Weekly Notes for EE2012 2013/14 – Week 11

T. J. Lim

April 6, 2014

Book sections covered this week: 5.4, 5.5.

1 Joint Probability Density Function

1.1 Concept

We saw in the previous week that the joint CDF $F_{X,Y}(x,y)$ can be used to find the probability of a product-form event:

$$P[x_1 < X \le x_2, y_1 < Y \le y_2] = F(x_2,y_2) - F(x_1,y_2) - F(x_2,y_1) + F(x_1,y_1). \quad (1)$$

However, this is not an easy expression to use for the following reasons.

- It is hard[1] to visualize what regions of the x-y plane are more likely to be encountered;
- It can in principle be used for all events including those that are not of product form, by splitting into an infinite number of product-form events. But this is not a practical method of computing probabilities.

When X and Y are both discrete, we have seen that the joint PMF $p_{X,Y}(x_j,y_k)$ allows us to compute all event probabilities. In this section, we deal with another case, where X and Y are said to be *jointly continuous*. This is defined as follows.

> X and Y are jointly continuous if and only if the function
>
> $$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$
>
> exists everywhere, without the need for Dirac delta functions. (Existence of the $(1,1)$-th derivative of $F_{X,Y}(x,y)$ $\Leftrightarrow$ the order of differentiation is immaterial.)

The function $f_{X,Y}(x,y)$ is known as the joint PDF of X and Y, and it has the following important properties:

[1] Even harder than the one-variable case!

1. It is non-negative everywhere –
$$f_{X,Y}(x,y) \geq 0, \quad \forall\, x,y \in \mathbb{R}. \tag{2}$$
2. It has a total volume of 1 –
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1. \tag{3}$$
3. The probability of any event A is given by the volume under the joint PDF enclosed by A –
$$P[A] = \int\int_A f_{X,Y}(x,y)dxdy. \tag{4}$$

We will prove Property 3, and then the other two will follow. Consider the product-form event
$$dA = \{x < X \leq x+dx, y < Y \leq y+dy\}. \tag{5}$$
By (1), we have
$$P[dA] = F(x+dx,y+dy) - F(x,y+dy) - F(x+dx,y) + F(x,y) \tag{6}$$
where we have dropped the $_{X,Y}$ subscript for notational convenience.

If we differentiate $F(x,y)$ with respect to y, we obtain
$$\frac{\partial}{\partial y}F(x,y) = \lim_{dy\to 0}\frac{F(x,y+dy)-F(x,y)}{dy} \tag{7}$$
Then if we differentiate with respect to x, we will have the $(1,1)$-th derivative, i.e.
$$\frac{\partial^2}{\partial x\partial y}F(x,y) = \lim_{dx,dy\to 0}\frac{F(x+dx,y+dy)-F(x+dx,y)-F(x,y+dy)+F(x,y)}{dx\cdot dy} \tag{8}$$
But the LHS is the joint PDF of X and Y, $f_{X,Y}(x,y)$, and the numerator on the RHS is $P[dA]$ from (6). Therefore, for small dx and dy,
$$P[dA] = f_{X,Y}(x,y)\cdot dx\cdot dy. \tag{9}$$

Note that $f_{X,Y}(x,y)dxdy$ is the volume under $f_{X,Y}(x,y)$ enclosed by the region dA. A general event A comprises a union of many contiguous and disjoint dA events, and hence the probability of A is the total volume under the joint PDF, enclosed by A. This proves Property 3. ■

Property 2 follows from Property 3 with $A = \mathbb{R}$, and Property 1 follows since $f_{X,Y}(x,y)dxdy$ is a probability, which must be non-negative, for any $dx > 0$ and $dy > 0$.

1.2 Marginal PDF and Joint CDF from the Joint PDF

If we have the joint PDF of a pair of random variables, how do we obtain their marginal PDFs and joint CDF? We start with the joint CDF:

$$F_{X,Y}(x,y) = P[X \leq x, Y \leq y] = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(x', y') dx' dy'. \tag{10}$$

Since the marginal PDFs $f_X(x)$ and $f_Y(y)$ are by definition the derivatives of the marginal CDFs $F_X(x)$ and $F_Y(y)$, and $F_X(x) = F_{X,Y}(x, \infty)$ and $F_Y(y) = F_{X,Y}(\infty, y)$, we have

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_{X,Y}(x, \infty) && (11) \\ &= \frac{d}{dx} \int_{-\infty}^{x} \int_{-\infty}^{\infty} f_{X,Y}(x', y') dy' dx' && (12) \\ &= \int_{-\infty}^{\infty} f_{X,Y}(x, y') dy'. && (13) \end{aligned}$$

The second line comes from (10), and the third from the fundamental theorem of calculus, i.e.

$$\frac{d}{dx} \int_{a}^{x} g(t) dt = g(x), \quad \forall\, x > a.$$

Similarly,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x', y) dx'. \tag{14}$$

In other words, the marginal PDF of X is obtained by "integrating away" the other variable Y, and the marginal PDF of Y is obtained by "integrating away" the other variable X. This idea can be extended to any number of random variables.

1.3 Examples

Example 1: Consider the jointly uniform PDF

$$f_{X,Y}(x,y) = \begin{cases} 1 & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

The joint CDF of X and Y is obtained after careful treatment of the various regions of the x-y plane.

Where $0 \leq x \leq 1, 0 \leq y \leq 1$ (you need to draw the x-y plane with all the regions), the intersection of $\{X \leq x, Y \leq y\}$ with the region having non-zero $f_{X,Y}(x,y)$ is a rectangle of width x and length y. Since the joint PDF is a constant value equal to 1 within this region, we have

$$F_{X,Y}(x,y) = x \times y \times 1 = xy, \quad 0 \leq x \leq 1, 0 \leq y \leq 1. \tag{16}$$

When $0 \le x \le 1, y > 1$, we have

$$F_{X,Y}(x,y) = P[0 \le X \le x] = x. \tag{17}$$

When $x > 1, 0 \le y \le 1$, we have

$$F_{X,Y}(x,y) = P[0 \le Y \le y] = y. \tag{18}$$

In total,

$$F_{X,Y}(x,y) = \begin{cases} 0 & x < 0 \text{ or } y < 0 \\ xy & 0 \le x \le 1, 0 \le y \le 1 \\ x & 0 \le x \le 1, y > 1 \\ y & x > 1, 0 \le y \le 1 \\ 1 & x > 1, y > 1. \end{cases} \tag{19}$$

Note how the joint CDF is so much more complicated than the joint PDF. ■

Example 2: Consider the joint PDF

$$f_{X,Y}(x,y) = ce^{-x}e^{-y}, \quad 0 \le y < x < \infty \tag{20}$$

and 0 elsewhere. To find c, we use the property that the total volume under the PDF is 1. To set up the integral representing the volume under $f_{X,Y}(x,y)$ requires that we draw the region over which it is non-zero. This would be a wedge formed by $y < x$ and $x \ge 0$.

We can compute the volume either by first integrating over y and then x, or over x then y. We illustrate both methods here.

<u>Method 1</u> Consider a thin slice of the $f(x', y')$ surface between the two lines $y' = y$ and $y' = y + dy$, for some $y > 0$. This thin slice has two parallel surfaces, each with the same area (because dy is small) given by

$$\int_y^\infty ce^{-x'}e^{-y}dx' = ce^{-2y}. \tag{21}$$

This comes from the fact that $f(x', y')$ when $y' = y$ is $f(x', y)$. By treating $f(x', y)$ as a function of x', its area is given by the above. The volume of the thin slice is therefore

$$dV_y = ce^{-2y}dy. \tag{22}$$

Now the entire volume of interest is obtained by "summing" together all the dV_y values over all y, which in the limit $dy \to 0$ is the integral

$$V = \int_0^\infty ce^{-2y}dy = c/2. \tag{23}$$

Since $V = 1$, we have $c = 2$.

<u>Method 2</u> We can also consider slicing $f(x', y')$ along the line $x' = x$, with a width of dx. Then the area on one face of the slice is

$$\int_0^x ce^{-x}e^{-y'}dy' = ce^{-x}(1 - e^{-x}). \tag{24}$$

The incremental volume of the slice at $x' = x$ is

$$dV_x = ce^{-x}(1 - e^{-x})dx, \tag{25}$$

and the total volume is

$$\begin{aligned} V &= \int_0^\infty ce^{-x}(1 - e^{-x})dx & (26)\\ &= c\left[-e^{-x} + \frac{1}{2}e^{-2x}\right]_0^\infty & (27)\\ &= c[1 - 1/2] = c/2. & (28)\end{aligned}$$

Thus the answer is the same as the one obtained via Method 1. ■

Example[2] **3:** In this example (not shown in class), we demonstrate the rather counter-intuitive phenomenon that two marginally continuous random variables X and Y may not be *jointly* continuous.

Let X have the PDF

$$f_X(x) = e^{-x}, \quad x > 0, \tag{29}$$

and define $Y = 2X$. The joint CDF $F_{X,Y}(x, y)$ in the first quadrant of the x-y plane has two forms – one in region $A = \{(x, y) : y > 2x\}$ and another in $B = \{(x, y) : y \leq 2x\}$.

In region A,

$$\begin{aligned} F_{X,Y}(x, y) &= P[X \leq x, Y \leq y] & (30)\\ &= P[X \leq x, X \leq y/2] \quad (\because\ Y = 2X) & (31)\\ &= P[X \leq x] \qquad (\because\ x < y/2) & (32)\\ &= 1 - e^{-x}. & (33)\end{aligned}$$

In region B,

$$\begin{aligned} F_{X,Y}(x, y) &= P[X \leq x, X \leq y/2] & (34)\\ &= P[X \leq y/2] \qquad (\because\ x \geq y/2) & (35)\\ &= 1 - e^{-y/2}. & (36)\end{aligned}$$

But this means that the $(1, 1)$-th derivative of $F_{X,Y}(x, y)$ will be zero everywhere except along $y = 2x$, at which even the first derivative does not exist[3]. Therefore, a joint PDF, whose volume over a region gives the probability of (X, Y) falling within that region, does not exist. X and Y are therefore not jointly continuous.

This observation holds for all cases where Y is a function of X, i.e. $Y = g(X)$. In such situations, we do not need the joint PDF, because any event involving the

[2] Not necessary for all students to study examples 3 and 4.

[3] You can check this by approaching $y = 2x$ from the left and from the right, and seeing that the results are different.

random vector (X,Y) is equivalent to an event involving only X or only Y, and these problems can be dealt with using the single-variable theory. ■

Example 4: Suppose X and Y are both discrete, can we define a joint PDF? Yes, we can if we define a two-dimensional version of the Dirac delta function $\delta(x,y)$, with the property that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\delta(x-x_0,y-y_0)dxdy = g(x_0,y_0),$$

and $\delta(x,y)=0$ for all (x,y) such that $x^2+y^2\neq 0$. We can then place these delta functions at every point in $S_{X,Y}$, with strengths given by the probability of (X,Y) being at that point. In other words,

$$f_{X,Y}(x,y) = \sum_j\sum_k p_{X,Y}(x_j,y_k)\delta(x-x_j,y-y_k), \tag{37}$$

where $p_{X,Y}(x_j,y_k)$ is the joint PMF of (X,Y), and $(x_j,y_k)\in S_{X,Y}$. ■

With the technicalities introduced in Examples 3 and 4, it is appropriate to tighten the definition of a jointly continuous (X,Y) to this:

> X and Y are jointly continuous if and only if the $(1,1)$-th derivative of $F_{X,Y}(x,y)$ exists for all $(x,y)\in\mathbb{R}^2$, *without* allowing the use of impulse functions.

And to observe that when $Y=g(X)$, the probability mass of the vector (X,Y) is concentrated on the line $y=g(x)$ in $\mathbb{R}^2$, and there is no mathematically precise way to represent such a joint distribution in terms of a density function. Nonetheless, since we have well-established tools with which to deal with functions of a random variable, the joint distribution of $(X,g(X))$ is not needed to solve problems of this nature.

2 Independence of X and Y

2.1 Concept

Recall that we had earlier introduced the concept of independence between two events A and B:

$$A,\ B \text{ independent } \Leftrightarrow\ P[A\cap B] = P[A]P[B].$$

What does it mean to say that two *random variables* are independent? It must be that *any* event $\{X\in A\}$ is independent of *any* event $\{Y\in B\}$, where A and B are subsets of $\mathbb{R}$. (With a slight abuse of notation we will for convenience write $P[A]$ in place of $P[X\in A]$ and $P[B]$ in place of $P[Y\in B]$ on occasion.)

In other words

$$X,\, Y \text{ independent } \Leftrightarrow P[X \in A, Y \in B] = P[A]P[B] \tag{38}$$

for any $A \subset \mathbb{R}$ and any $B \subset \mathbb{R}$. The conditional probability

$$P[X \in A | Y \in B] = \frac{P[X \in A, Y \in B]}{P[Y \in B]} = P[X \in A] \tag{39}$$

reveals that, no matter what information is given about Y, there is no effect on our knowledge of the values of X.

2.2 Discrete X and Y

How do we check if X and Y are independent, given that testing the condition (38) is an impossible task except for the simplest cases? For discrete X and Y, we have the following result:

$$X,\, Y \text{ independent } \Leftrightarrow p_{X,Y}(x_j, y_k) = p_X(x_j)p_Y(y_k) \tag{40}$$

for every $(x_j, y_k) \in S_{X,Y}$. The proof is as follows. Suppose X and Y are independent, in the sense of the definition in (38). Then $P[X = x_j, Y = y_k] = P[X = x_j]P[Y = y_k]$, i.e. $p_{X,Y}(x_j, y_k) = p_X(x_j)p_Y(y_k)$. Hence we have proven that independence between X and Y implies that the joint PMF factorizes.

Next suppose the joint PMF factorizes. Let A and B be any subsets of $\mathbb{R}$. Then $P[X \in A, Y \in B]$ is given by the sum of the $p_{X,Y}(x, y)$ values at all points $(x, y) \in \{X \in A, Y \in B\} \cap S_{X,Y}$. In other words,

$$P[X \in A, Y \in B] = \sum_{x \in A} \sum_{y \in B} p_{X,Y}(x, y). \tag{41}$$

But $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ by assumption, and hence

$$P[X \in A, Y \in B] = \sum_{x \in A} p_X(x) \sum_{y \in B} p_Y(y) = P[A]P[B]. \tag{42}$$

Therefore, if the joint PMF factorizes, then X and Y are independent. We have thus proven (40).

2.3 Non-Discrete X and Y

In general, we have that

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) \Leftrightarrow X \text{ and } Y \text{ are independent.} \tag{43}$$

A sketch of the proof is given in the appendix. From (43), assuming the joint PDF exists, we have from differentiating both sides that

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \Leftrightarrow X \text{ and } Y \text{ are independent.} \tag{44}$$

So whether in the jointly discrete or jointly continuous cases, if the joint distribution (CDF, PDF or PMF) factorizes into the product of marginal distributions, then X and Y are independent, and vice versa.

2.4 Examples

Example 5: Consider the joint PDF

$$f_{X,Y}(x,y) = 2e^{-x}e^{-y}, \quad 0 \le y < x < \infty.$$

The marginal PDFs of X and Y can be found as

$$\begin{aligned} f_X(x) &= \int_0^x f_{X,Y}(x,y)dy = 2e^{-x}(1-e^{-x}), \quad x > 0, && (45) \\ f_Y(y) &= \int_y^\infty f_{X,Y}(x,y)dx = 2e^{-2y}, \quad y > 0. && (46) \end{aligned}$$

Therefore $f_X(x)f_Y(y)$ is non-zero everywhere in the first quadrant, $x > 0, y > 0$. But the joint PDF $f_{X,Y}(1,2) = 0$ and so $f_{X,Y}(1,2) \neq f_X(1)f_Y(2)$. X and Y are therefore not independent. ■

Example 6: Let X and Y be jointly uniform in the square region $0 \le x \le 1, 0 \le y \le 1$, i.e.

$$f_{X,Y}(x,y) = 1, \quad 0 \le x \le 1, 0 \le y \le 1. \tag{47}$$

Then the marginal PDFs are

$$\begin{aligned} f_X(x) &= \int_0^1 1dy = 1, \quad 0 \le x \le 1, && (48) \\ f_Y(y) &= \int_0^1 1dx = 1, \quad 0 \le y \le 1, && (49) \end{aligned}$$

and 0 everywhere else. In this case, it should be clear that $f_X(x)f_Y(y) = f_{X,Y}(x,y)$ for all $x, y \in \mathbb{R}$. Therefore, X and Y are independent random variables.

Example 7: Let N be a geometric random variable, with the PMF

$$p_N(n) = (1-p)^{n-1}p, \quad n = 1, 2, 3, \ldots$$

and define Q as the quotient when N is divided by a constant m, and R as the remainder. Are Q and R independent?

Ans: We first note that the event $\{Q = q, R = r\}$ is equivalent to $\{N = qm+r\}$, and thus the joint PMF of Q and R is

$$p_{Q,R}(q,r) = p_N(qm+r) = (1-p)^{qm+r-1}p, \quad qm+r = 1, 2, \ldots$$

Note too that the set $\{(q,r) : qm+r = 1, 2, \ldots\}$ excludes $(0,0)$, i.e. $P[Q = 0, R = 0] = 0$. On the other hand, the marginal PMFs of Q and R evaluated at 0 are

$$\begin{aligned} p_Q(0) &= P[N \in \{1, 2, \ldots, m-1\}] = \sum_{n=1}^{m-1} p_N(n) \\ p_R(0) &= P[N \in \{m, 2m, 3m, \ldots\}] = \sum_{k=1}^{\infty} p_N(km). \end{aligned}$$

Clearly $p_Q(0) \neq 0$ and $p_R(0) \neq 0$, and therefore, $p_Q(0)p_R(0) \neq p_{Q,R}(0,0)$, and hence Q and R are not independent.

Interestingly, the answer changes completely, if we allow N to take values in $\{0, 1, 2, \ldots\}$. This example is shown in the book, in Examples 5.9 and 5.20. ■

3 Summary

- The joint PDF is a very useful characterization of a jointly continuous random vector (X, Y). The probability of (X, Y) lying within any region in the x-y plane is obtained from the volume of $f_{X,Y}(x, y)$ covering that region.
- Where the joint PDF has a relatively large value, (X, Y) is more likely to be in the neighbourhood of that point than in another neighbourhood of the same size elsewhere.
- Marginal PDFs can be obtained from the joint PDF, but not vice versa.
- Independence between X and Y means that no information about X can be obtained from Y and vice versa.
- Independence of X and Y holds if and only if their joint distribution (CDF, PDF or PMF) factorizes into the product of marginal distributions.

4 Diagnostic Questions

1. If the joint PDF of X and Y is

$$f_{X,Y}(x, y) = \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right)$$

   show that X and Y are independent, and identically distributed Gassian random variables.

2. Let X and Y be jointly uniform in the unit circle. Find $P[X > 0, Y > 0]$ and $P[0 < \tan^{-1}(Y/X) \leq \pi/4]$.
3. If X and Y are the outcomes of two independent dice rolls, find the joint PMF of X and Y.

A Outline of Proof of (43)

First, assume that X and Y are independent. Then clearly

$$F_{X,Y}(x, y) = P[X \leq x, Y \leq y] = P[X \leq x]P[Y \leq y] = F_X(x)F_Y(y). \tag{50}$$

Next, assume that $F_{X,Y}(x,y) = F_X(x)F_Y(y)$, and that $A = (x_1, x_2]$ and $B = (y_1, y_2]$. Then

$$\begin{aligned} P[X \in A, Y \in B] &= F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \\ &= F(x_2)F(y_2) - F(x_1)F(y_2) - F(x_2)F(y_1) + F(x_1)F(y_1) \\ &= F(x_2)[F(y_2) - F(y_1)] - F(x_1)[F(y_2) - F(y_1)] \\ &= [F(x_2) - F(x_1)][F(y_2) - F(y_1)] \\ &= P[X \in A]P[Y \in B], \end{aligned}$$

where it is to be understood that in the above, $F(x_i) = F_X(x_i)$ and $F(y_i) = F_Y(y_i)$. Thus, A and B are independent.

Now consider the case of A and B being arbitrary subsets of $\mathbb{R}$. Every such A is a union of subsets $(x_i, x_{i+1}]$ where $x_i \leq x_{i+1}$; similarly every B is a a union of $(y_i, y_{i+1}]$ where $y_i \leq y_{i+1}$, i.e.

$$A = \bigcup_{i=0}^{m-1} (x_i, x_{i+1}] \tag{51}$$

$$B = \bigcup_{j=0}^{n-1} (y_j, y_{j+1}]. \tag{52}$$

$\{X \in A, Y \in B\}$ is a union of mn elementary product-form events, e.g. $\{x_2 < X \leq x_3, y_4 < Y \leq y_5\}$. We have shown that when $m = n = 1$, factorization of $F_{X,Y}(x,y)$ implies $\{X \in A\}$ and $\{Y \in B\}$ are independent events. To generalize, we need to show that if $\{X \in A\}$ and $\{Y \in B\}$ are independent for some m and n, then incrementing m or n or both by 1 retains the independence property. This is doable but tedious. Assuming this step is taken, then we have shown by induction that $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ implies independence.