

Lecture 4. 20.10.2015 (no lectures Mon 19: Commemoration Day)

Higher dimensions; joint and marginal distributions

If $X = (X_1, \dots, X_n)$ is a random variable taking values in n -dimensional space – a random n -vector – then its distribution function F is defined as above, but coordinatewise. If $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$, we write

$$x \leq y \text{ iff } x_1 \leq y_1, \dots, x_n \leq y_n.$$

Then

$$F(x) := P(X \leq x) = P(X_1 \leq x_1, \dots, X_n \leq x_n).$$

This is also called the *joint* distribution of (X_1, \dots, X_n) , while

$$F_i(x_i) := P(X_i \leq x_i), \quad i = 1, \dots, n$$

is called the *marginal* distribution of X_i . Note that letting the j th argument $x_j \rightarrow \infty$ eliminates the condition $X_j \leq x_j$, and so leaves the joint distribution of the X s with X_j omitted. So the joint distribution of a random vector determines the joint distribution of any subvector, and the marginals of its coordinates, just by letting unwanted arguments go to $+\infty$. In sum: *the joint determines the marginals*.

Probability Integral Transformation (PIT).

As F is non-decreasing, it has an inverse function. We use

$$F^{-1}(x) := \inf\{x : F(x) \geq t\} = \min\{x : F(x) \geq t\}$$

(also non-decreasing, but left-continuous – so the infimum is attained, i.e. is a minimum). Write $X \sim F$ to mean that the random variable X has distribution F . Then if $U[0, 1]$ is the uniform distribution above (probability = length) and $U \sim U[0, 1]$, then U is uniformly distributed on $[0, 1]$; we shall use this standard notation below. The *Probability Integral Transformation (PIT)* uses U and F to generate X :

$$X := F^{-1}(U) \sim F. \quad (PIT).$$

Proof. $P(X = F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$. // The PIT is very useful in the context of Simulation (using computers to generate random numbers); see IS, I and p.2. It means that we only need random number tables for the uniform distribution $U[0, 1]$, and can then use (PIT)

to transform this data to have distribution F .

Copulas

The question arises of how to go in the reverse direction. It is helpful to think of the information in the joint distribution as composed of two parts: one on the marginals, the other on the *dependence* between the coordinates – often of great statistical importance! One needs a function that *couples* the marginals together to form the joint. This is called the *copula*.

A *copula* C in n dimensions is a probability distribution function on (= supported by – all its probability mass is on) the unit n -cube $[0, 1]^n$.

Sklar's Theorem (A. SKLAR, 1958). If $F(x)$ is a joint distribution in n dimensions, with marginals $F_i(x_i)$, there exists an n -dimensional copula C with

$$F(x) = F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)).$$

Conversely, given any copula C and marginals F_i , this formula gives a joint distribution F with marginals F_i . The correspondence between F and C is unique if the marginals F_i are continuous.

Absolute continuity and the Radon-Nikodym theorem

In the density case,

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u) du.$$

In the discrete case,

$$F(x) = P(X \leq x) = \sum_{n: x_n \leq x} f(x_n).$$

Each expresses a relationship between measures. In the density case, the measures are F and λ , Lebesgue measure:

$$\lambda(B) = 0 \quad \Rightarrow \quad F(B) = 0.$$

In the discrete case, the measures are F and counting measure on the set of values $\{n : x_n\}$ (think of $x_n = n$, say). In general: if P, Q are measures, we say Q is *absolutely continuous* w.r.t. P , written $Q \ll P$, if

$$P(A) = 0 \quad \Rightarrow \quad Q(A) = 0.$$

Then the *Radon-Nikodym theorem* states that $Q \ll P$ iff

$$Q(A) = \int_A f dP$$

for some measurable function f , called the *Radon-Nikodym (RN) derivative* of Q w.r.t. P , written $f = dQ/dP$. Thus each of the f s above is a RN derivative. See e.g. SP L7, or [S] Ch. 19.

II. DISTRIBUTIONS AND THEIR TRANSFORMS

1. Examples.

1. *Uniform* $U[a, b]$. This has density

$$f(x) = 1/(b-a) \quad (a \leq x \leq b), \quad 0 \quad \text{otherwise}$$

and distribution

$$F(x) = 0 \quad (x \leq a), \quad (x-a)/(b-a) \quad (a \leq x \leq b), \quad 1 \quad (x \geq b).$$

The case $U[0, 1]$ is basic – we have met this in I, and seen how to get any other distribution from it by the Probability Integral Transformation.

$U[a, b]$ forms a two-parameter family. It is statistically interesting, as maximum-likelihood estimation (MLE) of its parameters is *non-regular*: instead of getting a normal limit and convergence rate \sqrt{n} as usual, we get a symmetric exponential limit and convergence rate n ; see e.g. IS II. This is typical of situations, as here, where the *support* (smallest set carrying full probability, 1) depends on the parameters.

2. *Exponential* $E(\lambda)$, $\lambda > 0$. This has density

$$f(x) = \lambda e^{-\lambda x} \quad (x \geq 0), \quad 0 \quad (x < 0)$$

and distribution

$$F(x) = 1 - e^{-\lambda x} \quad (x \geq 0), \quad 0 \quad (x \leq 0).$$

Here the mean is $E[X] = 1/\lambda$. MLE is regular, and the MLE $\hat{\lambda} = 1/\bar{x}$, as one would expect (sample mean \bar{x} corresponds to population mean $1/\lambda$).

3. *Normal* $N(\mu, \sigma^2)$; μ real, $\sigma > 0$. Here the density is

$$f(x) := \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-\mu)^2/\sigma^2\right\}.$$

This is a density, and (as the notation suggests) it does indeed have mean μ and variance σ^2 [II.3 Example 1a, L7].

The case $\mu = 0, \sigma = 1$, the *standard normal* distribution $N(0, 1)$, is so

important it has special notation: the density and distribution function are written

$$\phi(x) := \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}x^2\}, \quad \Phi(x) = \int_{-\infty}^x \phi(u)du.$$

Note that $\Phi(0) = \frac{1}{2}$ by symmetry (and $\Phi(-\infty) = 0$, $\Phi(\infty) = 1$); for other values, we have to use tables.

The MLEs for the population mean and variance are the sample mean and variance:

$$\hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = \bar{S}^2 \quad (:= \frac{1}{n} \sum_1^n (X_k - \bar{X})^2).$$

Note that we use the "1/n" definition of the sample variance (so that "bar, or average, corresponds to expectation", rather than the alternative "1/(n-1)" definition (to get the sample variance unbiased). Always check!

4. *Chi-square with n degrees of freedom (df), $\chi^2(n)$* . This is the distribution of $X_1^2 + \dots + X_n^2$, where X_1, \dots, X_n are independent and identically distributed (iid) $N(0, 1)$. It has density

$$f(x) = \frac{1}{2^{\frac{1}{2}n} \Gamma(\frac{1}{2}n)} x^{\frac{1}{2}n-1} \exp\{-\frac{1}{2}x\} \quad (x > 0),$$

mean n and variance $2n$; see e.g. [BF], §2.1.

We quote (see e.g. [BF], Th. 2.4):

(i) \bar{X} and S^2 are independent; (ii) $\bar{X} \sim N(\mu, \sigma^2/n)$; (iii) $nS^2/\sigma^2 \sim \chi^2(n-1)$.
 5. *Student t-distribution with n df, $t(n)$* . This is defined as the distribution of

$$X := \sqrt{n}U/\sqrt{V},$$

where $U \sim N(0, 1)$, $V \sim \chi^2(n)$ and U, V are independent. It has distribution

$$f(x) = \frac{\Gamma(\frac{1}{2}n + \frac{1}{2})}{\sqrt{\pi n} \Gamma(\frac{1}{2}n)} \left(1 + \frac{x^2}{n}\right)^{-\frac{1}{2}(n+1)}.$$

By above,

$$\sqrt{n-1}(\bar{X} - \mu)/S \sim t(n-1).$$

This is very useful when estimating the mean μ without knowing the variance σ^2 (or standard deviation – SD – σ): the *nuisance parameter* σ cancels on forming the Student t ratio above.