_n) = P(X_1 \le x_1, \dots, X_n \le x_n) = \prod_{i=1}^{n} P(X_i \le x_i) = \prod_{i=1}^{n} F_i(x_i),$$

where F_i is the distribution function of X_i. The F_i are called the *marginal* distribution functions:

Random variables are independent iff their joint distribution function factorizes into the product of the marginals.

Then

$$\phi(t_1, \dots, t_k) = \int \dots \int \exp\{i(t_1 x_1 + \dots + t_k x_k\} dF(x_1, \dots, x_n)$$

$$= \prod_{j=1}^{n} \int \exp\{it_j x_j\} dF_j(x_j) = \prod_{j=1}^{n} \phi_j(t_j) :$$

Random variables are independent iff their joint CF factorizes into the product of the marginals.

Convolutions.

If X, Y are independent, with distribution functions F, G and CFs ϕ, ψ, the distribution of their sum $X+Y$ is called the *convolution* (German: Faltung) of their distributions. If $X+Y$ has distribution function H and CF χ,

$$\chi(t) := Ee^{it(X+Y)} = E[e^{itX}.e^{itY}] = E[e^{itX}].E[e^{itY}] = \phi(t).\psi(t),$$

by the Multiplication Theorem:

The CF of an independent sum is the product of the CFs.

So the CF turns the easy operation of adding independent random variables into the equally easy operation of multiplying CFs. By contrast, the situation for distribution functions is less simple. If X, Y, $X+Y$ have distribution functions F, G, H,

$$H(z) := P(X+Y \le z) = \int \int_{\{x+y \le z\}} dF(x) dG(y).$$

So

$$H(z) = \int F(z-y)dG(y) = \int G(z-x)dF(x);$$

we write either expression as $(F * G)(z)$. When F, G have densities f, g, H has density

$$h(x) = \int f(x-y)g(y)dy = \int g(x-y)f(y)dy.$$

In fact, if either of F, G has a density, so does $F * G$.

So by induction, if we add n independent random variables,
(i) the CFs multiply;
(ii) the distribution is a multiple convolution, involving $n-1$ integrations.
As n increases, $n-1$ integrations become intractable, so we use CFs.

Suppose now that $X_1, \ldots, X_n, \ldots$ are independent and identically distributed (iid) random variables, with distribution F, CF ϕ, mean μ and variance σ^2. Recall that the *variance* (variability) is a measure of randomness,

$$\sigma^2 := E[(X-EX)^2] = E[X^2 - 2EX.X + (EX)^2] = E[X^2] - 2EX.EX + [EX]^2 :$$

$$var\ X = E(X^2) - (EX)^2.$$

(We know from the definition that $var\ X \geq 0$; this also follows from the last equation by the Cauchy-Schwarz inequality.)

7. The Weak Law of Large Numbers (WLLN) and the Central Limit Theorem (CLT).

Recall that by Real Analysis,

$$(1 + \frac{x}{n})^n \to e^x \qquad (n \to \infty)$$

(this expresses compound interest, or exponential growth, as the limit of simple interest as the interest is compounded more and more often). This extends also to complex number z, and to $z_n \to z$:

$$(1 + \frac{z_n}{n})^n \to e^z \qquad (n \to \infty).$$

The next result is due to Lévy in 1925, and in more general form to the Russian probabilist A. Ya. KHINCHIN (1894-1956) in 1929 and to Kolmogorov in 1928/29.

Theorem (Weak Law of Large Numbers, WLLN). If X_i are iid with mean μ,

$$\frac{1}{n}\sum_1^n X_k \to \mu \qquad (n \to \infty) \qquad \text{in probability.}$$

Proof. If the X_k have CF $\phi(t)$, then as the mean μ exists $\phi(t) = 1 + i\mu t + o(t)$ as $t \to 0$. So $(X_1 + \ldots + X_n)/n$ has CF

$$E\exp\{it(X_1 + \ldots + X_n)/n\} = [\phi(t/n)]^n = [1 + \frac{i\mu t}{n} + o(1/n)]^n,$$

for fixed t and $n \to \infty$. By above, the RHS has limit $e^{i\mu t}$ as $n \to \infty$. But $e^{i\mu t}$ is the CF of the constant μ. So by Lévy's continuity theorem,

$$(X_1 + \ldots + X_n)/n \to \mu \qquad (n \to \infty) \qquad \text{in distribution.}$$

Since the limit μ is constant, by II.4 (L11), this gives

$$(X_1 + \ldots + X_n)/n \to \mu \qquad (n \to \infty) \qquad \text{in probability. //}$$

As the name implies, the Weak LLN can be strengthened, to the Strong LLN (with a.s. convergence in place of convergence in probability). We turn to this later, but proceed with a refinement of the method above, in which we retain one more term in the Taylor expansion of the CF. Note first that the CF of the standard normal distribution $\Phi = N(0,1)$, with density $\phi(x)$ and distribution function $\Phi(x)$

$$\phi(x) := \frac{e^{-x^2/2}}{\sqrt{2\pi}}, \qquad \Phi(x) := \int_\infty^x \phi(u)du$$

is $e^{-t^2/2}$. The easiest way to show this is to show

$$\int_{-\infty}^{+\infty} e^{tx}.e^{-x^2/2}dx/\sqrt{2\pi} = e^{t^2/2}$$

by completing the square, and then replace t by it by analytic continuation to get, for real t,

$$\int_{-\infty}^{+\infty} e^{itx}.e^{-x^2/2}dx/\sqrt{2\pi} = e^{-t^2/2}$$

Or, one can use contour integration and Cauchy's theorem. For both methods, see e.g. Bingham and Fry, p. 21.

Theorem (Central Limit Theorem, CLT). If $X_1, \ldots X_n, \ldots$ are iid with mean μ and variance σ^2, and $S_n := X_1 + \ldots + X_n$, then

$$(S_n - n\mu)/(\sigma\sqrt{n}) \to \Phi = N(0,1) \qquad (n \to \infty) \qquad \text{in distribution.}$$

Proof. When we subtract μ from each X_k, we change the mean from μ to 0 and the second moment from μ_2 to the variance σ^2. So by the moments property of CFs, $X_k - \mu$ has CF $1 - \frac{1}{2}\sigma^2 t^2 + o(t^2)$ as $t \to 0$. So $X_1 + \ldots + X_n - n\mu$ has CF

$$E \exp\{it(X_1 + \ldots + X_n - n\mu)\} = [1 - \frac{1}{2}\sigma^2 t^2 + o(t^2)]^n \qquad (t \to 0).$$

Replace t by $t/(\sigma\sqrt{n})$ and let $n \to \infty$:

$$E \exp\{it(X_1 + \ldots + X_n - n\mu)/(\sigma\sqrt{n})\} = [1 - \frac{1}{2}.\frac{t^2}{n} + o(1/n)]^n \to \exp\{-t^2/2\} \quad (n \to \infty),$$

by above. The left is the CF of $(S_n - n\mu)/(\sigma\sqrt{n})$; the right is the CF of $\Phi = N(0,1)$. By the continuity theorem for CFs, this gives

$$(S_n - n\mu)/(\sigma\sqrt{n}) \to \Phi = N(0,1) \qquad (n \to \infty) \qquad \text{in distribution. //}$$

The first result of this kind is the WLLN for Bernoulli trials (tossing a coin that falls heads with probability p, tails with probability $q := 1 - p$, due to Jakob BERNOULLI (1654-1705); *Ars conjectandi*, 1713, posth.) The general WLLN above, and its strengthening the SLLN below, constitute precise forms of the 'Law of Averages', known to the man in the street. The CLT for Bernoulli trials is due to Abraham de MOIVRE (1667-1754), *Doctrine of Cfhances* 1738 (de Moivre found the normal distribution in 1733), later extended by P. S. de LAPLACE (1749-1827), *Théorie Analytiques des Probabilités*, 1812. The general CLT is due to J. W. LINDEBERG (1876-1932) in 1922 (the name 'central limit theorem' is due to Pólya, also in 1922). The CLT is the precise form of the 'Law of Errors', known to the physicist in the street as saying 'errors are normally distributed about the mean'.

Note. 1. The CLT largely explains why the normal distribution is so ubiquitous in Statistics – basically, this is why Statistics works.

2. The CLT and the normal distribution are static. We shall need their dynamic counterparts. The stochastic process (dynamic counterpart) corresponding to the normal distribution is *Brownian motion* (Ch. IV); that of the CLT is the Erdös-Kac-Donsker *invariance principle*.