SMSTC (2007/08)

Probability

www.smstc.ac.uk

Contents

SMSTC (2007/08)
Probability
Lecture 4: Joint distributions and independence

[a]

www.smstc.ac.uk

Contents

[a]

4.1 Joint distributions

Consider a random element $(X, Y) : (\Omega, \mathscr{F}) \to (\mathbb{R}^2, \mathscr{B}(\mathbb{R}^2))$. This random variable is called TWO RANDOM VARIABLES.

> If $\mathbf{P}$ is a probability of $(\Omega, \mathscr{F})$, the joint distribution of (X, Y) is another name for the law $\mathbf{P}_{X,Y}$ of the random variable (X, Y).

The joint distribution function of (X, Y) is the function

$$F_{X,Y}(x, y) := \mathbf{P}(X \le x, Y \le y) = \mathbf{P}_{X,Y}((-\infty, x] \times (-\infty, y]), \quad (x, y) \in \mathbb{R}^2. \tag{4.1}$$

The law $\mathbf{P}_X$ of X is referred to as the first marginal of the law $\mathbf{P}_{X,Y}$. The distribution function F_X of X is referred to as the first marginal distribution function of the joint distribution function $F_{X,Y}$ and, of course,

$$F_X(x) = \lim_{y\to\infty} F_{X,Y}(x, y).$$

Note that we chose ti use $\le$ instead of $<$ in (4.1) for no good reason other than a mere arbitrary convention.

[a]These notes contain almost no proofs. For a complete set of notes, see [3]; alternatively, read the introductory book [4], or the more advanced book [1].

4.1.1 Knowledge of $F_{X,Y}$ implies knowledge of $\mathbf{P}_{X,Y}$

Explaining why knowledge of the function $F_{X,Y}$ implies knowledge of $\mathbf{P}_{X,Y}(B)$ for all $B \in \mathscr{B}(\mathbb{R}^2)$ is beyond the scope of these lectures. The explanation can be found in [1, 3, 4]. But we give some intuition. Consider a rectangle (with sides parallel to the axes–please think geometrically)

$$(a_1, b_1] \times (a_2, b_2] := \{(x, y) \in \mathbb{R}^2 : \; a_1 < x \le b_1, \; a_2 < y \le b_2\}. \tag{4.2}$$

We allow a_1, a_2 to take any value, including $-\infty$. Since $(-\infty, b_1] \times (-\infty, b_2]$ is the disjoint union of four rectangles, using additivity, we obtain

$$\mathbf{P}_{X,Y}((a_1, b_1] \times (a_2, b_2]) = F_{X,Y}(b_1, b_2) - F_{X,Y}(b_1, a_2) - F_{X,Y}(a_1, b_2) + F_{X,Y}(b_1, b_2).$$

Rectangles of the form (4.2) have the following nice properties: First, intersection of two of them is a rectangle of the same form. Thus, if $\mathscr{R}$ denotes the collection of these rectangles we have that $\mathscr{R}$ is closed under finite intersection, i.e. it is a π-system. Second, the complement of a rectangle from $\mathscr{R}$ is a finite union of disjoint rectangles from $\mathscr{R}$.

EXERCISE 1. Write, explicitly, the complement of $(a_1, b_1] \times (a_2, b_2]$ as the disjoint union of elements of $\mathscr{R}$. Notice we can do that in at least two different ways. Do so.

Now consider the class

$$\mathscr{C} := \{\text{finite disjoint unions of elements of } \mathscr{R}\}.$$

It is easy to visualise, geometrically, what kind of elements $\mathscr{C}$ contains. Then

EXERCISE 2. Show that $\mathscr{C}$ is a field, i.e. if $A \in \mathscr{C}$ then $A^c \in \mathscr{C}$ and if $A, B \in \mathscr{C}$ then $A \cup B \in \mathscr{C}$.

Hence, if $A \in \mathscr{C}$, we can write $A = \cup_{i=1}^m B_i$, where B_i are disjoint rectangles from $\mathscr{R}$, and, since for each such rectangle B_i we can use $F_{X,Y}$ to compute $\mathbf{P}_{X,Y}(B)$, we have that

$$\mathbf{P}_{X,Y}(A) = \sum_{i=1}^m \mathbf{P}_{X,Y}(B), \quad A \in \mathscr{C},$$

can be computed by using $F_{X,Y}$ only.
But there are many sets in $\mathscr{B}(\mathbb{R}^2)$ that do not belong to $\mathscr{C}$, so we wish to continue our endeavour. If it were true that every set in $\mathscr{B}(\mathbb{R}^2)$ was a limit of sets of $\mathscr{C}$ then we would be finished, by the continuity property of a probability. However, there are many elements in $\mathscr{B}(\mathbb{R}^2)$ that are not limits of sets in $\mathscr{C}$.

Example 4.1. The set of all points (x, y) where x, y are rationals cannot be obtained as a limit of elements of $\mathscr{C}^2$. To illuminate this point, consider "straightforward" procedure that places a little rectangle around each such point and then let the little rectangle shrink. Specifically, let $\mathbb{Q} = \{q_1, q_2, \ldots\}$ be an enumeration of the rationals. To each (q_m, q_n) associate the rectangle

$$I_{m,n}(\varepsilon) := (q_m - \varepsilon 2^{-m}, \; q_m + \varepsilon 2^{-m}] \times (q_n - \varepsilon 2^{-n}, \; q_n + \varepsilon 2^{-n}],$$

and let $I(\varepsilon) := \cup_{m,n} I_{m,n}(\varepsilon)$. Show that $\cap_{\varepsilon > 0, \varepsilon \in \mathbb{Q}} I(\varepsilon)$ is not equal to $\mathbb{Q} \times \mathbb{Q}$.
Hint: The set $I(\varepsilon)$ is uncountable.

We will denote by $\mathbf{P}_0$ the function $\mathbf{P}_{X,Y}$ restricted to $\mathscr{C}$. Clearly, $\mathbf{P}_0$ is uniquely specified by $F_{X,Y}$. If $\mathbf{P}_0$ can be uniquely extended to a probability on $(\mathscr{R}^2, \mathscr{B}(\mathbb{R}^2))$ then our claim that $F_{X,Y}$ completely specifies $\mathbf{P}_{X,Y}$ will be proved. The proof of this is contained in [3]. The most important step in the proof is to show that $\mathbf{P}_0$ is countably additive on $\mathscr{C}$.

EXERCISE 3. Let V, W be independent random variables with $F_V(t) = F_W(t) = 1 - e^{-t}$, $t > 0$. Let $X = 2V$, $Y = V - W$. Compute the joint distribution function $F_{X,Y}$ of the random element (X, Y).

4.2 Independence

Recall that X, Y are independent random variables on $(\Omega, \mathscr{F}\mathbf{P})$ if $\sigma(X)$, $\sigma(Y)$ are independent σ-fields. We would like to show that

Proposition 4.1. *X, Y are independent on $(\Omega, \mathscr{F}, \mathbf{P})$ if and only if $F_{X,Y}(x,y) := \mathbf{P}(X \leq x, Y \leq Y)$, $F_X(x) := \mathbf{P}(X \leq x)$, $F_Y(y) := \mathbf{P}(Y \leq y)$ are related by*

$$F_{X,Y}(x,y) = F_X(x)F_Y(y).$$

This is a consequence of the following:

Lemma 4.1. *Let $(\Omega, \mathscr{F}, \mathbf{P})$ be a probability space. Let $\mathscr{R}_1, \mathscr{R}_2$ be two classes of sets, each of which is closed under the intersection operation. Then $\sigma(\mathscr{R}_1), \sigma(\mathscr{R}_2)$ are independent if and only if*

$$\mathbf{P}(B_1B_2) = \mathbf{P}(B_1)\mathbf{P}(B_2), \quad B_1 \in \mathscr{R}_1,\ B_2 \in \mathscr{R}_2.$$

For a proof see [3].

Proof of Proposition 4.1: Let $\mathscr{R}_1 = \{X^{-1}(a,b] : \ a < b\}$, $\mathscr{R}_2 = \{Y^{-1}(a,b] : \ a < b\}$. These are both classes of sets closed under intersection. It is easy to see that the assumption $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ implies that $\mathbf{P}(B_1B_2) = \mathbf{P}(B_1)\mathbf{P}(B_2)$, $B_1 \in \mathscr{R}_1$, $B_2 \in \mathscr{R}_2$. Hence $\mathbf{P}(A_1A_2) = \mathbf{P}(A_1)\mathbf{P}(A_2)$, $A_1 \in \sigma(\mathscr{R}_1) = \sigma(X)$, $A_2 \in \sigma(\mathscr{R}_2) = \sigma(Y)$. $\square$

4.3 Joint density

The concept of joint density requires understanding of the Lebesgue integral of functions of two variables. You will not be asked to understand this, but you may think of it, roughly, as a Riemann integral. For a complete understanding, see [1,4], and also consult the notes [3].

We do mention, without proof of course, the following:

Theorem 4.1. *(i) If h is Riemann integrable on a rectangle $R = [a,b] \times [c,d]$ then its Lebesgue integral on R coincides with its Riemann integral on R.*
(ii) If h is bounded and measurable then it is Riemann integrable on R of the set of discontinuities D of h satisfies $\mathbf{P}((U_1, U_2) \in D) = 0$.

Given two distribution functions F_1, F_2 on $\mathbb{R}$, we can define the product measure $F_1 \times F_2$ on $(\mathbb{R}^2, \mathscr{B}(\mathbb{R}^2))$ by defining it first on rectangles,

$$(F_1 \times F_2)(B_1 \times B_2) := \mathbf{P}(F_1^{-1}(U_1) \in B_1,\ F_2^{-1}(U_2) \in B_2),$$

and then extending it to $\mathscr{B}(\mathbb{R}^2)$ using the procedure explained in [3]. One could define the Lebesgue-Stieltjes integral by

$$\int_{\mathbb{R}^2} h\ d(F_1 \times F_2) := \mathbf{E}h(F_1^{-1}(U_1), F_2^{-1}(U_2)),$$

where U_1, U_2 are independent random variables, uniform in $[0,1]$.

Theorem 4.2 (Fubini)**.** *If $h \geq 0$ or if $\int_{\mathbb{R}^2} |h|\ d(F_1 \times F_2) < \infty$, then*

$$\int_{\mathbb{R}^2} h\ d(F_1 \times F_2) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x,y) F_2(dy) \right) F_1(dx) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x,y) F_1(dx) \right) F_2(dy).$$

Proof [sketch]: The statement is obvious when $h(x,y) = \mathbf{1}_{B_1}(x)\mathbf{1}_{B_2}(y)$. Hence it is obvious for linear combinations of such functions. The general statement follows from an approximation procedure (see [1,4] and the notes [3].) $\square$

Occasionally, it so happens that there exists a function f such that $F_{X,Y}$ can be written as a Lebesgue integral:

$$F_{X,Y}(x,y) = \int_{(-\infty,x]\times(-\infty,y]} f(s,t)dsdt.$$

In such a case, we say that (X,Y) is absolutely continuous (jointly absolutely continuous, I suppose, if you want to be pedantic) and that $f_{X,Y}$ is a density. Using Fubini's theorem, we see that

$$f_X(x) = \int_{(-\infty,x]} f(s)ds$$

is a density for X (i.e. X is absolutely continuous, and so is Y).

Another consequence of Fubini's theorem is:

Lemma 4.2. *If X, Y are independent then*

$$\mathbf{E}(XY) = (\mathbf{E}X)(\mathbf{E}Y),$$

whenever the expectations are defined.

And another, useful consequence of Fubini's theorem is:

Lemma 4.3. *If X is a positive random variable then*

$$\mathbf{E}X = \int_0^\infty \mathbf{P}(X > x)\ dx.$$

A standard criterion for independence between X, Y, for absolutely continuous (X,Y), is:

Lemma 4.4. *Suppose that (X,Y) is absolutely continuous. Let $f_{X,Y}$ be a density of (X,Y). Let f_X, f_Y be densities of X, Y, respectively. Then X, Y are independent if and only if*

$$f(x,y) = f_X(x)f_Y(y),$$

for all (x,y) except, possibly, on a set of measure zero.

4.4 Joint moment generating function

When (X,Y) is a random variable in $\mathbb{R}^2$ we can define its JOINT MOMENT GENERATING FUNCTION by

$$M_{X,Y}(\eta,\theta) := \mathbf{E}e^{\eta X + \theta Y}, \quad \eta, \theta \in \mathbb{R}.$$

whenever it exists. Let also M_X, M_Y be the moment generating functions of X, Y. One can prove that:

Lemma 4.5. *Suppose that $M_{X,Y}$ exists in a (one sided) neighbourhood of zero. If $M_{X,Y}(\eta,\theta) = M_X(\eta)M_Y(\theta)$ then X, Y are independent.*

EXERCISE 4. Let X, Y be independent random variables with common moment generating function $M_X(\theta) = M_Y(\theta) = e^{\theta^2}$. Let $X' := X + Y$, $Y' := X - Y$. Compute the joint moment generating function of (X', Y'). Show that X', Y' are independent.

4.5 Correlations

We now consider random variables X on some probability space $(\Omega, \mathscr{F}, \mathbf{P})$ with $\mathbf{E}X^2 < \infty$. The aggregate of all these random variables will be denoted by

$$L^2(\Omega, \mathscr{F}, \mathbf{P}).$$

If $\mathbf{E}X^2 < \infty$, $\mathbf{E}Y^2 < \infty$, then $\mathbf{E}(X+Y)^2 < \infty$ (indeed, $(x+y)^2 \le 2x^2 + 2y^2$). This means that if $X, Y \in L^2(\Omega, \mathscr{F}, \mathbf{P})$ then, for any $a, b \in \mathbb{R}$, $aX + bY \in L^2(\Omega, \mathscr{F}, \mathbf{P})$. Hence $L^2(\Omega, \mathscr{F}, \mathbf{P})$ is a linear space (a linear subspace of $\mathbb{R}^\Omega$). The correlation between X and Y is defined by

$$\mathbf{E}(XY).$$

The covariance between X and Y is defined by

$$\operatorname{cov}(X, Y) = \mathbf{E}(X - \mathbf{E}X)(Y - \mathbf{E}Y).$$

Since, for $X_1, X_2, Y \in L^2(\Omega, \mathscr{F}, \mathbf{P})$, $a_1, a_2 \in \mathbb{R}$,

$$\operatorname{cov}(a_1X_1 + a_2X_2, Y) = a_1 \operatorname{cov}(X_1, Y) + a_2 \operatorname{cov}(X_2, Y),$$

the covariance is linear in each of its arguments when the other is kept fixed and it can thus be used to define an inner product:

$$\langle X, Y \rangle := \operatorname{cov}(X, Y).$$

We also define the semi-norm

$$||X|| := \sqrt{\operatorname{cov}(X, X)} = \sqrt{\operatorname{var}(X, X)},$$

where the word 'semi-norm' means that it has the following properties:

1. $||X|| \ge 0$
2. $||a_1X_1 + a_2X_2|| = |a_1|\ ||X_1|| + |a_2|\ ||X_2||$.
3. $||X + Y|| \le ||X|| + ||Y||$.

Another name for $||X||$ is 'standard deviation'. We also note that

If $||X|| = 0$ then $\mathbf{P}(X = 0) = 1$.

We <u>cannot</u> deduce, from $||X|| = 0$ alone that $X(\omega) = 0$ for all $\omega \in \Omega$, but only that $X(\omega) = 0$ for all ω except those in a set of probability zero. If we could deduce that $X(\omega) = 0$ for all $\omega \in \Omega$, we would say that $|| \cdot ||$ is a norm. To get around this problem, we merely identify all random variables in $L^2(\Omega, \mathscr{F}, \mathbf{P})$ which differ on a set of measure zero: That is, we let $[X]$ be the set of all Y such that $\mathbf{P}(X \ne Y) = 0$, and redefine $L^2(\Omega, \mathscr{F}, \mathbf{P})$ to be the collection of all such $[X]$. It is not hard to see that this is still a linear space and if we let $||[X]|| := ||X||$ (which is well defined), then this is a norm.

Being a normed space with an inner product, $L^2(\Omega, \mathscr{F}, \mathbf{P})$ has a structure much like the geometric structure of the usual Euclidean space, for instance, Pythagoras' theorem holds:

$$||X + Y||^2 = ||X||^2 + ||Y||^2 \quad \text{if } \langle X, Y \rangle = 0.$$

For more properties, see [3].

EXERCISE 5. Let U_1, U_2 be independent, both uniform in $[0, 1]$. Compute the correlation and covariance between $X_1 := U_1 + e^{U_2}$, $X_2 = U_1^2 - U_2^3$.

4.6 Conditioning

4.6.1 Naïve conditioning

Recall that, if $(\Omega, \mathscr{F}, \mathbf{P})$ is a probability space and $B \in \mathscr{F}$, $\mathbf{P}(B) \neq 0$, we define

$$\mathbf{P}(A|B) = \mathbf{P}(AB)/\mathbf{P}(B), \quad A \in \mathscr{F}.$$

This is a new probability, called $\mathbf{P}$ conditional on B. We easily see that $(\Omega, \mathscr{F}, \mathbf{P}(\cdot|B))$ is a new probability space.

The problem is that, in many cases, we want to condition with respect to an event B that has probability zero. This can be done. Loosely speaking, what saves us is the fact that if $\mathbf{P}(B) = 0$ then $\mathbf{P}(AB) = 0$ as well.

We can use many methods for defining conditioning in a more general sense. There is a geometric approach based on the structure of $L^2(\Omega, \mathscr{F}, \mathbf{P})$, and this can be found in any good book, like [1, 4], and also in the notes [3]. In Chapter 8 you will see another approach.

What we will do, instead, is to define conditioning in very special cases, using formulae.

4.6.2 Conditional density

Suppose that (X, Y) is an absolutely continuous random element of $\mathbb{R}^2$ with some joint density $f : \mathbb{R}^2 \to \mathbb{R}$. Then X is an absolutely continuous random variable. Let $f_1 : \mathbb{R} \to \mathbb{R}$ be a density for X. Similarly, Y is an absolutely continuous random variable. Let f_2 be a density for Y. We call f_1, f_2 MARGINAL DENSITIES of f. Define THE CONDITIONAL DENSITY OF X GIVEN Y by

$$f_{X|Y}(x|y) := \begin{cases} \dfrac{f(x,y)}{f_2(y)}, & \text{if } f_2(y) \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Lemma 4.6. *For $B, C \in \mathscr{B}$, let*

$$g_B(y) := \int_B f_{X|Y}(x|y)dx.$$

Then

$$P(X \in B, Y \in C) = \mathbf{E}[g_B(Y)\mathbf{1}_C(Y)],$$

This is hard to prove in general. But you can sort of see it is true by considering

EXERCISE 6. Convince yourselves that the lemma is true if B, C are intervals.

4.6.3 Conditional probability mass function

Let (X, Y) be a random variable on $(\Omega, \mathscr{F}, \mathbf{P})$ with values in the discrete set $(S_1 \times S_2, 2^{S_1 \times S_2})$. As usual, define its JOINT PROBABILITY MASS FUNCTION and MARGINAL PROBABILITY MASS FUNCTIONS by

$$p(x,y) := \mathbf{P}((X,Y) = (x,y)), \quad p_1(x) := \mathbf{P}(X = x), \quad p_2(y) := \mathbf{P}(Y = y), \quad (x,y) \in S_1 \times S_2.$$

Define the CONDITIONAL PROBABILITY MASS FUNCTION OF X GIVEN Y by [b]

$$p_{X|Y}(x|y) := \begin{cases} \dfrac{p(x,y)}{p_2(y)}, & \text{if } p_2(y) \neq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{4.3}$$

[b] The notation $p_{X|Y}$ is terrible. We only use it out of some respect to the undergraduate probability courses. The reason that the notation is terrible is that in the subscript '$X|Y$' in $p_{X|Y}(x|Y)$ plays a merely cosmetic rôle, as opposed to the essential rôle played by the last variable Y inside the parenthesis.

Lemma 4.7.

$$\mathbf{P}(X = x|Y = y) = p_{X|Y}(x|y)p_2(y).$$

Proof If $p_2(y) = 0$ then $\mathbf{P}(X = x, Y = y) = 0$. If $p_2(y) > 0$ then multiply both sides of (4.3) by $p_2(y)$ to obtain the desired result. $\square$

References

[1] B. Fristedt & L. Gray, *A Modern Approach to Probability Theory*, Birkhäuser, 1997.

[2] E. Hewitt & K. Stromberg, *Real and Abstract Analysis*, Springer, 1965.

[3] T. Konstantopoulos, Extended set of lecture notes, with proofs, `www.ma.hw.ac.uk/~takis`

[4] D. Williams, *Probability with Martingales*, Cambridge, 1991.