STA 623 – Fall 2013 – Dr. Charnigo

Section 4.1: Joint and Marginal Distributions

Discrete random vector and joint probability mass function. Recall that a random variable is a(n appropriately measurable) function from an underlying sample space Ω to the set of real numbers $\mathbb{R}$. We now define a k-dimensional random vector, k a positive integer, to be a(n appropriately measurable) function from Ω to the k-fold Cartesian product $\mathbb{R}^k$. For now we fix $k = 2$. The random vector may be denoted $\mathbf{X}$, or we may write X and Y to represent its two components. Note that X and Y are themselves random variables.

If $\mathbf{X}$ is a discrete random vector, which is to say that $\mathbf{X}$ realizes only countably many values with positive probability, then X and Y may be described by a joint probability mass function $f_{X,Y}(x, y) := P(X = x, Y = y)$ such that $P((X, Y)' \in A) = \sum_{\{(x,y)' \in A \cap \mathcal{S}\}} f_{X,Y}(x, y)$ for any (appropriately measurable) set $A \subset \mathbb{R}^2$, where $\mathcal{S} := \{(x, y)' \in \mathbb{R}^2 : f_{X,Y}(x, y) > 0\}$ is the support set of $\mathbf{X}$. The introduction of $\mathcal{S}$ here avoids the question of how to define an uncountable sum. Summation over the empty set is defined as 0. Note that we must have $f_{X,Y}(x, y) \geq 0$ and $\sum_{\{(x,y)' \in \mathcal{S}\}} f_{X,Y}(x, y) = 1$.

Example (discrete random vector and joint probability mass function). Suppose that X is the number of phone messages received in the next hour and has the Poisson distribution with mean $\lambda \in (0, \infty)$, while Y is the number of text messages received in the next hour and has the Poisson distribution with mean $\mu \in (0, \infty)$. If phone messages arrive independently of text messages, then for any nonnegative integers x and y we have

$$f_{X,Y}(x, y) =$$

This illustrates that, in some cases, the joint probability mass function for X and Y is simply the product of the probability mass function for X with the probability mass function for Y. However, this is not always true. Suppose that for $\{(x,y)' \in \mathbb{R}^2 : x \in \{0,1\}, y \in \{0,1\}\}$ we have

$$f_{X,Y}(x,y) = (x + 2y + 1)/10.$$

Clearly, $(x+2y+1)/10$ cannot be written as a product of a function of x with a function of y. Yet, $\sum_{\{(x,y)' \in \mathcal{S}\}}(x+2y+1)/10 = 1/10+2/10+3/10+4/10 = 1$, so $f_{X,Y}(x,y)$ is a valid joint probability mass function.

Marginal probability mass function. Since X and Y are themselves random variables, we may be interested in describing their distributions individually. If $\mathbf{X}$ is a discrete random vector, this can be accomplished by summing the joint probability mass function over appropriate sets. Explicitly, let u be any real number and put $A := \{(x,y)' \in \mathbb{R}^2 : x = u\}$. Then $\sum_{\{(x,y)' \in A \cap \mathcal{S}\}} f_{X,Y}(x,y)$ is simply $P(X = u)$ and may be labeled $f_X(u)$. In this way we recover the probability mass function of X, which is called a marginal probability mass function for a reason that will become clear presently. The marginal probability mass function of Y may be recovered similarly. I emphasize that, while marginal probability mass functions are uniquely determined from the joint probability mass function, the reverse is not true.

Example (marginal probability mass function). If the joint probability mass function was obtained by multiplying the marginal probability mass functions, then presumably there is no need to derive the marginal probability mass functions using the approach indicated in the last paragraph. However, suppose that the joint probability mass function is not the product of a function of x with a function of y, as in the previous example. Then the approach indicated in the last paragraph is useful. Indeed, with $f_{X,Y}(x,y) = (x + 2y + 1)/10$ for $\{(x,y)' \in \mathbb{R}^2 : x \in \{0,1\}, y \in \{0,1\}\}$ we have

$f_X(0) =$ $\qquad\qquad$ $f_X(1) =$

$f_Y(0) =$ $\qquad\qquad$ $f_Y(1) =$

When the joint probability mass function $f_{X,Y}(x, y)$ is listed in tabular form as below, the marginal probability mass functions $f_X(x)$ and $f_Y(y)$ may be read off the margins of the table.

$f_{X,Y}(x, y)$	$y = 0$	$y = 1$	$f_X(x)$
$x = 0$	$1/10$	$3/10$	
$x = 1$	$2/10$	$4/10$	
$f_Y(y)$			1

To clarify my point of emphasis in the preceding paragraph, can you exhibit a different joint probability mass function that is compatible with the above marginal probability mass functions?

Expected value. Let $g(x, y)$ be a(n appropriately measurable) real-valued function of x and y. We define the expected value $E[g(X, Y)]$ as $\sum_{\{(x,y)' \in \mathcal{S}\}} g(x, y) f_{X,Y}(x, y)$, provided that the sum is absolutely convergent. Even though we are now studying random vectors, expected value has the same linearity and monotonicity properties discussed earlier this semester.

Example (expected value). Continuing from the last example, put $g(X, Y) := (X + 1)(Y + 1)$. Then

$E[g(X, Y)] =$

Now put $g(X, Y) := X^2$. Then

$E[g(X, Y)] =$

However, since X^2 depends only on X (i.e., not also on Y), we can just as well calculate $E[X^2]$ using the marginal probability mass function of X. When we do so, we obtain

$E[X^2] =$

Continuous random vector and joint probability density function. With random vectors we still face an issue like that encountered in Section 1.5 on random variables, namely that non-discrete quantities are not necessarily continuous. Indeed, we can explicitly define continuous random vectors to be those for which the joint cumulative distribution function $F_{X,Y}(x,y) := P(X \le x, Y \le y)$ is continuous. However, a more practical (but stringent) working definition is that there exist a joint probability density function $f_{X,Y}(x,y)$ such that $P(\mathbf{X} \in A) = \int\int_A f_{X,Y}(x,y)\ dx\ dy$ for any (appropriately measurable) set $A \subset \mathbb{R}^2$. In particular, $F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)\ du\ dv$ and, if $f_{X,Y}(x,y)$ is continuous, $\frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y) = f_{X,Y}(x,y)$. Note that we must have $\int\int_{\mathbb{R}^2} f_{X,Y}(x,y)\ dx\ dy = 1$ and that we may as well require $f_{X,Y}(x,y) \ge 0$.

Example (continuous random vector and joint probability density function). Suppose that $\mathbf{X}$ is a continuous random vector with joint probability density function

$$f_{X,Y}(x,y) = 8xy1_{\{0<x<y<1\}}.$$

To verify that this is a valid joint probability density function, note that

$$\int\int_{\mathbb{R}^2} f_{X,Y}(x,y)\ dx\ dy = \int_0^1 \left\{ \int_0^y 8xy\ dx \right\}\ dy$$

$$= \int_0^1 8y\{x^2/2\}_0^y\ dy = \int_0^1 4y^3\ dy = \{y^4\}_0^1 = 1.$$

Suppose that we want to find $F_{X,Y}(x,y)$. Consider five cases:

1. $x \le 0$ or $y \le 0 \implies F_{X,Y}(x,y) = 0$.
2. $x \ge 1$ and $y \ge 1 \implies F_{X,Y}(x,y) = 1$.
3. $x \ge y$ and $0 \le y \le 1 \implies F_{X,Y}(x,y) = \int_0^y \left\{\int_0^v 8uv\ du\right\}\ dv$ $= \int_0^y 4v^3\ dv = y^4$.
4. $y \ge 1$ and $0 \le x \le 1 \implies F_{X,Y}(x,y) = \int_0^x \left\{\int_u^1 8uv\ dv\right\}\ du$ $= \int_0^x 4(u-u^3)\ du = 2x^2 - x^4$.
5. $0 \le x \le y \le 1 \implies F_{X,Y}(x,y) = \int_0^x \left\{\int_u^y 8uv\ dv\right\}\ du$

$= \int_0^x 4(uy^2 - u^3)\, du = 2x^2y^2 - x^4.$

Suppose that we want to find $P(XY < 1/2)$. This is

Marginal probability density function. Just as a marginal probability mass function is obtained by summing a joint probability mass function, a marginal probability density function is obtained by integrating a joint probability density function. Explicitly, we have $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$ and $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx$. An easy way to remember which is which is to note that the marginal probability density function of X must depend on x, so the y should be integrated out.

Example (marginal probability density function). Continuing from the previous example, we have

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy = \int_x^1 8xy\, dy$$

$$=$$

for $x \in (0,1)$ and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx = \int_0^y 8xy\, dx$$

$$=$$

for $y \in (0,1)$.

Expected value. Let $g(x, y)$ be a(n appropriately measurable) real-valued function of x and y. We define the expected value $E[g(X, Y)]$ as $\int\int_{\mathbb{R}^2} g(x, y) f_{X,Y}(x, y)\, dx\, dy$, provided that the integral is absolutely convergent. Again, expected value has the same linearity and monotonicity properties discussed earlier this semester.

Example (expected value). Continuing from the previous example, put $g(X, Y) := X^\gamma Y^\delta$ for constants $\gamma \in [0, \infty)$ and $\delta \in [0, \infty)$. We have

$E[g(X, Y)] =$

Starting a new example, suppose that $f_{X,Y}(x, y)$ has the form $f_X(x) f_Y(y)$, which is to say that the joint probability density function is the product of the marginal probability density functions. I claim that, in this case, $Var[X+Y] = Var[X]+Var[Y]$ if all of these quantities exist as finite numbers. To prove my claim, I begin by noting that

$$E[XY] = \int\int_{\mathbb{R}^2} xy f_{X,Y}(x, y)\, dx\, dy = \int_{\mathbb{R}} x f_X(x)\, dx \int_{\mathbb{R}} y f_Y(y)\, dy = E[X]E[Y].$$

Then, putting $\mu := E[X]$ and $\nu := E[Y]$, I have

$$\begin{aligned} Var[X + Y] &= E[(X + Y)^2] - (E[X + Y])^2 \\ &= E[X^2 + 2XY + Y^2] - (\mu^2 + 2\mu\nu + \nu^2) \\ &= E[X^2] - \mu^2 + E[Y^2] - \nu^2 + 2E[XY] - 2\mu\nu \\ &= E[X^2] - \mu^2 + E[Y^2] - \nu^2 + 2E[X]E[Y] - 2\mu\nu \\ &= E[X^2] - \mu^2 + E[Y^2] - \nu^2 \\ &= Var[X] + Var[Y]. \end{aligned}$$

You will learn next week that X and Y are called independent if their joint probability density function decomposes into the product of their marginal probability density functions.