Queen Mary
University of London

MAS 108 Probability I

Notes 12 Autumn 2005

Conditional random variables

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

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

$$P(X = x \mid B) = \frac{P(B \text{ holds and } X = x)}{P(B)}.$$

This defines the probability mass function of the *conditional random variable* $X \mid B$, called "X given B".

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

$$E(X \mid B) = \sum_x x P(X = x \mid B).$$

Example I throw two fair six-sided dice independently. Let X be the number showing on the first die, and let C be the event that at least one number showing is odd. In Notes 3 we saw that $P(C) = 3/4$.

If i is odd and $1 \le i \le 6$ then

$$P(X = i \mid C) = \frac{P(X = i \text{ and } C)}{P(C)} = \frac{P(X = i)}{P(C)} = \frac{1/6}{3/4} = \frac{2}{9}.$$

On the other hand, if i is even and $1 \le i \le 6$ then

$$P(X = i \mid C) = \frac{P(X = i \text{ and } C)}{P(C)} = \frac{P(X = i \text{ and second is odd})}{P(C)} = \frac{(1/6)\times(1/2)}{3/4} = \frac{1}{9}.$$

Therefore the p.m.f. for $X \mid C$ is

i	1	2	3	4	5	6
$P(X = i \mid C)$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

and $E(X \mid C) = 10/3$.

Example Suppose that I go fishing all day, and that the number X of fish that I catch in a day has the Poisson(λ) distribution. Supopose that you know that I have caught at least one fish. Then, if $m \neq 0$ we get

$$P(X = m \mid X \neq 0) = \frac{P(X = m)}{P(X \neq 0)} = \frac{\mathrm{e}^{-\lambda}\frac{\lambda^m}{m!}}{1 - \mathrm{e}^{-\lambda}} = \frac{\lambda^m}{m!(e^{\lambda} - 1)};$$

of course, the probability is 0 if $m = 0$.

Now the event B in $X \mid B$ might itself be defined by another random variable; for example, B might be the event that Y takes the value y. In this case, we have

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

In other words, we have taken the column of the joint p.m.f. table of X and Y corresponding to the value $Y = y$. The sum of the entries in this column is just $P(Y = y)$, one entry in 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).

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	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

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

In Theorem 1, we saw that independence of events can be characterised in terms of conditional probabilities: A and B are independent if and only if $P(A \mid B) = P(A)$. A similar result holds for independence of random variables.

Proposition Let X and Y be discrete random variables. Then X and Y are independent of each other if and only if, for all values x and y of X and Y respectively, we have

$$P(X = x \mid Y = y) = P(X = x).$$

This is obtained by applying Theorem 1 to the events $X = x$ and $Y = y$. It can be stated in the following way:

> X and Y are independent if and only if the the *conditional* p.m.f. of $X \mid (Y = y)$ is equal to the *marginal* p.m.f. of X, for every value y of Y.

In general, these distributions will *not* be the same!

If we know the *conditional expectation* of X for all values of Y, we can find the expected value of X. This is what the *conditional expectation theorem* tells us.

Theorem 11 (Theorem of Conditional Expectation) Let $A_1, \ldots, A_n$ be mutually exclusive events whose union is the whole sample space, with $P(A_i) > 0$ for $i = 1, \ldots, n$. If X is a random variable defined on the same sample space then

$$E(X) = \sum_i P(A_i) E(X \mid A_i).$$

Proof

$$\begin{aligned}
E(X) &= \sum_x x P(X = x) \\
&= \sum_x x \left[\sum_{i=1}^n P(A_i) P(X = x \mid A_i) \right] && \text{by the Theorem of Total Probability for each event “}X = x\text{”} \\
&= \sum_x \sum_{i=1}^n P(A_i) x P(X = x \mid A_i) \\
&= \sum_{i=1}^n \sum_x P(A_i) x P(X = x \mid A_i) && \text{rearranging the order of summation} \\
&= \sum_{i=1}^n P(A_i) \left[\sum_x x P(X = x \mid A_i) \right] \\
&= \sum_{i=1}^n P(A_i) E(X \mid A_i). \quad \blacksquare
\end{aligned}$$

In the above example, we have

$$E(X) = E(X \mid Y = 0) P(Y = 0) + E(X \mid Y = 1) P(Y = 1) = (4/3) \times (1/2) + (2/3) \times (1/2) = 1.$$

Example A bottle of liquid contains an unknown number of infectious units. The scientists want to know how many. This is how they try to find out. First, they mix the liquid thoroughly and divide it into 15 equal portions, called *aliquots*. An aliquot is said to be *infectious* if it contains any of the infectious units. Then five aliquots are chosen at random. Each chosen aliquot is injected into one animal. If the aliquot is infectious then the animal will become ill; otherwise the animal remains healthy. The scientists count how many of the five animals become ill. Let X be the number of ill animals.

Suppose that the original bottle of liquid contains exactly three germs (infectious units). When the liquid is mixed and divided into aliquots, each infectious unit is equally likely to end up in any of the fifteen aliquots, independently of all the other infectious units. Find $E(X)$.

Solution Let

$$\begin{aligned} A_1 &= \text{“exactly one aliquot is infectious”} \\ A_2 &= \text{“exactly two aliquots are infectious”} \\ A_3 &= \text{“exactly three aliquot are infectious”.} \end{aligned}$$

For each aliquot, the probabilty that all three germs end up in it is $(1/15)^3$. There are 15 aliquots, so

$$P(A_1) = 15 \times \left(\frac{1}{15}\right)^3 = \frac{1}{225}.$$

There are ${}^{15}\mathrm{C}_3$ ways of choosing three aliquots to be infected. For each of these ways there are 3! ways of matching the order of the germs to the order of the aliquots. Hence

$$P(A_3) = {}^{15}\mathrm{C}_3 \times 3! \times \left(\frac{1}{15}\right)^3 = \frac{182}{225}.$$

By subtraction, $P(A_2) = 42/225$.

Now, randomly choosing five aliquots from 15 when a certain number of the aliquots is infectious gives us a hypergeometric random variable for the number of infected animals. Thus $X \mid A_1 \sim \mathrm{Hg}(5,1,15)$ so $E(X \mid A_1) = 5/15 = 1/3$. Similarly, $X \mid A_2 \sim \mathrm{Hg}(5,2,15)$ so $E(X \mid A_2) = (5\times 2)/15 = 2/3$ and $X \mid A_3 \sim \mathrm{Hg}(5,3,15)$ so $E(X \mid A_3) = (5\times 3)/15 = 1$. Therefore

$$E(X) = \frac{1}{225}\times\frac{1}{3} + \frac{42}{225}\times\frac{2}{3} + \frac{182}{225}\times 1 = \frac{631}{675}.$$

Example Here is another example. Suppose that I roll a fair die with faces numbered from 1 to 6. If the number shown is n, I then toss a fair coin n times and count the number of heads. What is the expected value of the number of heads?

Let N be the random number shown on the die, and X the number of heads. It is possible to calculate the p.m.f. of X directly, though it is quite a lot of work. For example, what is the probability that $X = 5$? This can only occur if either $N = 5$ and we get heads five times (with probability $(1/6)\times(1/32)$), or if $N = 6$ and we get five out of six heads (with probability $(1/6)\times({}^6\mathrm{C}_5\times(1/64))$, total $1/48$. When we have done all these laborious calculations, we can find $E(X)$ directly.

However, the conditional expectation theorem makes it much easier. If the number on the die is n, then X is a binomial random variable $\mathrm{Bin}(n,1/2)$, with expected value $n/2$. That is, $E(X \mid N = n) = n/2$. By Theorem 11,

$$\begin{aligned} E(X) &= E(X \mid N=1)P(N=1)+\cdots+E(X \mid N=6)P(N=6) \\ &= \frac{1}{2}\cdot\frac{1}{6}+\cdots+\frac{6}{2}\cdot\frac{1}{6} \\ &= \frac{21}{2}\cdot\frac{1}{6} \\ &= \frac{7}{4}. \end{aligned}$$

Mean and variance of geometric

Let us revisit the geometric random variable and calculate its expected value and variance. 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. More generally, we continue with independent Bernoulli trials until the first success.

Let S be the event that the first trial is a success. If S occurs then we stop then and there; so $X = 1$, and we have $E(X \mid S) = 1$. On the other hand, if S does not occur then the sequence of trials from that point on has the same distribution as the original X; so $E(X \mid S') = 1+E(X)$ (the 1 counting the first trial). So

$$\begin{aligned} E(X) &= P(S)E(X \mid S)+P(S')E(X \mid S') \\ &= p\times 1+(1-p)\times(1+E(X)). \end{aligned}$$

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

Similarly,

$$E(X^2) = P(S)E(X^2 \mid S)+P(S')E(X^2 \mid S').$$

If S occurs then $X = 1$ so $X^2 = 1$. If not, put $Y = X - 1$, so that $Y \sim \text{Geom}(p)$. Then

$$E(X^2 \mid S') = E((Y+1)^2) = E(Y^2) + 2E(Y) + 1 = E(X^2) + \frac{2}{p} + 1.$$

Substitution gives

$$E(X^2) = p \times 1 + (1-p) \times \left(E(X^2) + \frac{2}{p} + 1 \right),$$

which can be rearranged to give $E(X^2) = (2-p)/p^2$. Then

$$\text{Var}(X) = E(X^2) - (E(X))^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2},$$

as we found before.

Two continuous random variables

If X and Y are continuous random variable defined on the same sample space, they have a *joint probability density function* $f_{X,Y}(x,y)$ such that

$$P\left((X,Y) \in \text{some region } A \text{ of } \mathbb{R}^2\right) = \iint_{(x,y)\in A} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

Since probabilities are positive, we have $f_{X,Y}(x,y) \geq 0$ for all real x and y. Putting $A = \mathbb{R}^2$ gives

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

It can also be shown that if δx and δy are small then

$$P(x \leq X \leq x + \delta x \text{ and } y \leq Y \leq y + \delta y) \approx f_{X,Y}(x,y)\,\delta x\,\delta y.$$

The *joint cumulative distribution function* $F_{X,Y}$ is given by

$$F_{X,Y}(x,y) = P(X \leq x \text{ and } Y \leq y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}(t,u)\,\mathrm{d}t\,\mathrm{d}u.$$

The *marginal probability density functions* are obtained by integrating over the other variable:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y; \\ f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x. \end{aligned}$$

If g is a real function of two variables and if X and Y have a joint continuous distribution then

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

Theorem 7 ($E(X+Y) = E(X)+E(Y)$) still holds; the proof uses integration instead of summation.

For continuous random variables, $P(X=x) = P(Y=y) = 0$, so we have to adapt the definition of independence. We say that X and Y are *independent* of each other if $f_{X,Y}(x,y) = f_X(x)\,f_Y(y)$ for all real x and y.

If X and Y are independent of each other then

$$\begin{aligned}
P(x_1 \le X \le x_1 \text{ and } y_1 \le Y \le y_2) &= \int_{y_1}^{y_2}\int_{x_1}^{x_2} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{y_1}^{y_2}\int_{x_1}^{x_2} f_X(x)\,f_Y(y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{y_1}^{y_2} f_Y(y)\left[\int_{x_1}^{x_2} f_X(x)\,\mathrm{d}x\right]\mathrm{d}y \\
&= \int_{y_1}^{y_2} f_Y(y) P(x_1 \le X \le x_2)\,\mathrm{d}y \\
&= P(x_1 \le X \le x_2)\int_{y_1}^{y_2} f_Y(y)\,\mathrm{d}y \\
&= P(x_1 \le X \le x_2) P(y_1 \le Y \le y_2)
\end{aligned}$$

so the events "$x_1 \le X \le x_2$" and "$y_1 \le Y \le y_2$" are independent of each other.

The following hold for joint continous random variables as well as for joint discrete random variables.

Theorem 8 (X and Y independent $\Longrightarrow E(XY) = E(X)E(Y)$) The proof uses integration instead of summation.

Definition of covariance This uses just E, so is unchanged.

Theorem 9 (properties of covariance) Unchanged.

Definition and propeties of correlation Unchanged.

Theorem 10 and its corollary (variance of $aX+bY$) Unchanged.

Theorem 11 (Theorem of Conditional Expectation) The proof uses integration instead of summation.

The definition of conditional random variables like $X \mid Y = y$ must be different, because $P(Y = y) = 0$. If X and Y have a joint continuous distirbution then we define the conditional random variable $X \mid Y = y$ to have probability density function

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

so long as $f_Y(y) \neq 0$. Similarly, if $f_X(x) \neq 0$ then the conditional random variable $Y \mid X = x$ is defined to have probability density function

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

If X and Y are independent of each other then the conditional distributions are equal to the marginal distributions.