Summary of basic probability theory, part 3

D. Joyce, Clark University

Math 218, Mathematical Statistics, Jan 2008

Joint distributions. When studing two related real random variables X and Y, it is not enough just to know the distributions of each. Rather, the pair (X, Y) has a joint distribution. You can think of (X, Y) as naming a single random variable that takes values in the plane $\mathbf{R}^2$.

Joint and marginal probability mass functions. Let's consider the discrete case first where both X and Y are discrete random variables. The probability mass function for X is $f_X(x) = P(X{=}x)$, and the p.m.f. for Y is $f_Y(y) = P(Y{=}y)$. The joint random variable (X, Y) has its own p.m.f. denoted $f_{(X,Y)}(x, y)$, or more briefly $f(x, y)$:

$$f(x,y) = P((X,Y){=}(x,y)) = P(X{=}x \text{ and } Y{=}y),$$

and it determines the two individual p.m.f.s by

$$f_X(x) = \sum_y f(x,y), \qquad f_Y(y) = \sum_x f(x,y).$$

The individual p.m.f.s are usually called *marginal probability mass functions.*

For example, sssume that the random variables X and Y have the joint probability mass function given in this table.

		Y			
		-1	0	1	2
	-1	0	1/36	1/6	1/12
X	0	1/18	0	1/18	0
	1	0	1/36	1/6	1/12
	2	1/12	0	1/12	1/6

By adding the entries row by row, we find the the marginal function for X, and by adding the entries column by column, we find the marginal function for Y. We can write these marginal functions on the margins of the table.

		Y				f_X
		-1	0	1	2	
	-1	0	1/36	1/6	1/12	5/18
X	0	1/18	0	1/18	0	1/9
	1	0	1/36	1/6	1/12	5/18
	2	1/12	0	1/12	1/6	1/3
f_Y		5/36	1/18	17/36	1/3	

Discrete random variables X and Y are independent if and only if the joint p.m.f is the product of the marginal p.m.f.s

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

In the example above, X and Y aren't independent.

Joint and marginal cumulative distribution functions. Besides the p.m.f.s, there are joint and marginal cumulative distribution functions. The c.d.f. for X is $F_X(x) = P(X{\le}x)$, while the c.d.f. for Y is $F_Y(y) = P(Y{\le}y)$. The joint random variable (X, Y) has its own c.d.f. denoted $F_{(X,Y)}(x, y)$, or more briefly $F(x, y)$:

$$F(x,y) = P(X{\le}x \text{ and } Y{\le}y),$$

and it determines the two marginal p.m.f.s by

$$F_X(x) = \lim_{y\to\infty} F(x,y), \qquad F_Y(y) = \lim_{x\to\infty} F(x,y).$$

Joint and marginal probability density functions. Now let's consider the continuous case where X and Y are both continuous. The last paragraph on c.d.f.s still applies, but we'll have marginal probability density functions $f_X(x)$ and $f_Y(y)$, and a joint probability density function $f(x,y)$ instead of probability mass functions. Of course, the derivatives of the marginal c.d.f.s are the density functions

$$f_X(x) = \frac{d}{dx}F_X(x) \qquad f_Y(y) = \frac{d}{dy}F_Y(y)$$

and the c.d.f.s can be found by integrating the density functions

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt \qquad F_Y(y) = \int_{-\infty}^{y} f_Y(t)\,dt.$$

The joint probability density function $f(x,y)$ is found by taking the derivative of F twice, once with respect to each variable, so that

$$f(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y}F(x,y).$$

(The notation ∂ is substituted for d to indicate that there are other variables in the expression that are held constant while the derivative is taken with respect to the given variable.) The joint cumulative distribution function can be recovered from the joint density function by integrating twice

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)\,dt\,ds.$$

Furthermore, the marginal density functions can be found by integrating joint density function.

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy, \quad f_Y(x) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

Continuous random variables X and Y are independent if and only if the joint density function is the product of the marginal density functions

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

Covariance and correlation. The *covariance* of two random variables X and Y is defined as

$$\operatorname{Cov}(X,Y) = \sigma_{XY} = E((X-\mu_X)(Y-\mu_Y)).$$

It can be shown that

$$\operatorname{Cov}(X,Y) = E(XY) - \mu_X\mu_Y.$$

When X and Y are independent, then $\sigma_{XY} = 0$, but in any case

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + 2\operatorname{Cov}(X,Y) + \operatorname{Var}(Y).$$

Covariance is a bilinear operator, which means it is linear in each coordinate

$$\begin{aligned}
\operatorname{Cov}(X_1+X_2,Y) &= \operatorname{Cov}(X_1,Y) + \operatorname{Cov}(X_2,Y) \\
\operatorname{Cov}(aX,Y) &= a\operatorname{Cov}(X,Y) \\
\operatorname{Cov}(X,Y_1+Y_2) &= \operatorname{Cov}(X,Y_1) + \operatorname{Cov}(X,Y_2) \\
\operatorname{Cov}(X,bY) &= b\operatorname{Cov}(X,Y)
\end{aligned}$$

The *correlation*, or *correlation coefficient*, of X and Y is defined as

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}.$$

Correlation is always a number between -1 and 1.