Probability I

Lecture 9

Davide La Vecchia

Spring Semester 2020

Joint Probability Functions

- Let X and Y be a pair of discrete random variables
- Their **joint probability mass function** (joint PMF) expresses the probability that simultaneously X takes on the specific value x and Y takes on the specific value y.
- It is denoted by

$$p_{X,Y}(x,y) = P(\{X = x \cap Y = y\})$$

  thought of as a function of x and y.
- The joint PMF has two essential properties:
  1. $p_{X,Y}(x,y) \geq 0$ for all possible pairs (x,y) (its value is always non-negative)
  2. $\sum_x \sum_y P(\{X = x \cap Y = y\}) = 1$ (its sum over all combinations of x and y values is equal to one)

Marginal probability (mass) functions

Definition

The probability (mass) function of the *discrete* random variable X is called its marginal probability (mass) function. It is obtained by summing the joint probabilities relating to pairs (X, Y) over all possible values of Y:

$$p_X(x) = \sum_y p_{X,Y}(x, y).$$

Similarly, the probability (mass) function of the *discrete* random variable Y is called its marginal probability (mass) function. It is obtained by summing the joint probabilities relating to pairs (X, Y) over all possible values of X:

$$p_Y(y) = \sum_x p_{X,Y}(x, y).$$

First Example

Example (caplets: this probability course is giving me headache)

- Two caplets are selected at random from a bottle containing three aspirins, two sedatives and two placebo caplets. We are assuming that the caplets are well mixed and that each has an equal chance of being selected.
- Let X and Y denote, respectively, the numbers of aspirin caplets, and the number of sedative caplets, included among the two caplets drawn from the bottle.

First Example

Example (cont'd)

Tabulating the *joint* probabilities as follows, we can easily work out the *marginal* probabilities

$y \backslash x$	0	1	2	$P\{Y = y\}$
0	1/21	6/21	3/21	10/21
1	4/21	6/21	0	10/21
2	1/21	0	0	1/21
$P\{X = x\}$	6/21	12/21	3/21	1

Empirical Example

Example

- Two production lines manufacture a certain type of item.
- Suppose that the capacity (on any given day) is 5 items for *Line I* and 3 items for *Line II*.
- Assume that the number of items actually produced by either production line varies from one day to the next.
- Let (X, Y) represent the 2-dimensional random variable yielding the number of items produced by *Line I* and *Line II*, respectively, on any one day.
- In practical applications of this type the joint probability (mass) function $P(\{X = x \cap Y = y\})$ is unknown more often than not!

Empirical Example

Example

- The joint probability (mass) function $P(\{X = x \cap Y = y\})$ for all possible values of x and y can be approximated however.
- By the observing the long-run relative frequency with which different numbers of items are actually produced by either production line.

$y \backslash x$	0	1	2	3	4	5	$P\{Y = y\}$
0	0	0.01	0.03	0.05	0.07	0.09	0.25
1	0.01	0.02	0.04	0.05	0.06	0.08	0.26
2	0.01	0.03	0.05	0.05	0.05	0.06	0.25
3	0.01	0.02	0.04	0.06	0.06	0.05	0.24
$P\{X = x\}$	0.03	0.08	0.16	0.21	0.24	0.28	1

- e.g. $P(\{X = 5 \cap Y = 0\}) \approx 0.09 = \frac{\#\{X=5 \cap Y=0\}\text{days}}{\#\text{days}}$

Conditional probability mass function

Recall that the *conditional* probability mass function of the *discrete* random variable Y, *given* that the random variable X takes the value x, is given by

$$p_{Y|X}(y|x) = \frac{P\{X = x \cap Y = y\}}{P_X(X = x)}$$

Note this is a probability mass function for y, with x viewed as fixed. Similarly,

Definition

The *conditional* probability mass function of the *discrete* random variable X, *given* that the random variable Y takes the value y, is given by

$$p_{X|Y}(x|y) = \frac{P\{X = x \cap Y = y\}}{P_Y(Y = y)}$$

Note this is a probability mass function for x, with y viewed as fixed.

Independence

- Two random variables X and Y are **independent** if

$$p_{X,Y}(x,y) = p_X(x)p_Y(y) \qquad \text{(discrete)}$$

for *all* values of x and y.

- Note that independence also implies that

$$p_{X|Y}(x|y) = p_X(x) \text{ and } p_{Y|X}(y|x) = p_Y(y) \qquad \text{(discrete)}$$

for *all* values of x and y.

Expectations for Jointly Distributed Discrete RVs

Example

Example. Suppose that (X, Y) is a bivariate discrete random variable such that the point $(1, 2)$ occurs with probability 1/8, $(1, 3)$ with probability 3/8, $(2, 3)$ with probability 1/4, and $(3, 1)$ with probability 1/4. Then (X, Y) assumes as values only one of these for points.

	$Y = 1$	$Y = 2$	$Y = 3$	marginal of X
$X = 1$	0	1/8	3/8	1/2
$X = 2$	0	0	1/4	1/4
$X = 3$	1/4	0	0	1/4
marginal of Y	1/4	1/8	5/8	1

Note that, similarly to the univariate case, (i) all the probabilities must be non-negative and (ii) $\sum_{\mathrm{x} \in \mathbb{R}} P[\mathrm{X} = \mathrm{x}] = 1$ (for both marginal and joint probabilities).

Expectations for Jointly Distributed Discrete RVs

Example

Example. Continued. We compute the conditional probability function of Y given $X = 1$. Note that $P[Y = y \mid X = 1] = 0$ except for $y = 2, 3$. Thus,

$$P[Y = 2 \mid X = 1] = \frac{P[X = 1, Y = 2]}{P[X = 1]} = \frac{1/8}{1/2} = 1/4;$$
$$P[Y = 3 \mid X = 1] = \frac{P[X = 1, Y = 3]}{P[X = 1]} = \frac{3/8}{1/2} = 3/4.$$

Note that once again $\sum_y P[Y = y \mid X = 1] = 1$.

Expectations for Jointly Distributed Discrete RVs

Definition

Let $h(x, y)$ be a function of x and y. We define the **expected value** of $h(X, Y)$ as

$$E[h(X, Y)] = \sum_y \sum_x h(x, y) \, p_{X,Y}(x, y)$$

Example

Example. Let X and Y be two discrete random variables with joint probability function

	$Y=0$	$Y=1$	$Y=2$	$Y=3$	marginal of X
$X=0$	h	$2h$	$3h$	$4h$	$10h$
$X=1$	$4h$	$6h$	$8h$	$2h$	$20h$
$X=2$	$9h$	$12h$	$3h$	$6h$	$30h$
marginal of Y	$14h$	$20h$	$14h$	$12h$	$\sum_{(x,y)} = 60h$

Hence, $h = 1/60$. We compute all moments up to order 2:

$$E[X] = \sum_x x p_X(x) = 0 \cdot 10h + 1 \cdot 20h + 2 \cdot 30h = 80h = 4/3;$$

$$E[Y] = \sum_y y p_Y(y) = 0 \cdot 14h + 1 \cdot 20h + 2 \cdot 14h + 3 \cdot 12h = 84h = 7/5;$$

$$E[X^2] = \sum_x x^2 p_X(x) = 0^2 \cdot 10h + 1^2 \cdot 20h + 2^2 \cdot 30h = 140h = 7/3;$$

$$E[Y^2] = \sum_y y^2 p_Y(y) = 0^2 \cdot 14h + 1^2 \cdot 20h + 2^2 \cdot 14h + 3^2 \cdot 12h = 184h = 46/15;$$

$$E[XY] = \sum_{(x,y)} xy p_{(X,Y)}(x,y) = 5/3.$$

Thus $E[X] = 4/3, E[Y] = 7/5, Var(X) = 7/3 - (4/3)^2 = 5/9, Var(Y) = 46/15 - (7/5)^2 = 83/75$ and $Cov(X,Y) = 5/3 - 4/3 \cdot 7/5 = -1/5, \rho(X,Y) = \frac{-1/5}{\sqrt{(5/9)(83/75)}} = -0.255.$

Expectations for Jointly Distributed Discrete RVs

Definition

The **conditional expectation** of $h(X, Y)$ *given* $Y = y$ is defined as

$$E\left[h(X, Y) \,|y\right] = \sum_{x} h(x, y)\, p_{X|Y}(x|y).$$

The **conditional expectation** of $h(X, Y)$ *given* $X = x$ is defined as

$$E\left[h(X, Y) \,|x\right] = \sum_{y} h(x, y)\, p_{Y|X}(y|x).$$

Expectations for Jointly Distributed Discrete RVs

Example (continuing the example at page 10)

30. First, we compute the conditional mean of Y given that $X = 1$:

$$E[Y \mid X = 1] = 2 \cdot 1/4 + 3 \cdot 3/4 = 11/4.$$

In addition, we compute the conditional mean of X given that $Y = 3$. The conditional distribution of X given $Y = 3$ is

$$p_{X|Y=3}(1) = 3/5; p_{X|Y=3}(2) = 2/5; p_{X|Y=3}(3) = 0.$$

Thus $E[X \mid Y = 3] = 1 \cdot 3/5 + 2 \cdot 2/5 = 7/5$.

Iterated Expectations

Definition

The *law of the iterated expectation* is often stated in the form

$$E[h(X,Y)] = E[E[h(X,Y)|Y]] = E[E[h(X,Y)|X]].$$

- This notation emphasises that whenever we write down $E[\cdot]$ for an expectation we are taking that expectation with respect to the distribution implicit in the formulation of the argument.
- The above formula is perhaps more easily understood using the more explicit notation

$$E_{(X,Y)}[h(X,Y)] = E_{(Y)}[E_{(X|Y)}[h(X,Y)]] = E_{(X)}[E_{(Y|X)}[h(X,Y)]].$$

- The latter notation makes it clear what distribution is being used to evaluate the expectation, the joint, the marginal or the conditional.

Expectations for Jointly Distributed Discrete RVs

Definition

Let X and Y be two discrete random variables. The **covariance** between X and Y is given by $E\left[h\left(X,Y\right)\right]$ when

$$h\left(X,Y\right)=\left(X-E\left[X\right]\right)\left(Y-E\left[Y\right]\right),$$

i.e. $Cov\left(X,Y\right)=E\left[\left(X-E\left[X\right]\right)\left(Y-E\left[Y\right]\right)\right]$

Alternative formula[1] for $Cov(X,Y)$ is

$$\boxed{Cov\left(X,Y\right)=E\left[XY\right]-E\left[X\right]E\left[Y\right]\ .} \tag{1}$$

[1] To get it, expand

$$\left(X-E\left[X\right]\right)\left(Y-E\left[Y\right]\right)=XY-E\left[X\right]Y-XE\left[Y\right]+E\left[X\right]E\left[Y\right]$$

and make use of the properties of expectation.

Expectations for Jointly Distributed Discrete RVs

So, to compute the covariance from a table describing the joint behaviour of X and Y, you have to:

- compute the joint expectation $E[XY]$—you get it making use of the joint probability;
- compute $E[X]$ and $E[Y]$—you get using the marginal probability for X and Y;
- combine these expected values as in formula (1).

See example on page 13 for an illustrative computation.

Some Properties of Covariances

- The Cauchy-Schwartz Inequality states

$$(E\left[XY\right])^2 \leq E\left[X^2\right] E\left[Y^2\right],$$

with equality if, and only if, $P(Y = cX) = 1$ for some constant c.

- Let $h(a) = E[(Y - aX)^2]$ where a is any number. Then

$$0 \leq h(a) = E[(Y - aX)^2] = E[X^2]a^2 - 2E[XY]a + E[Y^2].$$

This is a quadratic in a, and

- if $h(a) > 0$ the roots are real and $4(E[XY])^2 - 4E[X^2]E[Y^2] < 0$,
- if $h(a) = 0$ for some $a = c$ then $E[(Y - cX)^2] = 0$, which implies that $P(Y - cX = 0) = 1$.

Some Properties of Covariances

Remark

If two random variables are independent, their covariance is equal to zero. Note that the converse is not necessarily true: a zero covariance between two random variables does not imply that the variables are independent. This asymmetry[a] *follows because* **the covariance is a 'measure' of linear dependence.**

[a] Independence $\Rightarrow Cov(X, Y) = 0$ but $Cov(X, Y) = 0 \not\Rightarrow$ independence.

Example

Let us consider two discrete random variable X and Y, such that

$$P(\{X = 0\}) = P(\{X = 1\}) = P(\{X = -1\}) = \frac{1}{3},$$

while $Y = 0$ if $X \neq 0$ and $Y = 1$, if $X = 0$. So we have $E[X] = 0$ and $XY = 0$. This implies

$$Cov(X, Y) = E[XY] - E[X]E[Y] = 0,$$

although X and Y are NOT independent: they are related in a nonlinear way.

Some Properties of Covariances

Building on this remark, we have

- $Cov(X, Y) > 0$ if
  - large values of X tend to be *linearly* associated with large values of Y
  - small values of X tend to be *linearly* associated with small values of Y
- $Cov(X, Y) < 0$ if
  - large values of X tend to be *linearly* associated with small values of Y
  - small values of X tend to be *linearly* associated with large values of Y
- When $Cov(X, Y) = 0$, X and Y are said to be uncorrelated.

Some Properties of Covariances

- If X and Y are two random variables (either discrete or continuous) with $Cov(X, Y) \neq 0$, then:

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

  <u>Compare this expression with the formula on page 25, Lecture 3-4,</u> where we read that in the case of independent random variables X and Y we have

$$Var(X + Y) = Var(X) + Var(Y),$$

  which trivially follows from (2)—indeed, for independent random variables, $Cov(X, Y) \equiv 0$.

- The covariance depends upon the unit of measurement.

A remark

- If we scale X and Y, the covariance changes: For $a, b > 0$

$$Cov\,(aX, bY) = abCov\,(X, Y)$$

Thus, we introduce the **correlation** between X and Y is

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

which *does not* depend upon the unit of measurement.

An important property of correlation

Remark

The Cauchy-Schwartz Inequality implies that

$$-1 \leq corr\,(X, Y) \leq 1$$

The correlation is typically denoted by the Greek letter ρ, so we have

$$\rho(X, Y) = corr\,(X, Y)\,.$$