Probability & Data
Week 4 – Two Random Variables

DR NICK FEWSTER-YOUNG

Bivariate Distributions

We have explored discrete probability distributions of one random variable, say *X*. In this section, we'll extend many of the definitions and concepts in which we have two random variables, say *X* and *Y*. More specifically, we will:

Topics

- extend the definition of a probability distribution of one random variable to the **joint probability distribution** of two random variables;
- learn how to use the **correlation coefficient** as a way of quantifying the extent that two random variables are linearly related;
- extend the definition of the conditional probability of events in order to find the **conditional probability distribution** of a random variable *X* given that *Y* has occurred.

Introduction

- We will learn how to extend the concept of a probability distribution of one random variable X to a joint probability distribution of two random variables X and Y.

- We will stay in the mind set of discrete random variables, X and Y may both be discrete random variables. For example, suppose X denotes whether a plane is late and Y denotes whether it rains. We might want to know if there is a relationship between X and Y. Or, we might want to know the probability that X takes on a particular value x and Y takes on a particular value y. That is, we might want to know $P(X = x , Y = y)$.

In more depth…

- Formal definition of a joint probability mass function of two discrete random variables.
- Finding a marginal probability mass function of a discrete random variable X from the joint probability mass function of X and Y.
- Independence of two random variables X and Y.
- Expectation of a function of the discrete random variables X and Y using their joint probability mass function.
- Variances of the discrete random variables X and Y using their joint probability mass function.
- Real world problems and applications in R.

Example

Suppose we toss a pair of fair, four-sided dice, in which one of the dice is **RED** and the other is **BLACK**. We'll let:

X = the outcome on the **RED** die = {1, 2, 3, or 4}

Y = the outcome on the **BLACK** die = {1, 2, 3, or 4}

What is the probability that X takes on a particular value x, and Y takes on a particular value y? That is, what is $P(X = x , Y = y)$?

$\boldsymbol{P_{X,Y}}$	$Y = 1$	$Y = 2$	$Y = 3$	$Y = 4$
$X = 1$	$^1/_{16}$	$^1/_{16}$	$^1/_{16}$	$^1/_{16}$
$X = 2$	$^1/_{16}$	$^1/_{16}$	$^1/_{16}$	$^1/_{16}$
$X = 3$	$^1/_{16}$	$^1/_{16}$	$^1/_{16}$	$^1/_{16}$
$X = 4$	$^1/_{16}$	$^1/_{16}$	$^1/_{16}$	$^1/_{16}$

Example (Flights and Rains)

The next example are our flights, consider the case in which the two random variables, *X* and *Y* are both discrete. Looking at the whether planes and the weather example, let's slowly extend the definitions for one discrete random variable, such as the probability mass function, mean and variance, to the case in which we have two discrete random variables. So $P_{X,Y}$ to start ☺

$$\mathrm{X} = \begin{cases} 1, & \textit{if the plane is late} \\ 0, & \textit{if the plane is ontime} \end{cases}, \quad \mathrm{Y} = \begin{cases} 1, & \textit{if it rains} \\ 0, & \textit{if does not rain} \end{cases}.$$

These random variables create all the possible outcomes by $X \times Y$.

$\boldsymbol{p_{X,Y}}$	$Y = 0$	$Y = 1$
$X = 0$	$\frac{14}{20}$	$\frac{1}{20}$
$X = 1$	$\frac{2}{20}$	$\frac{3}{20}$

Probability Mass

Definition. Let X and Y be two discrete random variables, and let S denote the two-dimensional support of X and Y. Then, the function $f(x, y) = P(X = x\ ,\ Y = y)$ is a **joint probability mass function** (abbreviated p.m.f.) if it satisfies the following three conditions:

$$0 \le f(x, y) \le 1$$

$$\sum_{(x,y)\in S}\sum f(x, y) = 1$$

$$P(\ (X, Y) \in A) = \sum_{(x,y)\in A}\sum f(x, y) = 1$$

where A is a subset of the space S.

Marginal Probability

Definition. Let X be a discrete random variable with support S_X, and let Y be a discrete random variable with support S_Y. Let X and Y have the joint probability mass function $f(x,y)$ with support S. Then, the probability mass function of X alone, which is called the **marginal probability mass function of *X***, is defined by:

$$P(X = x) = f_X(x) = \sum_{y \in S_Y} f(x, y), \;\; x \in S_X$$

where, for each x in the support S_X, the summation is taken over all possible values of y. Similarly, the probability mass function of Y alone, which is called the **marginal probability mass function of *Y***, is defined by:

$$P(Y = y) = f_Y(y) = \sum_{x \in S_X} f(x, y), \;\; y \in S_Y$$

where, for each y in the support S_Y, the summation is taken over all possible values of x.

Example

Example: Calculating the marginal probabilities for flights and rain. Recall the probability mass distribution.

$p_{X,Y}$	Y=0	Y=1
$X = 0$	$\frac{14}{20}$	$\frac{1}{20}$
$X = 1$	$\frac{2}{20}$	$\frac{3}{20}$

Solution: We want to calculate the marginal probability, if say Y is fixed then what is

$P(X = x) = f_X(x) :=$

$X = 0$	$X = 1$
$3/4$	$1/4$

If X is fixed then we have

$P(Y = y) = f_Y(y) :=$

$Y = 0$	$Y = 1$
$4/5$	$1/5$

Conditional Probability Mass Function

Definition: The conditional probability mass function of Y given X, where X, Y are discrete random variables defined on the same probability space, is given by

$$p_{Y|X}(y|x) = P(Y = y \,|X = x)$$
$$= \frac{p_{X,Y}(x,y)}{p_X(x)}, \qquad\qquad if\ p_X(x) > 0.$$

Let's now check out that flight and rain example!

Example

Example: Calculate the conditional probabilities for flights and rain. Recall the probability mass distribution.

$p_{X,Y}$	Y=0	Y=1
$X = 0$	$\frac{14}{20}$	$\frac{1}{20}$
$X = 1$	$\frac{2}{20}$	$\frac{3}{20}$

Solution: We want to calculate the conditional probability of x given that $\mathrm{Y} = y$ has occurred;

$$p_{X|Y}(x|y) = P(X = x\,|Y = y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} =$$

$Y = 0$	$Y = 1$
7/8	1/4
1/8	3/4

If X is fixed then we have

$$p_{Y|X}(y|x) = P(Y = y|X = x) = \frac{p_{X,Y}(x,y)}{p_X(x)} =$$

$X = 0$	$X = 1$
14/15	2/5
1/15	3/5

Independence

Definition. The random variables X and Y are **independent** if and only if:

$$P(X = x, Y = y) = P(X = x) \times P(Y = y)$$

for all $x \in S_X$, $y \in S_Y$. Otherwise, X and Y are said to be **dependent**.

Example: From the flights and rain example, if we multiple the marginal probabilities together and check the definition then

$$\begin{aligned} P(X = 0, Y = 0) = p(0,0) &= \frac{14}{20} \\ &\neq \frac{3}{5} \\ &= p_X(0) \times p_Y(0) = P(X = 0) \times P(Y = 0) \end{aligned}$$

Therefore, X and Y are dependent.

Expectation

Definition. Let X be a discrete random variable with support S_X, and let Y be a discrete random variable with support S_Y. Let X and Y be discrete random variables with joint p.m.f. $f(x,y)$ on the space S. If $u(X,Y)$ is a function of these two random variables, then:

$$E\big(u(X,Y)\big) = \sum_{(x,y)\in S}\sum u(x,y)f(x,y)$$

if it exists, is called the **expected value** of $u(X,Y)$. If $u(X,Y) = X$, then:

$$E(x) = \sum_{x\in S_X} xf(x,y)$$

if it exists, is the **mean of *X***.

Expectation Results

- (Linearity of expectation). For any real valued constant a, any function g and any discrete (or continuous) random variable X

$$E\big(ag(X)\big) = aE(g(X));$$

For any real valued constants a, b, any functions f, g and any discrete (or continuous) random variables X and Y

$$E\big(a\,f(X,Y) + b\,g(X,Y)\big) = a\,E\big(g(X,Y)\big) + b\,E\big(f(X,Y)\big).$$

- (Expectation of functions of independent random variables). If X, Y are **independent** random variables defined on the same probability space, and g, h are real-valued functions, then

$$E\big(g(X)h(Y)\big) = E\big(g(X)\big)E\big(h(Y)\big)$$

Covariance of random variables

Definition. The covariance of two random variables, X, Y is defined as

$$Cov(X,Y) := E\Big(\big(X - E(X)\big)\big(Y - E(Y)\big)\Big) = E(XY) - E(X)E(Y)$$

If $Cov(X,Y) = 0$, then X, Y are uncorrelated (no linear association).

- Joining this up with variance, if $X = Y$ then this gives the variance of X, σ_X. Therefore, we define the variance of bivariate distributions as

$$\sigma_{X,Y} \coloneqq Cov(X,Y).$$

Correlation

Previously, we introduced the joint probability distribution of two random variables *X* and *Y.* Now, we extend our investigation of the relationship between two random variables by how to quantify the *extent* or *degree* to which two random variables *X* and *Y* are associated or **correlated**.

For example, suppose *X* denotes the number of cups of hot chocolate sold daily at a local café, and *Y* denotes the number of apple cinnamon muffins sold daily at the same café. Then, the manager of the café might benefit from knowing whether *X* and *Y* are highly correlated or not. If the random variables are highly correlated, then the manager would know to make sure that both are available on a given day.

If the random variables are not highly correlated, then the manager would know that it would be okay to have one of the items available without the other. A statistical `measure of the strength is known as the **correlation coefficient** which allows us to quantify the degree of correlation between two random variables *X* and *Y*.

Question?

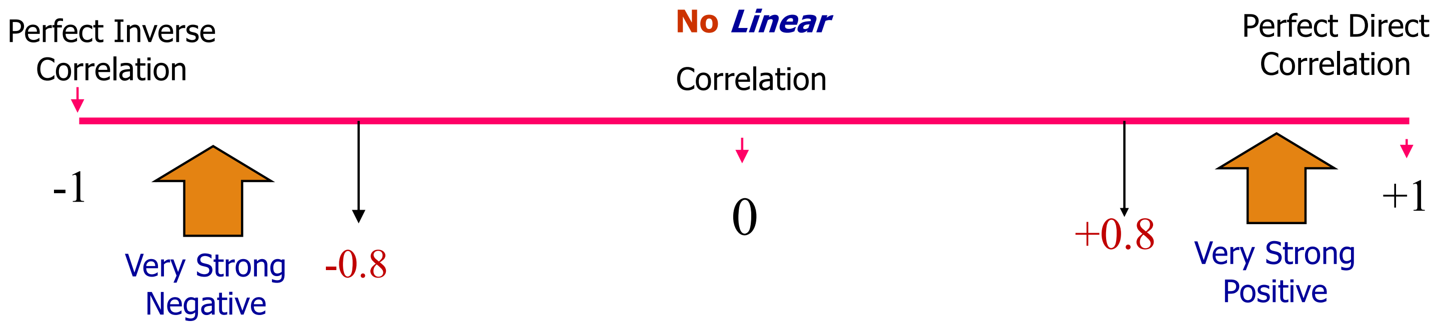
Similarly, we can go back to our example on “flights and rain”, we can ask the question, what is the strength of the correlation between the two variables?

The Correlation coefficient

Definition. Let *X* and *Y* be any two random variables (discrete or continuous) with standard deviations σ_X and σ_Y, respectively. The **correlation coefficient** of *X* and *Y*, denoted **Corr(*X*,*Y*)**

$$\rho_{X,Y} = Corr(X,Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Interpretation of the ρ coefficient



Sign of ρ	Very Strong	Strong	Moderate	Weak/None
+ Positive (Direct) relationship	0.80 to 1.00	0.50 to 0.79	0.30 to 0.49	0.00 to 0.29
- Negative (Inverse) relationship	-0.80 to -1.00	-0.50 to -0.79	-0.30 to -0.49	-0.00 to -0.29

Examples

Example: Calculate the correlation coefficient for “flights and rain” problem. Recall the probability mass distribution.

$p_{X,Y}$	Y=0	Y=1
$X = 0$	$\frac{14}{20}$	$\frac{1}{20}$
$X = 1$	$\frac{2}{20}$	$\frac{3}{20}$

Solution: Calculating the expected values, we have

$$E(X) = 0 \times f_X(0) + 1 \times f_X(1) = \frac{1}{4}, \qquad E(X^2) = 1^2 \times f_X(1) = \frac{1}{4} \quad \rightarrow Var(X) = \frac{3}{16}$$

$$E(Y) = 0 \times f_Y(0) + 1 \times f_Y(1) = \frac{1}{5}, \qquad E(Y^2) = 1^2 \times f_Y(1) = \frac{1}{5} \quad \rightarrow \quad Var(Y) = \frac{4}{25}$$

$$Cov(X,Y) = E(XY) - E(X)E(Y) = f_{X,Y}(1,1) - \frac{1}{4} \times \frac{1}{5} = \frac{1}{10}.$$

Thus, we obtain

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = 0.57735$$

Trinomial Distribution (Special)

Recall the binomial distribution which describes the behaviour of one discrete random variable X, where X is the number of successes in n trials when either we have success or failure.

What happens if there are three possible outcomes? This gives us the trinomial distribution. A rather fitting name, I might say!

Example: Suppose $n = 20$ students are selected at random:

- Let A be the event that a randomly selected student went to the football game on Saturday with $P(A) = 0.20 = p_1$.
- Let B be the event that a randomly selected student watched the football game on TV on Saturday with $P(B) = 0.50 = p_2$.
- Let C be the event that a randomly selected student completely ignored the football game on Saturday with $P(C) = 0.30 = 1 - p_1 - p_2$.

Example

Let *X* denote the number in the sample who went to the football game on Saturday;

Let *Y* denote the number in the sample who watched the football game on TV on Saturday; and

Let *Z* denote the number in the sample who completely ignored the football game.

A random simulation yields: **BBACB CCBAB CAAAC BCABB**

This translates to in terms of X, Y, Z as

X = 6

Y = 8

Z = 20 – X – Y = 6

So....

See that **Z** is controlled by the values of X and Y and capped at n.

Thus, for this example we can calculate the $P(X = 6\,, \mathrm{Y} = 8)$ by using our brains…!

The probability of the event A occurring 6 times in a specific order is p_1^6;

The probability of the event B occurring 8 times in a specific order is p_2^8;

The probability of the event A occurring 6 times in a specific order is $(1 - p_1 - p_2)^6$.

If we remove the order of arrangement and combine the combination of A, B, C then

$$P(X = 6\,, Y = 8) = \frac{20!}{6!\,8!\,(20 - 6 - 8)!} p_1^6 p_2^8 (1 - p_1 - p_2)^6$$

This leads us to the definition for the probability mass function for any X and Y.

Formal Definition of Trinomial

Definition. Suppose we repeat an experiment *n* independent times, with each experiment ending in one of three mutually exclusive and exhaustive ways (success, first kind of failure, second kind of failure). If we let *X* denote the number of times the experiment results in a success, let *Y* denote the number of times the experiment results in a failure of the first kind, and let *Z* denote the number of times the experiment results in a failure of the second kind, then the joint probability mass function of *X* and *Y* is:

$$P(X = k\ , Y = m) = \frac{n!}{k!\, m!\, (n-k-m)!} p_1^k p_2^m (1 - p_1 - p_2)^{n-k-m}$$

with:

k = 0, 1, …, *n; m* = 0, 1, …, *n and k + m* ≤ *n.*

Example

Example: In a plant, it is detected that 95% of items are in a very good condition, 4% are in satisfactory condition, and 1% are defective. Each hour an inspector randomly chooses 20 items to inspect. Let X be the number of items in good condition and Y be the number of items in satisfactory condition. Compute the probability of the inspector finds no defective items and only 2 items in satisfactory condition.

Solution: This is a trinomial distribution because there are three variables and one variable is based on the number of X and Y. For the scenario above,

$$k = 20 - 2 = 18, m = 2.$$

Thus, we can use the distribution definition to obtain

$$P(X = 18\,, Y = 2) = \frac{20!}{18!\,2!\,(0)!} 0.95^{18} 0.04^{2}\,(1 - 0.95 \quad - 0.04 \quad)^{0}$$

Does this answer look familiar?

Let’s begin Markov Chains

- The Markov property is satisfied by any random process for which the future is conditionally independent from the past given the present.
- A Markov chain is a random process that satisfies the Markov property. Here we consider discrete-time Markov chains with a finite state space, which means that the process can only take a finite number of values at any given time point. To specify a Markov chain, we only need to dene the probability mass function of the random process at its starting point (which we will assume is at i = 0) and its transition probabilities.

Definition 7.0.1 (Markov property). A random process satisfies the **Markov property** if $X(t_{i+1})$ is conditionally independent of $X(t_1), \dots, X(t_{i-1})$ given $X(t_i)$ for any $t_1 < t_2 < \ \dots < t_{i+1}$.

If the state space of the random process is discrete, then for any $x_1, x_2 \dots, x_{i+1}$:

$$p_{X(t_{n+1})|\, X(t_1),X(t_2),\dots,X(t_i)}(x_n|x_1, x_2, \dots., x_n) = p_{X(t_{i+1})|X(t_i)}(x_{i+1} \,|\, x_i).$$

Graphically – what does this mean?

Example (Car Rental)

(Car rental). A car-rental company hires you to model the location of their cars. The company operates in Los Angeles, San Francisco and San Jose. Customers regularly take a car in a city and drop it off in another city. It would be very useful for the company to be able to compute how likely it is for a car to end up in a given city. You decide to model the location of the car as a Markov chain, where each time step corresponds to a new customer taking the car. The company allocates new cars evenly between the three cities. The transition probabilities, obtained from past data, are given by

$$\begin{array}{cc} \begin{array}{ccc} \text{San Francisco} & \text{Los Angeles} & \text{San Jose} \end{array} & \\ \begin{pmatrix} 0.6 & 0.1 & 0.3 \\ 0.2 & 0.8 & 0.3 \\ 0.2 & 0.1 & 0.4 \end{pmatrix} & \begin{array}{c} \text{San Francisco} \\ \text{Los Angeles} \\ \text{San Jose} \end{array} \end{array}$$

Example (car rental)

To be clear, the probability that a customer moves the car from **San Francisco to LA** is 0.2, the probability that the car **stays in San Francisco** is 0.6, and so on.

Question: What is the probability of a car moving from LA to Jose?

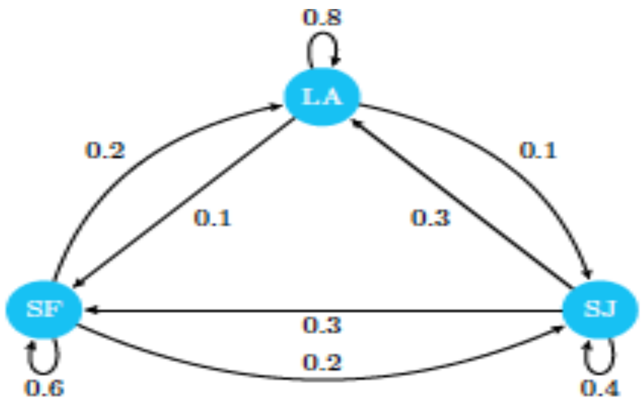
The initial state vector and the transition matrix of the Markov chain are

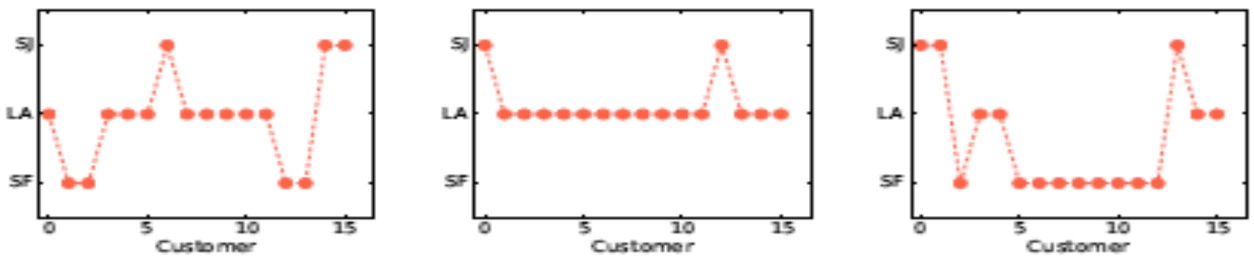
$$p_{X(0)} := \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}, \quad T_X := \begin{bmatrix} 0.6 & 0.1 & 0.3 \\ 0.2 & 0.8 & 0.3 \\ 0.2 & 0.1 & 0.4 \end{bmatrix},$$

State 1 is assigned to San Francisco, state 2 to Los Angeles and state 3 to San Jose can be seen in the state diagram of the Markov chain. The figure shows some realizations of the Markov chain.

Question: The company wants to find out the probability that the car starts in San Francisco, but is in San Jose right after the second customer.

Example (Car rental)





Car Rental

Solution: We start with the car in San Fran, where it can move (returned) freely with the first customer, but once the second customer picks up the car, it is dropped off in San Jose. This is given by the probability statement:

$$p_{X(0),X(2)}(1,3) = \sum_{i=0}^{3} p_{X(0),X(1),X(2)}(1,i,3)$$

$$= \sum_{i=1}^{3} p_{X(0)}(1)\, p_{X(1)|X(0)}(i|1) p_{X(2)|X(1)}(3|i)$$

$$= p_{X(0)}(1) \sum_{i=1}^{3} (T_{X(1)}[1,i])(T_{X(2)}[i,3])$$

$$= \frac{1}{3} \times [\, 0.6 \times 0.2 + 0.2 \times 0.1 + 0.2 \times 0.4 \,] = 7.33\%$$

Thus, we can see for this scenario it boils down to a 7.33% chance.

R and Applications

- Bivariate simulations and calculations of conditional probabilities

- To be announced on the day!

- Tutorial questions!

❑ Not to worry, the R workshops will be put online as a quick video for Week 5!!! ☺