Probability & Statistics

Dongjiao Ge

djge@cityu.edu.mo

Session 07

Chapter 4
More on joint distribution

We have seen the joint *p.m.f.* of two discrete random variables X and Y, and we have learned what it means for X and Y to be ***independent***.

In this chapter, we examine this further to see measures of ***non-independence*** and ***conditional distributions*** of random variables.

4.1 Covariance and correlation

we consider a pair of discrete random variables X and Y. Remember that X and Y are independent if

$$P(X = a_i, Y = b_j) = P(X = a_i) \cdot P(Y = b_j)$$

holds for any pair (a_i, b_j) of values of X and Y. We introduce a number (called the covariance of X and Y) which gives a measure of how far they are from being independent.

Theorem 3.2(b), where we showed that if X and Y are independent then $Var(X + Y) = Var(X) + Var(Y)$. We found that, in any case,

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2(E(XY) - E(X)E(Y)),$$

and then proved that if X and Y are independent then $E(XY) = E(X)E(Y)$, so that the last term is zero.

We define the covariance of X and Y to be $E(XY) - E(X)E(Y)$. We write $Cov(X, Y)$ for this quantity.

Theorem 4.1 *(a)* $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$.

(b) If X and Y are independent, then $\mathrm{Cov}(X,Y) = 0$.

In fact, a more general version of (a), proved by the same argument, says that

$$\mathrm{Var}(aX + bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X,Y). \tag{4.1}$$

Another quantity closely related to covariance is the correlation coefficient, $corr(X, Y)$, which is just a “normalised” version of the covariance. It is defined as follows:

$$\mathrm{corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}.$$

Theorem 4.2 *Let X and Y be random variables. Then*

(a) $-1 \leq \operatorname{corr}(X,Y) \leq 1$;

(b) if X and Y are independent, then $\operatorname{corr}(X,Y) = 0$;

(c) if $Y = mX + c$ for some constants $m \neq 0$ and c, then $\operatorname{corr}(X,Y) = 1$ *if $m > 0$, and* $\operatorname{corr}(X,Y) = -1$ *if $m < 0$.*

Part (a): consider the variance below

$$0 \leq \mathrm{Var}\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\right)$$

$$\begin{aligned}
\mathrm{Var}\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\right) &= \mathrm{Var}\left(\frac{X}{\sigma_X}\right) + \mathrm{Var}\left(\frac{Y}{\sigma_Y}\right) \pm 2\,\mathrm{Cov}\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right) \\
&= \frac{1}{\sigma_X^2}\mathrm{Var}(X) + \frac{1}{\sigma_Y^2}\mathrm{Var}(Y) \pm 2\,\frac{1}{\sigma_X\sigma_Y}\,\mathrm{Cov}(X,Y) \\
&= \frac{1}{\sigma_X^2}\sigma_X^2 + \frac{1}{\sigma_Y^2}\sigma_Y^2 \pm 2\,\frac{1}{\sigma_X\sigma_Y}\,\sigma_{XY}\ .
\end{aligned}$$

$$\mathrm{Var}\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\right) = 1 + 1 + \pm 2\,\mathrm{Corr}(X,Y)\ . \implies 0 \leq 2 \pm 2\,\mathrm{Corr}(X,Y) \implies -1 \leq \mathrm{Corr}(X,Y) \leq +1\ .$$

Part (b) follows immediately from part (b) of the Theorem 4.1.

For part (c), suppose that $Y = mX + c$. Let $E(X) = \mu$ and $\mathrm{Var}(X) = \alpha$, so that $E(X^2) = \mu^2 + \alpha$. Now we just calculate everything in sight.

$$
\begin{aligned}
E(Y) &= E(mX+c) = mE(X) + c = m\mu + c \\
E(Y^2) &= E(m^2X^2 + 2mcX + c^2) = m^2(\mu^2+\alpha) + 2mc\mu + c^2 \\
\mathrm{Var}(Y) &= E(Y^2) - E(Y)^2 = m^2\alpha \\
E(XY) &= E(mX^2 + cX) = m(\mu^2+\alpha) + c\mu; \\
\mathrm{Cov}(X,Y) &= E(XY) - E(X)E(Y) = m\alpha \\
\mathrm{corr}(X,Y) &= \mathrm{Cov}(X,Y)/\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)} = m\alpha/\sqrt{m^2\alpha^2} \\
&= \begin{cases} +1 & \text{if } m > 0, \\ -1 & \text{if } m < 0. \end{cases}
\end{aligned}
$$

Thus the correlation coefficient is a measure of the extent to which the two variables are related.

- It is +1 if Y increases linearly with X;
- It is 0 if there is no relation between them;
- It is -1 if Y decreases linearly as X increases.

More generally, a positive correlation indicates a tendency for larger X values to be associated with larger Y values; a negative value, for smaller X values to be associated with larger Y values.

Example I have two red pens, one green pen, and one blue pen, and I choose two pens without replacement. Let X be the number of red pens that I choose and Y the number of green pens. Then the joint p.m.f. of X and Y is given by the following table:

		$Y=0$	$Y=1$
X	0	0	$\frac{1}{6}$
	1	$\frac{1}{3}$	$\frac{1}{3}$
	2	$\frac{1}{6}$	0

From this we can calculate the marginal p.m.f. of X and of Y and hence find their expected values and variances:

$$E(X) = 1, \qquad \mathrm{Var}(X) = 1/3,$$
$$E(Y) = 1/2, \qquad \mathrm{Var}(Y) = 1/4.$$

Also, $E(XY) = 1/3$, since the sum

$$E(XY) = \sum_{i,j} a_i b_j P(X = a_i, Y = b_j)$$

contains only one term where all three factors are non-zero. Hence

$$\text{Cov}(X,Y) = 1/3 - 1/2 = -1/6,$$

and

$$\text{corr}(X,Y) = \frac{-1/6}{\sqrt{1/12}} = -\frac{1}{\sqrt{3}}.$$

The negative correlation means that small values of X tend to be associated with larger values of Y. Indeed, if $X = 0$ then Y must be 1, and if $X = 2$ then Y must be 0, but if $X = 1$ then Y can be either 0 or 1.

Example We have seen that if X and Y are independent then $Cov(X, Y) = 0$. <u>However, it doesn't work the other way around.</u> Consider the following joint p.m.f.

$$
\begin{array}{cc|c|ccc|}
 & & \multicolumn{4}{c}{Y} \\
\cline{3-5}
 & & -1 & 0 & 1 \\
\cline{2-5}
 & -1 & \frac{1}{5} & 0 & \frac{1}{5} \\
X & 0 & 0 & \frac{1}{5} & 0 \\
 & 1 & \frac{1}{5} & 0 & \frac{1}{5} \\
\cline{2-5}
\end{array}
$$

Now calculation shows that $E(X) = E(Y) = E(XY) = 0$, so $\text{Cov}(X, Y) = 0$. But X and Y are not independent: for $P(X = -1) = 2/5$, $P(Y = 0) = 1/5$, but $P(X = -1, Y = 0) = 0$.

We call two random variables X and Y ***uncorrelated*** if $Cov(X,Y) = 0$ (in other words, if $corr(X,Y) = 0$). So we can say:

Independent random variables are uncorrelated, but uncorrelated random variables need not be independent.

4.2 Conditional random variables

Remember that the *conditional probability* of event B given event A is $P(B \mid A) = P(A \cap B)/P(A)$.

Suppose that X is a discrete random variable. Then the conditional probability that X takes a certain value a_i, given A, is just

$$P(X = a_i \mid A) = \frac{P(A \text{ holds and } X = a_i)}{P(A)}.$$

This defines the probability mass function of the <u>*conditional random variable*</u> $X \mid A$.

So we can, for example, talk about the **conditional expectation**

$$E(X \mid A) = \sum_i a_i P(X = a_i \mid A).$$

Now the event A might itself be defined by a random variable; for example, A might be the event that Y takes the value b_j. In this case, we have

$$P(X = a_i \mid Y = b_j) = \frac{P(X = a_i, Y = b_j)}{P(Y = b_j)}.$$

In other words, we have taken the column of the joint p.m.f. table of X and Y corresponding to the value $Y = b_j$. The sum of the entries in this column is just $P(Y = b_j)$, the marginal distribution of Y. We divide the entries in the column by this value to obtain a new distribution of X (whose probabilities add up to 1).

In particular, we have

$$E(X \mid Y = b_j) = \sum_i a_i P(X = a_i \mid Y = b_j).$$

Example I have two red pens, one green pen, and one blue pen, and I choose two pens without replacement. Let X be the number of red pens that I choose and Y the number of green pens. Then the joint p.m.f. of X and Y is given by the following table:

		$Y=0$	$Y=1$
X	0	0	$\frac{1}{6}$
	1	$\frac{1}{3}$	$\frac{1}{3}$
	2	$\frac{1}{6}$	0

In this case, the conditional distributions of X corresponding to the two values of Y are as follows:

a	0	1	2
$P(X = a \mid Y = 0)$	0	$\frac{2}{3}$	$\frac{1}{3}$

a	0	1	2
$P(X = a \mid Y = 1)$	$\frac{1}{3}$	$\frac{2}{3}$	0

We have

$$E(X \mid Y = 0) = \frac{4}{3}, \qquad E(X \mid Y = 1) = \frac{2}{3}.$$

If we know the conditional expectation of X for all values of Y, we can find the expected value of X:

Proposition 4.3 $E(X) = \sum_j E(X \mid Y = b_j) P(Y = b_j).$

Proof:

$$\begin{aligned} E(X) &= \sum_i a_i P(X = a_i) \\ &= \sum_i a_i \sum_j P(X = a_i \mid Y = b_j) P(Y = b_j) \\ &= \sum_j \left(\sum_i a_i P(X = a_i \mid Y = b_j \right) P(Y = b_j) \\ &= \sum_j E(X \mid Y = b_j) P(Y = b_j). \end{aligned}$$

Example Let us revisit the geometric random variable and calculate its expected value. Recall the situation: I have a coin with probability p of showing heads; I toss it repeatedly until heads appears for the first time; X is the number of tosses.

Let Y be the Bernoulli random variable whose value is 1 if the result of the first toss is heads, 0 if it is tails. If $Y = 1$, then we stop the experiment then and there; so if $Y = 1$, then necessarily $X = 1$, and we have $E(X \mid Y = 1) = 1$.

On the other hand, if $Y = 0$, then the sequence of tosses from that point on has the same distribution as the original experiment; so $E(X \mid Y = 0) = 1 + E(X)$ (the 1 counting the first toss). So

$$\begin{aligned} E(X) &= E(X \mid Y=0)P(Y=0) + E(X \mid Y=1)P(Y=1) \\ &= (1+E(X)) \cdot q + 1 \cdot p \\ &= E(X)(1-p) + 1; \end{aligned}$$

rearranging this equation, we find that $E(X) = 1/p$, confirming our earlier value.

Proposition 4.4 *Let X and Y be discrete random variables. Then X and Y are independent if and only if, for any values a_i and b_j of X and Y respectively, we have*

$$P(X = a_i \mid Y = b_j) = P(X = a_i).$$

Note that Proposition 4.4 holds only if for any b_j the probability $P(Y = b_j) > 0$

Q1. How to proof?

Q2 An urn has 2 red, 5 white, and 3 green balls. Select 3 balls at random without replacement and let X be the number of red balls and Y the number of white balls. Determine

(a) joint p. m. f. of (X, Y),
(b) marginal p. m. f.
(c) $P(X \geq Y)$
(d) $P(X = 2|X \geq Y)$.

4.3 Joint distribution of continuous r.v.s

Let X and Y be continuous random variables. The joint cumulative distribution function of X and Y is the function $F_{X,Y}$ of two real variables given by

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

We define X and Y to be independent if $P(X \le x, Y \le y) = P(X \le x) \cdot P(Y \le y)$, for any x and y, that is, $F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y)$.

(Note that, just as in the one variable case, X is part of the name of the function, while x is the argument of the function.)

The joint probability density function of X and Y is

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y).$$

The probability that the pair of values of (X, Y) corresponds to a point in some region of the plane is obtained by taking the double integral of $f_{X,Y}$ over that region. For example,

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f_{X,Y}(x,y)\mathrm{d}x\,\mathrm{d}y$$

The marginal p.d.f. of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}y,$$

and the marginal p.d.f. of Y is similarly

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}x.$$

Then the conditional p.d.f. of $X|Y$ is

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

联合密度除以边际密度

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x, y) \mathrm{d}x \, \mathrm{d}y,$$

$$\operatorname{Cov}(X, Y) = E(XY) - E(X)E(Y), \qquad \operatorname{corr}(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X) \operatorname{Var}(Y)}}.$$

The continuous random variables X and Y are independent if and only if

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y).$$

As usual this holds if and only if the *conditional* p.d.f. of $X|Y$ is equal to the *marginal* p.d.f. of X

Also, if X and Y are independent, then $Cov(X,Y) = corr(X,Y) = 0$ (but not conversely!)

4.4 Transformation of random variables

If a continuous random variable Y is a function of another r.v. X, we can find the distribution of Y in terms of that of X.

Example Let X and Y be random variables. Suppose that $X \sim U[0,4]$ (uniform on $[0,4]$) and $Y = \sqrt{X}$. What is the support of Y? Find the cumulative distribution function and the probability density function of Y.

Solution (a) The support of X is $[0,4]$, and $Y = \sqrt{X}$, so the support of Y is $[0,2]$.

(b) We have $\cancel{f_X}(x) = x/4$ for $0 \leq x \leq 4$. Now

Fx(x)

$$
\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(X \leq y^2) \\
&= F_X(y^2) \\
&= y^2/4
\end{aligned}
$$

for $0 \leq y \leq 2$; of course $F_Y(y) = 0$ for $y < 0$ and $F_Y(y) = 1$ for $y > 2$. (Note that $Y \leq y$ if and only if $X \leq y^2$, since $Y = \sqrt{X}$.)

(c) We have

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \begin{cases} y/2 & \text{if } 0 \leq y \leq 2, \\ 0 & \text{otherwise.} \end{cases}$$

The argument in (b) is the key. If we know Y as a function of X, say $Y = g(X)$, where g is an increasing function, then the event $Y \leq y$ is the same as the event $X \leq h(Y)$, where h is the *inverse function* of g. This means that $y = g(x)$ if and only if $x = h(y)$. (In our example, $g(x) = \sqrt{x}$, and so $h(y) = y^2$.) Thus

$$F_Y(y) = F_X(h(y)),$$

and so, by the Chain Rule,

$$f_Y(y) = f_X(h(y))h'(y),$$

where h' is the derivative of h. (This is because $f_X(x)$ is the derivative of $F_X(x)$ with respect to its argument x, and the Chain Rule says that if $x = h(y)$ we must multiply by $h'(y)$ to find the derivative with respect to y.)

Applying this formula in our example we have

$$f_Y(y) = \frac{1}{4} \cdot 2y = \frac{y}{2}$$

for $0 \leq y \leq 2$, since the p.d.f. of X is $f_X(x) = 1/4$ for $0 \leq x \leq 4$.

Here is a formal statement of the result.

Theorem 4.5 *Let X be a continuous random variable. Let g be a real function which is either strictly increasing or strictly decreasing on the support of X, and which is differentiable there. Let $Y = g(X)$. Then*

(a) the support of Y is the image of the support of X under g;

(b) the p.d.f. of Y is given by $f_Y(y) = f_X(h(y))|h'(y)|$, where h is the inverse function of g.

For example, here is the proof of Proposition 3.6: *if $X \sim N(\mu, \sigma^2)$ and $Y = (X - \mu)/\sigma$, then $Y \sim N(0, 1)$.* Recall that

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\, \mathrm{e}^{-(x-\mu)^2/2\sigma^2}.$$

We have $Y = g(X)$, where $g(x) = (x - \mu)/\sigma$; this function is everywhere strictly increasing (the graph is a straight line with slope $1/\sigma$), and the inverse function is $x = h(y) = \sigma y + \mu$. Thus, $h'(y) = \sigma$, and

$$f_Y(y) = f_X(\sigma y + \mu) \cdot \sigma = \frac{1}{\sqrt{2\pi}}\, \mathrm{e}^{-y^2/2},$$

the p.d.f. of a standard normal variable.

Please note: If the transforming function g is not monotonic (that is, not either increasing or decreasing), then life is a bit more complicated.

For example, if X is a random variable taking both positive and negative values, and $Y = X^2$, then a given value y of Y could arise from either of the values $\sqrt{y}$ and $-\sqrt{y}$ of X, so we must work out the two contributions and add them up.

Example $X \sim N(0,1)$ and $Y = X^2$. Find the p.d.f. of Y.

The p.d.f. of X is $(1/\sqrt{2\pi})\mathrm{e}^{-x^2/2}$. Let $\Phi(x)$ be its c.d.f., so that $P(X \leq x) = \Phi(x)$, and

$$\Phi'(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}.$$

Now $Y = X^2$, so $Y \leq y$ if and only if $-\sqrt{y} \leq X \leq \sqrt{y}$. Thus

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\
&= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) \\
&= \Phi(\sqrt{y} - (1 - \Phi(\sqrt{y})) \qquad \text{(by symmetry of } N(0,1)\text{)} \\
&= 2\Phi(\sqrt{y}) - 1.
\end{aligned}$$

So

$$\begin{aligned} f_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) \\ &= 2\Phi'(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} \qquad \text{(by the Chain Rule)} \\ &= \frac{1}{\sqrt{2\pi y}} \mathrm{e}^{-y/2}. \end{aligned}$$

Of course, this is valid for $y > 0$; for $y < 0$, the p.d.f. is zero.

Exercises

EX1 Two numbers X and Y are chosen independently from the uniform distribution on the unit interval [0,1]. Let Z be the maximum of the two numbers. Find the p.d.f. of Z, and hence find its expected value, variance and median.

EX2 I roll a fair die bearing the numbers 1 to 6. If N is the number showing on the die, I then toss a fair coin N times. Let X be the number of heads I obtain.

(a) Write down the p.m.f. for X.

(b) Calculate $E(X)$ without using this information.

EX3 Let

$$f(x,y) = \begin{cases} c\,x^2 y & \text{if } x^2 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Determine (a) the constant c, (b) $P(X \geq Y)$, (c) $P(X = Y)$, and (d) $P(X = 2Y)$, (e) Compute marginal densities and determine whether X and Y are independent.

EX4 Assume that you are waiting for two phone calls, from Alice and from Bob. The waiting time T_1 for Alice's call has expectation 10 minutes and the waiting time T_2 for Bob's call has expectation 40 minutes. Assume T_1 and T_2 are independent exponential random variables. What is the probability that Alice's call will come first?

EX5 The joint density of (X, Y) is given by

$$f(x,y) = \begin{cases} 3x & \text{if } 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Compute the conditional density of Y given $X = x$.

(b) Are X and Y independent?