Probability 1

Lecture 10: Bivariate Discrete Random Variables

Dr. Daniel Flores-Agreda,

(based on the notes of Prof. Davide La Vecchia)

Spring Semester 2021

Objectives

- Introduce the Notion of Joint PMF's for discrete RV and the main implications
    - Difference between Conditional and Marginal PMF
    - Joint and Conditional Expectations
    - Covariance and Correlation

Outline

1 Joint Probability Functions

2 Conditional Probability

3 Expectations

4 Covariance and Correlation

Joint Probability Functions

Definition (Joint Probability Mass Function)

Let X and Y be a pair of discrete random variables

Their **Joint Probability Mass Function (Joint PMF)** expresses the probability that X takes on a specific value x and Y takes on the specific value y **simultaneously**.

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

thought of as a **function** of x and y.

Joint Probability Functions

The joint PMF has two essential properties:

1. The value of the Joint PMF is always non-negative

$$p_{X,Y}(x, y) \geq 0 \text{ for all possible pairs } \quad (x, y)$$

2. The sum over all combinations of x and y values is equal to one

$$\sum_x \sum_y P(\{X = x \cap Y = y\}) = 1$$

Joint Probability Functions

Definition (Marginal Probability Mass Functions)

The Probability Mass Function (PMF) of the *discrete* random variable X is called its *Marginal* PMF.
It can be obtained by summing the joint probabilities relating to pairs (X, Y) over **all possible values of** Y:

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

Equivalently, the Marginal PMF of Y can be obtained by summing the joint probabilities relating to pairs (X, Y) over **all possible values of** X:

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

Joint Probability Functions

Example (Caplets)

- Two caplets are selected at random from a bottle containing **3 aspirins**, **2 sedatives** and **2 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:
  - X : the numbers of aspirin caplets,
  - Y : the number of sedative caplets,

  that we pick when we draw when we draw two caplets from the bottle.

Joint Probability Functions

Example (Caplets)

Number of sets of 2 caplets out of 7:

$$\binom{7}{2} = \frac{7!}{2! \times (7-2)!} = \frac{7!}{2! \times 5!} = \frac{6 \times 7}{2} = 3 \times 7 = 21$$

Joint Probability Functions

Example (continued)

The Joint Probabilities can be found with Combinatorial Formulae:

- $p_{X,Y}(0,0) = \binom{3}{0}\binom{2}{0}\binom{2}{2} / 21 = 1/21$
- $p_{X,Y}(1,0) = \binom{3}{1}\binom{2}{0}\binom{2}{1} / 21 = 6/21$
- $p_{X,Y}(2,0) = \binom{3}{2}\binom{2}{0}\binom{2}{0} / 21 = 3/21$
- $p_{X,Y}(0,1) = \binom{3}{0}\binom{2}{1}\binom{2}{1} / 21 = 4/21$
- $p_{X,Y}(1,1) = \binom{3}{1}\binom{2}{1}\binom{2}{0} / 21 = 6/21$
- $p_{X,Y}(2,1) = 0$ since $2+1=3>2$
- $p_{X,Y}(0,2) = \binom{3}{0}\binom{2}{2}\binom{2}{0} / 21 = 1/21$
- $p_{X,Y}(1,2) = 0$ since $1+2=3>2$
- $p_{X,Y}(2,2) = 0$ since $2+2=4>2$

Joint Probability Functions

Example

Notice that we can infer the following Joint PMF:

$$P(X = x \cap Y = y) = p(x, y) = \begin{cases} \binom{3}{x}\binom{2}{y}\binom{2}{2-x-y} \Big/ \binom{7}{2} & x + y \leq 2 \\ 0 & \textit{otherwise} \end{cases}$$

Example (continued)

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

y \ 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

Example (Empirical 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!

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

Outline

1 Joint Probability Functions

2 Conditional Probability

3 Expectations

4 Covariance and Correlation

Conditional Probability

Definition

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

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

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

Similarly, the *conditional* PMF of Y, *given* $X = x$ is given by:

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

Again, **this is a PMF for** y, with x viewed as **fixed**.

- Two discrete random variables X and Y are **independent** if the **joint PMF is the product of the marginals**, i.e.

$$p_{X,Y}(x,y) \quad = \quad p_X(x)p_Y(y)$$

  for *all* values of x and y.

- Note that independence also implies that

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

  for *all* values of x and y.

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) $\sum_{\mathbf{x} \in \mathbb{R}} P[\mathbf{X} = \mathbf{x}] = 1$ (for both marginal and joint probabilities).

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 $\sum_y P[Y = y \mid X = 1] = 1$.

Outline

1 Joint Probability Functions

2 Conditional Probability

3 Expectations

4 Covariance and Correlation

Definition (Expectation of a Function of Two discrete RV)

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

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

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 $Cov(X,Y) = 5/3 - 4/3 \cdot 7/5 = -1/5, \rho(X,Y) = \frac{-1/5}{\sqrt{(5/9)(83/75)}} = -0.255.$

Definition (Conditional Expectation)

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

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

Equivalently, the **conditional expectation** of $h(X, Y)$ *given* $X = x$ is:

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

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

Definition (Law of the iterated expectation)

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 when we write down $E[\cdot]$ we take the expectation with respect to the **distribution implied by the argument**.
- Let's be more explicit:

$$\begin{aligned} E_{(X,Y)}[h(X, Y)] &= E_Y\left[E_{X|Y}[h(X, Y)|Y]\right] \\ &= E_X[E_{Y|X}[h(X, Y)|X]]. \end{aligned}$$

Outline

1. Joint Probability Functions
2. Conditional Probability
3. Expectations
4. Covariance and Correlation

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.

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).

Remark

If two random variables are independent, their covariance is equal to zero. Note that the converse is not necessarily true, ie. zero covariance between

two random variables does not imply that the variables are independent. This asymmetry follows because the covariance is a 'measure' of linear dependence.

$$Independence \Rightarrow Cov(X, Y) = 0 \text{ } but \text{ } Cov(X, Y) = 0 \nRightarrow independence.$$

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.

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.

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

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

  Compare this expression with the formula on Lecture 3-4, i.e. in the case of independent rv 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.

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

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

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

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

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

Remark

The Cauchy-Schwartz Inequality implies that

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

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

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

Wrap-up

- Two Random Variables can be Jointly distributed
- This implies Marginal and Conditional Distributions
- Expectations can also be conditional and can be calculated
- Covariance and Correlation are measures of *linear* association

Thank You for your Attention!

“See you” Next Week