Weekly Notes for EE2012 2014/15 – Week 10

T. J. Lim

March 19, 2015

Book sections covered this week: 5.2–5.4.

1 Both X and Y are Discrete

1.1 Joint Probability Mass Function (PMF)

When X and Y are both discrete (as in Example B), then we would normally use the joint PMF to handle them:

$$p_{X,Y}(x_j, y_k) = P[X = x_j, Y = y_k], \quad j = 1, 2, \ldots, k = 1, 2, \ldots \tag{1}$$

where $S_X = \{x_1, x_2, \ldots\}$ and $S_Y = \{y_1, y_2, \ldots\}$. Once we have the joint PMF, all events involving (X, Y) will have known probabilities, because

$$P[(X, Y) \in A] = \sum_{(x,y)\in A}\sum p_{X,Y}(x, y) \tag{2}$$

where $A \subset S_{X,Y}$. Clearly, if $A = S_{X,Y}$, then $P[A] = 1$, and hence we have

$$\sum_{(x,y)\in S_{X,Y}}\sum p_{X,Y}(x, y) = 1. \tag{3}$$

Example 1: For the scenario of Example B, we can easily find the joint PMF of (X, Y) as shown in class. The final result is

$$p_{X,Y}(j, k) = \begin{cases} \frac{1}{36} & j = k \\ \frac{1}{18} & j > k \end{cases} \tag{4}$$

where j and k both belong in $\{1, 2, 3, 4, 5, 6\}$. Define the following events:

$$A = \{X \leq 2, Y \leq 3\}. \; B = \{X > 4, Y \leq 3\} \text{ and } C = \{2 \leq X \leq 4, 1 \leq Y \leq 3\}. \tag{5}$$

The probabilities of A, B and C are found using (2) as

$$P[A] = p_{X,Y}(1,1) + p_{X,Y}(2,1) + p_{X,Y}(2,2) = \frac{1}{9} \tag{6}$$

$$P[B] = \sum_{j=5}^{6}\sum_{k=1}^{3} p_{X,Y}(j,k) = \frac{1}{3} \tag{7}$$

$$P[C] = \sum_{j=2}^{4}\sum_{k=1}^{3} p_{X,Y}(j,k) = \frac{7}{18}. \tag{8}$$

Similarly, the probability of any other event can be found by first identifying the points with non-zero probability within the region of interest, and then summing up their PMF values. ■

1.2 Marginal PMF

From the joint PMF, we can find the marginal PMFs of X and of Y – these are the single-variable PMFs that we have studied in depth in previous lessons. The "marginalizing" of a joint PMF makes use of the idea of partitioning a sample space into mutually exclusive events. By considering the partition $\bigcup_{y_k \in S_Y}\{Y = y_k\} = S_{X,Y}$, we get the marginal PMF of X:

$$p_X(x_j) = P[X = x_j] = \sum_{y_k \in S_Y} P[X = x_j, Y = y_k] \tag{9}$$

$$= \sum_{y_k \in S_Y} p_{X,Y}(x_j, y_k). \tag{10}$$

Similarly, the partition $\bigcup_{x_j \in S_X}\{X = x_j\} = S_{X,Y}$ yields the marginal PMF of Y:

$$p_Y(y_k) = \sum_{x_j \in S_X} p_{X,Y}(x_j, y_k). \tag{11}$$

In other words, when finding the marginal PMF of X, sum the joint PMF over $y_k \in S_Y$; when finding the marginal PMF of Y, sum it over $x_j \in S_X$.

Note that it is not possible to obtain the joint PMF from the marginal PMFs of X and Y without some additional information or assumptions (e.g. that X and Y are independent). This is because many joint PMFs will lead to the same marginal PMFs, and therefore we cannot uniquely identify the joint PMF from the marginals.

2 Joint Cumulative Distribution Function (CDF)

If X and Y are both discrete, the joint PMF is all we need. Unfortunately, that scenario is rather limiting, and often we have to deal with both X and Y being continuous (as in Example A), or one being discrete and the other continuous (as in Example C). To handle such cases, we need to first introduce the joint CDF.

We define the joint CDF of X and Y as

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y). \tag{12}$$

It has the following properties:

1. It is non-decreasing with x and with y, i.e. for any $x_1 \leq x_2$ and $y_1 \leq y_2$, we must have
$$F_{X,Y}(x_1,y_1) \leq F_{X,Y}(x_2,y_2). \tag{13}$$
Graphically, $F_{X,Y}(x,y)$ is non-decreasing when looking towards the "north east" standing at any point (x,y).

2. $F_{X,Y}(x,-\infty) = F_{X,Y}(-\infty,y) = 0$ for any x and y, and $F_{X,Y}(\infty,\infty) = 1$. This is because $F_{X,Y}(x,-\infty) = P[X \leq x, Y \leq -\infty]$, but since $Y \leq -\infty$ is impossible, $F_{X,Y}(x,-\infty) = 0$. For a similar reason, $F_{X,Y}(-\infty,y) = 0$. At the opposite extreme, $\{X \leq \infty, Y \leq \infty\}$ is the certain event, and thus $F_{X,Y}(\infty,\infty) = 1$.

3. The marginal CDFs of X and Y are
$$F_X(x) = F_{X,Y}(x,\infty) \quad \text{and} \quad F_Y(y) = F_{X,Y}(\infty,y). \tag{14}$$
This is because $P[X \leq x] = P[X \leq x, Y \leq \infty]$ and $P[Y \leq y] = P[X \leq \infty, Y \leq y]$.

4. The joint CDF is continuous from the "north east", i.e.
$$\begin{aligned}
\lim_{dx \to 0^+} F_{X,Y}(x+dx,y) &= F_{X,Y}(x,y) \\
\lim_{dy \to 0^+} F_{X,Y}(x,y+dy) &= F_{X,Y}(x,y) \\
\lim_{dx \to 0^+} \lim_{dy \to 0^+} F_{X,Y}(x+dx,y+dy) &= F_{X,Y}(x,y).
\end{aligned}$$
This just means that if there is a step discontinuity in $F_{X,Y}$ at (x',y'), then $F_{X,Y}(x',y') = F_{X,Y}(x'+\epsilon_x, y'+\epsilon_y)$ in the limit $\epsilon_x, \epsilon_y \to 0$ from above.

5. The probability of (X,Y) lying in any rectangular region, with edges parallel to the axes[1], can be obtained from the joint CDF:
$$P[x_1 < X \leq x_2, y_1 < Y \leq y_2] = F(x_2,y_2) - F(x_1,y_2) - F(x_2,y_1) + F(x_1,y_1) \tag{15}$$
where the subscript ${}_{X,Y}$ has been dropped from the joint CDF notation for convenience. This was proven in class.

[1]This is known as a product-form event.

Property 5 is particularly important, because it allows us to claim that the probability of any event involving (X, Y) can be found from $F_{X,Y}(x, y)$. The reason is that any region in the x-y plane can be approximated arbitrarily well with a set of non-overlapping rectangles, which represent mutually exclusive events. Therefore the probability of (X, Y) lying in a region is the sum of the probabilities of (X, Y) lying in those rectangles, which can be found using (15) in principle.

Example 2: Consider the joint CDF

$$F_{X,Y}(x,y) = \begin{cases} 0, & x < 0 \text{ or } y < 0 \\ xy, & 0 \le x \le 1, 0 \le y \le 1 \\ x, & 0 \le x \le 1, y > 1 \\ y, & 0 \le y \le 1, x > 1 \\ 1, & x \ge 1, y \ge 1 \end{cases} \tag{16}$$

By identifying the regions in the x-y plane in which $F(x, y)$ takes distinct forms, and then "slicing" the surface representing the function along lines such as $y = 0.2$ and $y = 0.4$, we saw that for a given y value, $F(x, y)$ is a function of x which can be plotted in the usual way. For instance,

$$\begin{aligned} F_{X,Y}(x, 0.2) &= \begin{cases} 0 & x < 0 \\ 0.2x & 0 \le x \le 1 \\ 0.2 & x > 1 \end{cases} \\ F_{X,Y}(x, 0.4) &= \begin{cases} 0 & x < 0 \\ 0.4x & 0 \le x \le 1 \\ 0.4 & x > 1 \end{cases} \end{aligned}$$

We can use $F_{X,Y}(x, y)$ to find various probabilities, as well as the marginal CDFs of X and Y:

$$\begin{aligned} F_X(x) &= F_{X,Y}(x, \infty) = \begin{cases} 0 & x < 0 \\ x & 0 \le x \le 1 \\ 1 & x > 1 \end{cases} \\ P[0.2 < X \le 1, 0.5 < Y \le 1.5] &= F(1, 1.5) - F(0.2, 1.5) - F(1, 0.5) \\ &\quad +F(0.2, 0.5) \\ &= 1 - 0.2 - 0.5 + 0.1 \\ &= 0.4. \end{aligned}$$

For general non-product form events, the CDF cannot be used directly and we have to resort to the joint PDF, to be discussed next, or the joint PMF in the case of discrete (X, Y). ■

Example 3: Let X and Y have the joint CDF

$$F_{X,Y}(x, y) = (1 - e^{-\alpha x})(1 - e^{-\beta y}) \tag{17}$$

for $x > 0$ and $y > 0$. Then we can find $P[X > 1, Y > 2]$ as

$$\begin{aligned} P[X > 1, Y > 2] &= 1 - P[\{X \le 1\} \cup \{Y \le 2\}] \\ &= 1 - (P[X \le 1] + P[Y \le 2] - P[X \le 1, Y \le 2]) \\ &= 1 - F(1, \infty) - F(\infty, 2) + F(1, 2) \\ &= 1 - (1 - e^{-\alpha}) - (1 - e^{-2\beta}) + (1 - e^{-\alpha})(1 - e^{-2\beta}) \\ &= e^{-(\alpha + 2\beta)} \end{aligned}$$

2.1 Comments on Joint CDF

As in the single-variable case, the CDF in the two-variable case is not very easy to manipulate, and it does not offer a quick visualization of the likely values of the random vector (X, Y). The joint PMF is better in that respect, but exists only for discrete random vectors. Next, we will study the joint PDF, which describes *jointly continuous* random variables X and Y, and also briefly touch on how to handle the case of X being discrete and Y being continuous.

3 Joint Probability Density Function

3.1 Concept

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

$$P[x_1 < X \le x_2, y_1 < Y \le y_2] = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1). \quad (18)$$

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

- It is hard[2] 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.)

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

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. It is non-negative everywhere –
$$f_{X,Y}(x, y) \geq 0, \quad \forall\, x, y \in \mathbb{R}. \tag{19}$$
2. It has a total volume of 1 –
$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = 1. \tag{20}$$
3. The probability of any event A is given by the volume under the joint PDF enclosed by A –
$$P[A] = \int \int_A f_{X,Y}(x, y) dx dy. \tag{21}$$

We will prove Property 3, and then the other two will follow. Consider the product-form event
$$dA = \{x < X \leq x + dx, y < Y \leq y + dy\}. \tag{22}$$
By (18), we have
$$P[dA] = F(x + dx, y + dy) - F(x, y + dy) - F(x + dx, y) + F(x, y) \tag{23}$$
where we have dropped the $_{X,Y}$ subscript for notational convenience.

If we differentiate $F(x, y)$ with respect to y, we obtain
$$\frac{\partial}{\partial y} F(x, y) = \lim_{dy \to 0} \frac{F(x, y + dy) - F(x, y)}{dy} \tag{24}$$
Then if we differentiate with respect to x, we will have the $(1, 1)$-th derivative, i.e.
$$\frac{\partial^2}{\partial x \partial y} F(x, y) = \lim_{dx,dy \to 0} \frac{F(x + dx, y + dy) - F(x + dx, y) - F(x, y + dy) + F(x, y)}{dx \cdot dy} \tag{25}$$
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 (23). Therefore, for small dx and dy,
$$P[dA] = f_{X,Y}(x, y) \cdot dx \cdot dy. \tag{26}$$

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

3.2 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:

$$F_{X,Y}(x,y) = P[X \leq x, Y \leq y] = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(x',y')dx'dy'. \tag{27}$$

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

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_{X,Y}(x,\infty) && (28)\\ &= \frac{d}{dx}\int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{X,Y}(x',y')dy'dx' && (29)\\ &= \int_{-\infty}^{\infty} f_{X,Y}(x,y')dy'. && (30) \end{aligned}$$

The second line comes from (27), 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,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x',y)dx'. \tag{31}$$

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 in fact be extended to any number of random variables.

3.3 Examples

Example 4: Consider the jointly uniform PDF

$$f_{X,Y}(x,y) = \begin{cases} 1 & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{32}$$

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

$$F_{X,Y}(x,y) = x \times y \times 1 = xy, \quad 0 \leq x \leq 1, 0 \leq y \leq 1. \tag{33}$$

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

$$F_{X,Y}(x,y) = P[0 \leq X \leq x] = x. \tag{34}$$

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

$$F_{X,Y}(x,y) = P[0 \leq Y \leq y] = y. \tag{35}$$

In total,

$$F_{X,Y}(x,y) = \begin{cases} 0 & x < 0 \text{ or } y < 0 \\ xy & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ x & 0 \leq x \leq 1, y > 1 \\ y & x > 1, 0 \leq y \leq 1 \\ 1 & x > 1, y > 1. \end{cases} \tag{36}$$

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

Example 5: Consider the joint PDF

$$f_{X,Y}(x,y) = ce^{-x}e^{-y}, \quad 0 \leq y < x < \infty \tag{37}$$

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 $y < x$, $y \geq 0$ and $x \geq 0$.

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

Method 1 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

$$\int_y^\infty ce^{-x'}e^{-y}dx' = ce^{-2y}. \tag{38}$$

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

$$dV_y = ce^{-2y}dy. \tag{39}$$

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

$$V = \int_0^\infty ce^{-2y}dy = c/2. \tag{40}$$

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

Method 2 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

$$\int_0^x ce^{-x}e^{-y'}dy' = ce^{-x}(1 - e^{-x}). \tag{41}$$

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

$$dV_x = ce^{-x}(1 - e^{-x})dx, \tag{42}$$

and the total volume is

$$\begin{aligned} V &= \int_0^\infty ce^{-x}(1 - e^{-x})dx & (43) \\ &= c\left[-e^{-x} + \frac{1}{2}e^{-2x}\right]_0^\infty & (44) \\ &= c[1 - 1/2] = c/2. & (45) \end{aligned}$$

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

4 Diagnostic Questions

1. In Example B, if we re-define X and Y to be the outcomes of the two dice respectively, find the joint PMF of X and Y.

2. Let N be uniform in $\{1, 2, \ldots, 10\}$, and let X be binomial conditioned on $N = n$, in the sense that

$$P[X = k|N = n] = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

   What is the range of (X, N)? Sketch this on a two-dimensional plot.

3. For X and Y having the joint CDF of Example 3, find the marginal CDF of X and hence the expected value of X.

4. For the PDF in Example 5, find $P[Y < 2]$ and $P[X > 3, Y < 2]$.