**Lecture 3**. 11.10.2012 (half-hour:problems)

## 3. Distributions and distribution functions

The distribution function $F(x) := P(X \leq x)$ of $X$ is a Lebesgue-Stieltjes measure function; it determines the corresponding Lebesgue-Stieltjes measure by (denoting this also by $F$ to save letters – $\mu_F$ is the other common notation)

$$F((a, b]) = F(b) - F(a)$$

(and hence we can extend from such intervals to general Borel sets). Now

$$F(x) = P(X \leq x) = P(X \in (-\infty, x]) = P(X^{-1}(-\infty, x]),$$

or (taking $a = -\infty, b = x$ above)

$$F((-\infty, x]) = P(X^{-1}(-\infty, x]).$$

Extending as above,

$$F(B) = P(X^{-1}(B))$$

for any Borel set $B$. We may write the RHS as the composite $(P \circ X)(B)$. We thus then have

$$F = P \circ X^{-1} :$$

$F$, the distribution of $X$, is the *image measure* of the probability measure $P$ under the inverse map $X^{-1}$ (or more briefly, 'under $X$').

*Expectations*

In Lecture 2, we defined $E[X]$ as $\int_\Omega X dP$, and similarly $E[g(X)] = \int_\Omega g(X) dP$, for Borel measurable $g$.

In your first course on Probability and/or Statistics, you defined

$$E[g(X)] := \int_{-\infty}^{\infty} g(x) dF(x),$$

at least in the two main cases:

$$\int_{-\infty}^{\infty} g(x) f(x) dx \quad \text{(density case, density } f); \quad \sum_n g(x_n) f(x_n) \quad \text{(discrete case)}$$

(we now know that there is no need to handle these separately, we can handle them together – and, that we must restrict to the case of *absolute* convergence

1

in all sums and integrals).

It seems that we now have two different ways of defining $E[g(X)]$ – as a $P$-integral over the sample space $\Omega$ or as an $F$-integral over the line. As one might expect, these two are the same. This follows from the transformation formula for integrals in Measure Theory; see SP L7.

*The discrete case.*

If $X$ takes (finitely or) countably many values $x_n$, write

$$f(x_n) := P(X = x_n) \qquad (n = 1, 2, \ldots).$$

Then the distribution function

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

is a jump-function, increasing by $f(x_n)$ at $x_n$ and constant elsewhere.

*The Lebesgue decomposition.*

The discrete and density cases are not exhaustive – though they are all one usually encounters in practice in Statistics. We quote: the general distribution function $F$ has a *Lebesgue decomposition*

$$F = c_{ac}F_{ac} + c_dF_d + c_sF_s,$$

where the constants $c_i$ are non-negative and sum to 1 (the RHS is called a *mixture*), $F_{ac}$ is an *absolutely continuous* distribution, $F_d$ is a *discrete* distribution (with density $f$, say), and $F_s$ is a *continuous singular* distribution (no jumps, but increases only on a Lebesgue-null set). We will not encounter such $F_s$ in practice, so we do not discuss them further.

This reduces the number of components on the right to two. Actually, we will only encounter one at a time here – usually the density case (see below)

*Discrete v. continuous.*

Statistics is dominated by the density case: normal, chi-square, Student $t$, Fisher $F$, uniform, exponential, Gamma, Beta etc. But the discrete case also occurs – e.g., the *Poisson* distribution. The density case corresponds to *measurement* data, the discrete case to *count* data. Mathematically, the density case involves *integrals*, the discrete case *sums*. We have chosen our notation $f(.)$ to fit both cases. This is more than a formal analogy: distributions with densities are absolutely continuous w.r.t. *Lebesgue measure*; discrete ones are absolutely continuous w.r.t. *counting measure* (SP L4).