pfsl12.tex

Lecture 12. 6.11.2013

Take expectations: as $Ey = \mu$, $Eg(y) \sim g(\mu)$. So

$$g(y) - g(\mu) \sim g(y) - Eg(y) \sim g'(\mu)(y - \mu).$$

Square both sides:

$$[g(y) - g(\mu)]^2 \sim [g'(\mu)]^2 (y - \mu)^2.$$

Take expectations: as $Ey = \mu$ and $Eg(y) \sim g(\mu)$, this says

$$var(g(y)) \sim [g'(\mu)]^2 var(y).$$

Regression. So if

$$E(y_i|x_i) = \mu_i, \qquad var(y_i|x_i) = \sigma_i^2,$$

we use EDA to try to find some link between the means μ_i and the variances σ_i^2. Suppose we try $\sigma_i^2 = H(\mu_i)$, or

$$\sigma^2 = H(\mu).$$

Then by above,

$$var(g(y)) \sim [g'(\mu)]^2 \sigma^2 = [g'(\mu)]^2 H(\mu).$$

We want *constant variance*, c^2 say. So we want

$$[g'(\mu)]^2 H(\mu) = c^2, \qquad g'(\mu) = \frac{c}{\sqrt{H(\mu)}}, \qquad g(y) = c \int \frac{dy}{\sqrt{H(y)}}.$$

Note. The idea of variance-stabilising transformations (like so much else in Statistics) goes back to Fisher (R. A. (Sir Ronald) FISHER (1890-1962)). He found the density of the sample correlation coefficient r^2 in the bivariate normal distribution – a complicated function involving the population correlation coefficient ρ^2, simplifying somewhat in the case $\rho = 0$ (see e.g. [KS1], §16.27, 28). But Fisher's z-transformation of 1921 ([KS1], §16.33)

$$r = \tanh z, \qquad z = \frac{1}{2}\log(\frac{1+r}{1-r}), \qquad \rho = \tanh \zeta, \qquad \zeta = \frac{1}{2}\log(\frac{1+\rho}{1-\rho})$$

gives z approximately normal, with variance almost independent of ρ:

$$z \sim N(0, 1/(n-1)).$$

4. Infinite divisibility; self-decomposability; stability: $I \supset SD \supset S$

In the CLT, the limit distribution is normal, $N(0,1)$, CF $\exp\{-\frac{1}{2}t^2\}$. But

$$\exp\{-\frac{1}{2}t^2\} = [\exp\{-\frac{1}{2}t^2/n\}]^n \qquad (n = 1, 2, \ldots)$$

expresses the CF of the limit law $N(0,1)$ as the nth power of the CF of another probability law, $N(0,1/n)$. So $N(0,1)$ is the nth convolution of $N(0,1/n)$. We think of this as 'splitting $N(0,1)$ up into n independent parts': $N(0,1)$ is n times 'divisible'. We can do this for each n, so $N(0,1)$ is 'infinitely divisible'.

Similarly for X Poisson $P(\lambda)$: the CF is

$$E[e^{itX}] = \sum_{n=0}^{\infty} e^{-\lambda}\lambda^n.e^{itn}/n! = \exp\{-\lambda(1-e^{it})\} = [\exp\{-(\lambda/n)(1-e^{it})\}]^n,$$

so $P(\lambda)$ is the n-fold convolution of $P(\lambda/n)$, for each n. So the Poisson distributions are infinitely divisible (id).

We can extend this to the *compound Poisson distribution* $CP(\lambda, F)$, which is very important in the actuarial/insurance industry. Suppose that the number of claims is Poisson $P(\lambda)$, and that the claim sizes are iid, with distribution F and CF ϕ. Then conditional on the number of claims being n, the total claimed in the nth convolution F^{*n}, and the CF is ϕ^n. So the total X claimed has CF

$$E[e^{itX}] = \sum_{n=0}^{\infty} e^{-\lambda}\lambda^n.\phi(t)^n/n! = \exp\{-\lambda(1-\phi(t))\} = [\exp\{-(\lambda/n)(1-\phi(t))\}]^n.$$

So $CP(\lambda, F)$ is the n-fold convolution of $CP(\lambda/n, F)$ for each n, so is id.
But this holds much more generally.
Definition. We say that a random variable X, or its distribution F, is *infinitely divisible* (id) if for each $n = 1, 2, \ldots$, X has the same distribution as the sum of n independent identically distributed random variables. We write I for the class of infinitely divisible distributions.

It turns out that I is also the class of limit laws of row-sums of triangular arrays, as follows. We say that $\{x_{nk}\}$ $(k = 1, \ldots, k_n,\ n = 1, 2, \ldots)$ is a

triangular array if for each n, the X_{nk} are independent;
we say that the array is *uniformly asymptotically negligible (uan*, more briefly *negligible*, if for all $\epsilon > 0$,

$$P(\max_{1 \leq k \leq k_n} |X_{nk}| > \epsilon) \to 0 \qquad (n \to \infty).$$

The following are equivalent:
(i) X is infinitely divisible, $X \in I$;
(ii) X is the limit law of the row-sums $\sum_k X_{nk}$ of some negligible triangular array.
The classic reference for this material is Gnedenko and Kolmogorov [GnK].

It turns out also that the CFs of distributions in I can be characterised explicitly: they are those of the form

$$E[e^{itX}] = \exp\{iat - \frac{1}{2}\sigma^2 t^2 + \int_{-\infty}^{\infty} \Big(e^{ixt} - 1 - ixtI_{(-1,1)}\Big) d\nu(t)\}, \qquad (LK)$$

where the (positive) measure ν, the *Lévy measure*, satisfies

$$\int \min(1, |x|^2) d\nu(x) < \infty$$

(here we omit 0 from the range of the integration – or, we can include it, perhaps at the cost of changing σ), a, the *drift*, is real, and σ, the *Gaussian component*, is ≥ 0; (a, σ, ν) is called the *characteristic triplet* of X.

Equation (LK) above is called the *Lévy-Khintchine formula* (Lévy in 1934, Khintchine[1] in 1937, following work of de Finetti in 1929 and 1930, Kolmogorov in 1932). We return to it (Ch. VI) in connection with stochastic processes – *Lévy processes*. It gives a *semi-parametric* representation – think of (a, σ) as the parametric part and ν as the non-parametric part.[2]
Note. 1. In the integrand, we need three terms near the origin, but only two terms away from the origin. As we shall see later, the Lévy measure ν governs the *jumps* of the relevant Lévy process. We distinguish between the 'big' jumps (only finitely many of these in finite time), and the 'little' jumps (there may be infinitely many in finite time!) We 'compensate' the little

[1] Khintchine as he wrote here in French; Khinchin is the usual transliteration of his name into English

[2] Here we follow the British usage of regarding a parameter as finite-dimensional. In Russian usage, the triplet would be a parametric description.

jumps by subtracting the mean – hence the $I_{(-1,1)}$. Actually, the '1' here is arbitrary: any $c \in (0, \infty)$ would do, but $c = 1$ is customary and convenient.
2. The a in the triplet corresponds to a *deterministic* part, at (t is the time), called the *drift*; the σ part corresponds to a Gaussian component (Brownian motion – see VI.1). Any of these three components may be absent.

A triangular array is a *two-suffix* entity (needing a matrix of distributions). If we specialise to the *one-suffix* case (needing a sequence of distributions), then in each row, all the X_{nk} have the same distribution. This restricts the class I of infinitely divisible distributions, and we obtain now the class SD of *self-decomposable* distributions. These have CFs of the more restricted form, where

$$\nu(dx) = k(x)dx/|x|, \quad k \quad \text{increasing on } (-\infty, 0), \text{ decreasing on } (0, \infty).$$

Again, this is a semi-parametric description.

We can specialise even further, and have an array depending on only *one* distribution, F say. We have $X_1, X_2, \ldots$ iid with law F, and form the sequence of partial sums

$$S_n := X_1 + \ldots + X_n;$$

then $S := \{S_n\}$ is called a *random walk* with *step-length* distribution F, or *generated by* F, $\{S_n\} \sim F$. Just as in the CLT, we seek to centre and scale so as to get a non-degenerate limit law. we ask for a non-degenerate limit of

$$(S_n - a_n)/b_n,$$

with a_n real, $b_n > 0$ (in the CLT $a_n = n\mu$ and $b_n = \sigma\sqrt{n}$ with μ the mean and σ^2 the variance, but here we need not have a mean or variance). So we get a *parametric* description, with four parameters – two essential, two not.
Type: location and scale.

In one dimension, the mean μ gives us a natural measure of *location* for a distribution. The variance σ^2, or standard deviation (SD) σ, give us a natural measure of *scale*.
Note. The variance has much better mathematical properties (e.g., it adds over independent, or even uncorrelated, summands). But the SD has the *dimensions* of the random variable, which is better from a physical point of view. As moving between them is mathematically trivial, we do so at will, without further comment.