pfsl5(15).tex

Lecture 5. 26.10.2015

6. *Fisher F-distribution*, $F(m,n)$. This is defined as that of the ratio

$$F := \frac{U/m}{V/n},$$

where $U \sim \chi^2(m)$, $V \sim \chi^2(n)$ and U, V are independent. This has density

$$f(x) = \frac{m^{\frac{1}{2}m} n^{\frac{1}{2}n}}{B(\frac{1}{2}m, \frac{1}{2}n)} . \frac{x^{\frac{1}{2}(m-2)}}{(mx+n)^{\frac{1}{2}(m+n)}} \qquad (m, n > 0, \quad x > 0).$$

See e.g. [BF], 3.3, and for its uses in, e.g., hypothesis testing, [BF] Ch. 6.

7. *Gamma distributions*, $\Gamma(\alpha)$. These are defined, for $\alpha > 0$, as having density

$$f(x) := x^{\alpha-1} e^{-x} / \Gamma(\alpha) \qquad (x > 0).$$

They are the 'default option' for modelling non-negative errors – just as the Normal is the 'default option' for modelling errors with either sign (in *Generalized Linear Models (GLMs)*; see e.g. [BF], Ch. 8.

8. *Beta distributions* $B(\beta, \alpha)$. These are defined, for $\alpha, \beta > 0$, as having density

$$f(x) := \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \qquad 0 \leq x \leq 1.$$

They are often used for modelling *proportions*.

Finally, we mention some standard discrete distributions.

9. *Bernoulli* $B(p)$. This models a biased coin, that falls heads (or success, 1) with probability $p \in (0,1)$, tails (or failure, or 0) with probability $qx := 1-p$ (a "p-coin"):

$$f(x) = p^x (1-p)^{1-x}, \qquad x = 0, 1.$$

This has mean p and variance pq.

10. *Binomial* $B(n,p)$, $p \in (0,1)$, $n = 1, 2, \ldots$. This models the number of times a p-coin falls heads in n independent tosses:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n.$$

It has mean np and variance npq.

11. *Poisson* $P(\lambda)$, $\lambda > 0$. Here

$$f(x) = e^{-\lambda} \lambda^x / x!, \qquad x = 0, 1, 2, \ldots$$

This has mean λ and variance λ. It models *counts*, of such things as accidents, insurance claims, earthquakes etc. We shall meet it later in connection with the *Poisson process*.

2. Convolutions

In the density case,

$$P(X \in A) = \int_A f(x)dx.$$

Here we may be in any number of dimensions n: if $n > 1$, X is a random vector, x is a vector, $\int$ is an n-fold integral, and its element of integration dx is also n-dimensional.

Recall the definition of independence: events $A_1, \dots, A_n$ are *independent* if the probability of the intersection of any subset of them is the product of their separate probabilities. Random variables $X_1, \dots, X_n$ are *independent* iff the events $X_1 \leq x_1, \dots, X_n \leq x_n$ are independent, for all $x_1, \dots, x_n$. So independence holds iff

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

i.e.

$$F(x_1, \dots, x_n) = \prod_{1}^{n} F_i(x_i),$$

in which case the copula is

$$C(u_1, \dots, u_n) \equiv u_1 u_2 \dots u_n \qquad (u_i \in [0,1]).$$

That is, independence holds iff the joint distribution *factorizes* into the product of the marginals.

In the density case: the LHS above is

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(u_1, \dots, u_n) du_1 \dots du_n,$$

and similarly for the RHS. As the two integrals are equal, the two integrands are equal a.e., giving: independence holds iff

$$f(x_1, \dots, x_n) = \prod_{1}^{n} f_i(x_i), \qquad a.e.:$$

the joint density *factorises* into the product of the marginal densities.

When $n = 2$, we write the random vector as (X, Y), and write f, g for the densities of X, Y. In the independent case, the joint density of (X, Y) is thus $f(x)g(y)$.

In the density case, *one evaluates probabilities by integrating the relevant density over the relevant region.* So if $Z := X + Y$ has density h and distribution function H,

$$H(z) := P(Z \leq z) = P(X + Y \leq z) = \int_{-\infty}^{z} h(u)du$$

$$= \int\int_{\{(x,y):x+y\leq z\}} f(x)g(y)dxdy = \int_{-\infty}^{\infty} f(x)dx \int_{-\infty}^{z-x} g(y)dy.$$

(Here in the last integral we have the usual choices of notation in calculus of several variables: $\int f(x)dx \int \ldots$ as here, $\int f(x)[\int \ldots]dx$, or – arguably the most logical – $\int dx f(x) \int dy \ldots$. We shall use whatever seems most convenient – the meaning is clear.)

We now want to extract the density h. At least a.e. in z, we can get this by differentiating H. As the z-dependence on the RHS is under the integral sign, we want to *differentiate under the integral sign* on the right. This can be justified (in this case, because $g \geq 0$, so the inner integral is increasing in z). We obtain

$$h(z) = \int_{-\infty}^{\infty} f(x)g(z - x)dx.$$

Changing variables and using symmetry:

$$h(x) = \int_{-\infty}^{\infty} f(y)g(x - y)dy = \int_{-\infty}^{\infty} g(y)f(x - y)dy.$$

We call this the *convolution* h of f and g, and write

$$h = f * g \ (= g * f).$$

Note. 1. When $X, Y \geq 0$, f, vanish on $(-\infty, 0)$, and the above reduces to

$$(f * g)(x) = \int_{0}^{x} f(y)g(x - y)dy.$$

2. The distribution-function version of convolution is

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

3. Think of convolution as a *smoothing* process (it involves an integration, and integration is a smoothing process). We quote: we do not need *both* X and Y to have a density for $X+Y$ to have a density – it is enough if *one* of them does.
4. In Statistics, we are constantly *averaging*, over the n readings in our sample; n may be large – the larger, the better. Averaging entails two steps: adding over the readings, and dividing by the number of readings. The second is trivial, so this reduces effectively to the first. If (as is usual) the readings are independent, this means we are forming an n-fold convolution. This involves $n-1$ integrations. This becomes impossibly unwieldy for large n! So, we need an alternative – to which we turn in II.3c below.
Example: Gamma distributions and the Beta integral. If $X \sim \Gamma(\alpha)$, $Y \sim \Gamma(\beta)$ and X, Y are independent, we have

$$f(x) = e^{-x}x^{\alpha-1}/\Gamma(\alpha), \qquad g(y) = e^{-y}y^{\beta-1}/\Gamma(\beta), \qquad x, y > 0.$$

So

$$\begin{aligned} (f*g)(x) &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x e^{-y}e^{-(x-y)}y^{\alpha-1}(x-y)^{\beta-1}dy \\ &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} e^{-x} \int_0^x y^{\alpha-1}(x-y)^{\beta-1}dy. \end{aligned}$$

Put $y := xu$

$$h(x) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)} e^{-x}x^{\alpha+\beta-1} \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du.$$

We know that h is a density (of $X+Y$). But on the RHS, we have the functional form of a density – the Gamma density $\Gamma(\alpha+\beta)$. As always with densities, it suffices to work modulo a multiplicative constant: if we get the variable part right, we can always pick up the constant at the end – it is the value needed to make the density integrate to 1. So the constant on the RHS – $\Gamma(\alpha)\Gamma(\beta)\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du$ – is in fact $\Gamma(\alpha+\beta)$. Recalling the definition of the Beta integral, this proves

$$B(\alpha, \beta) := \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

– Euler's integral for the Beta function, of II.1 above! Note how remarkable this is: we have obtained a purely analytic result by a purely probabilistic argument.