Joint and Marginal Distributions

October 23, 2008

We will now consider more than one random variable at a time. As we shall see, developing the theory of **multivariate** distributions will allow us to consider situations that model the actual collection of data and form the foundation of inference based on those data.

1 Discrete Random Variables

We begin with a pair of discrete random variables X and Y and define the **joint (probability) mass function**

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

Example 1. *For X and Y each having finite range, we can display the mass function in a table.*

		x				
		0	1	2	3	4
	0	0.02	0.02	0	0.10	0
	1	0.02	0.04	0.10	0	0
y	2	0.02	0.06	0	0.10	0
	3	0.02	0.08	0.10	0	0.05
	4	0.02	0.10	0	0.10	0.05

As with **univariate** random variables, we compute probabilities by adding the appropriate entries in the table.

$$P\{(X, Y) \in A\} = \sum_{(x,y)\in A} f_{(X,Y)}(x, y).$$

Exercise 2. *Find*

1. $P\{X = Y\}$
2. $P\{X + Y \leq 3\}$.
3. $P\{XY = 0\}$.
4. $P\{X = 3\}$.

As before, the mass function has two basic properties.

- $f_{X,Y}(x, y) \geq 0$ for all x and y.

- $\sum_{x,y} f_{X,Y}(x,y) = 1$.

The distribution of an individual random variable is call the **marginal distribution**. The **marginal mass function** for X is found by summing over the appropriate column and the marginal mass function for Y can be found be summing over the appropriate row.

$$f_X(x) = \sum_y f_{X,Y}(x,y), \quad f_Y(y) = \sum_x f_{X,Y}(x,y)$$

The marginal mass functions for the example above are

x	$f_X(x)$
0	0.10
1	0.30
2	0.20
3	0.30
4	0.10

y	$f_Y(y)$
0	0.14
1	0.16
2	0.18
3	0.25
4	0.27

Exercise 3. *Give two pairs of random variables with different joint mass functions but the same marginal mass functions.*

The definition of expectation in the case of a finite sample space S is a straightforward generalization of the univarate case.

$$Eg(X,Y) = \sum_{s \in S} g(X(s), Y(s))P\{s\}.$$

From this formula, we see that expectation is again a positive linear functional. Using the distributive property, we have the formula

$$Eg(X,Y) = \sum_{x,y} g(x,y) f_{X,Y}(x,y).$$

Exercise 4. *Compute EXY in the example above.*

2 Continuous Random Variables

For continuous random variables, we have the notion of the **joint (probability) density function**

$$f_{X,Y}(x,y)\Delta x \Delta y \approx P\{x < X \leq x + \Delta x, y < Y \leq y + \Delta y\}.$$

We can write this in integral form as

$$P\{(X,Y) \in A\} = \int\int_A f_{X,Y}(x,y)\, dy dx.$$

The basic properties of the joint density function are

- $f_{X,Y}(x,y) \geq 0$ for all x and y.

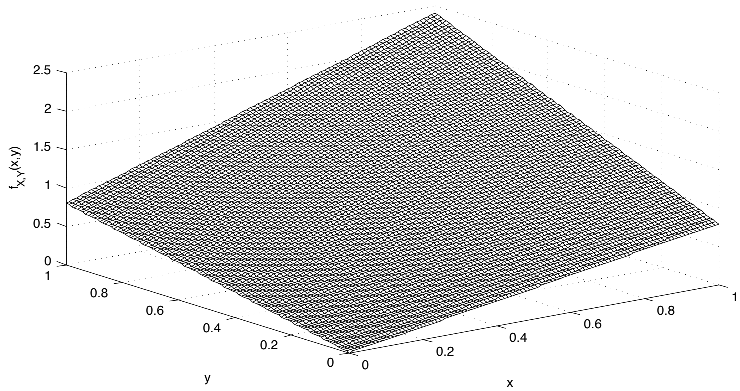

Figure 1: Graph of density $f_{X,Y}(x,y) = 4(xy + x + y)/5$, $0 \le x, y \le 1$

- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dydx = 1$.

Example 5. *Let* (X,Y) *have joint density*

$$f_{X,Y}(x,y) = \begin{cases} c(xy + x + y) & \text{for } 0 \le x \le 1, 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dydx &= \int_0^1 \int_0^1 c(xy + x + y)\, dydx \\ &= c \int_0^1 \left(\frac{1}{2}xy^2 + xy + \frac{1}{2}y^2 \right) \Big|_0^1 dx = c \int_0^1 \left(\frac{3}{2}x + \frac{1}{2} \right) dx \\ &= c \left(\frac{3}{4}x^2 + \frac{1}{2}x \right) \Big|_0^1 = \frac{5c}{4} \end{aligned}$$

and $c = 4/5$

$$\begin{aligned} P\{X \ge Y\} &= \int_0^1 \int_0^x \frac{4}{5}(xy + x + y)\, dydx = \frac{4}{5} \int_0^1 \left(\frac{1}{2}xy^2 + xy + \frac{1}{2}y^2 \right) \Big|_0^x dx \\ &= \frac{4}{5} \int_0^1 \left(\frac{1}{2}x^3 + \frac{3}{2}x^2 \right) dx = \frac{4}{5} \left(\frac{1}{8}x^4 + \frac{1}{2}x^3 \right) \Big|_0^1 = \frac{4}{5} \cdot \frac{5}{8} = \frac{1}{2}. \end{aligned}$$

The **joint cumulative distribution function** is defined as

$$F_{X,Y}(x,y) = P\{X \le x, Y \le y\}.$$

For the case of continuous random variables, we have

$$F_{X,Y}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(s,t)\, dtds.$$

By two applications of the fundamental theorem of calculus, we find that

$$\frac{\partial}{\partial y}F_{X,Y}(x,y) = \int_{-\infty}^{x} f_{X,Y}(s,y)\,dt \quad \text{and} \quad \frac{\partial^2}{\partial x \partial y}F_{X,Y}(x,y) = f_{X,Y}(x,y).$$

Example 6. *For the density introduced above,*

$$\begin{aligned}
F_{X,Y}(x,y) &= \int_0^y \int_0^x \frac{4}{5}\left(st+s+t\right)\,dtds = \int_0^y \frac{4}{5}\left(\frac{1}{2}st^2+st+\frac{1}{2}t^2\right)\Big|_0^y\,ds \\
&= \int_0^y \frac{4}{5}\left(\frac{1}{2}sy^2+sy+\frac{1}{2}y^2\right)\,ds = \frac{4}{5}\left(\frac{1}{4}s^2y^2+\frac{1}{2}s^2y+\frac{1}{2}sy^2\right)\Big|_0^x \\
&= \frac{4}{5}\left(\frac{1}{4}x^2y^2+\frac{1}{2}x^2y+\frac{1}{2}xy^2\right)
\end{aligned}$$

Notice that $F_{X,Y}(1,1) = 1$

$$P\{X \le \frac{1}{2}, Y \le \frac{1}{2}\} = F_{X,Y}\left(\frac{1}{2},\frac{1}{2}\right) = \frac{4}{5}\left(\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}+\frac{1}{2}\cdot\frac{1}{4}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{4}\right) = \frac{4}{5}\cdot\frac{9}{64} = \frac{9}{80}.$$

The joint cumulative distribution function is right continuous in each variable. It has limits at $-\infty$ and $+\infty$ similar to the univariate cumulative distribution function.

- $\lim_{y\to-\infty} F_{X,Y}(x,y) = 0$ and $\lim_{x\to-\infty} F_{X,Y}(x,y) = 0$.
- $\lim_{x,y\to\infty} F_{X,Y}(x,y) = 1$.

In addition,

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

Thus,

$$F_X(x) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{X,Y}(s,t)\,dtds \quad \text{and} \quad F_Y(y) = \int_{-\infty}^{y}\int_{-\infty}^{\infty} f_{X,Y}(s,t)\,dsdt.$$

Now use the fundamental theorem of calculus to obtain the **marginal densities**.

$$f_X(x) = F_X'(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,t)\,dt \quad \text{and} \quad f_Y(y) = F_Y'(y) = \int_{-\infty}^{\infty} f_{X,Y}(s,y)\,ds.$$

Example 7. *For the example density above, the marginal densities*

$$f_X(x) = \int_0^1 \frac{4}{5}\left(xt+x+t\right)\,dt = \frac{4}{5}\left(\frac{1}{2}xt^2+xt+\frac{1}{2}t^2\right)\Big|_0^1 = \frac{4}{5}\left(\frac{3}{2}x+\frac{1}{2}\right)$$

and

$$f_Y(y) = \frac{4}{5}\left(\frac{3}{2}y+\frac{1}{2}\right).$$

The formula for expectation for jointly continuous random variables is dervied by discretizing X and Y, creating a double Rieman sum and taking a limit. This yields the identity

$$Eg(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)\,dydx.$$

Exercise 8. *Compute EXY in the example above.*

As in the one-dimensional case, we can give a comprehensive formula for expectation using Riemann-Steiltjes integrals

$$Eg(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, dF_{X,Y}(x,y).$$

These can be realized as the limit of Riemann-Steiltjes sums

$$S(g,F) = \sum_{i=1}^{m}\sum_{j=1}^{n} g(x_i,y_j)\Delta F_{X,Y}(x_i,y_j).$$

Here,

$$\Delta F(x_i,y_j) = P\{x_i < X \leq x_i + \Delta x, y_j < Y \leq y_j + \Delta y\}$$

Exercise 9. *Show that*

$$\begin{aligned} & P\{x_i < X \leq x_i + \Delta x, y_j < Y \leq y_j\} \\ = \; & F_{X,Y}(x_i + \Delta x, y_j + \Delta y) - F_{X,Y}(x_i, y_j + \Delta y) - F_{X,Y}(x_i + \Delta x, y_j) + F_{X,Y}(x_i + \Delta x, y_j + \Delta y). \end{aligned}$$