Data Analytics

ECON 1008, Semester 1, 2019

Giulio Zanella
University of Adelaide
School of Economics

Chapter 7, continues…

Random variables and discrete probability distributions

Laws of Expected Value and Variance

Let X be a random variable and c and b are constants.

Laws of expected value

- $E(c) = c$
- $E(\mathbf{X} + c) = E(\mathbf{X}) + c$
- $E(c\mathbf{X}) = cE(\mathbf{X})$
- $E(c\mathbf{X} + b) = cE(\mathbf{X}) + b$

Laws of variance

- $V(c) = 0$
- $V(\mathbf{X} + c) = V(\mathbf{X})$
- $V(c\mathbf{X}) = c^2V(\mathbf{X})$
- $V(c\mathbf{X} + b) = c^2V(\mathbf{X})$

Example 4

With the probability distribution of cars sold per week, assume a salesman earns a fixed weekly wage of \$150 plus \$200 commission for each car sold. What is his expected wage, and the variance of the wage, for the week?

Solution

- The weekly wage $Y = 200X + 150$
- $E(Y) = E(200X + 150) = 200E(X) + 150$ $\quad E(cX + b) = cE(X) + b$

  $= 200(2.4) + 150 = \$630$
- $V(Y) = V(200X + 150) = 200^2 V(X)$ $\quad V(cX+b) = c^2V(X)$

  $= 200^2(1.24) = 49\,600\ (\$)^2$
- $SD\ (Y) = \sqrt{49600} = \$222.71$

Example 5

The monthly sales at a computer store have a mean of \$25,000 and a standard deviation of \$4,000. Profits are 30% of the sales less fixed costs of \$6,000.

Find the mean and standard deviation of the monthly profit.

Solution

- E(Sales) = \$25000

  SD(Sales) = \$4000

  V(Sales) = $(4000)^2$ $(\$)^2$
- Profit = 0.30(Sales) – 6000

Example 5: Solution

- Profit = 0.30(sales) – 6000
- E(Profit) = E[0.30(Sales) – 6000]

$E(X + c) = E(X) + c$

$E(c\mathbf{X}) = cE(\mathbf{X})$

= E[0.30(Sales)] – 6000
= 0.30 E(Sales) – 6000
= (0.30)(25000) – 6000 = \$1 500

- V(Profit) = V(0.30(Sales) – 6000]

$V(X + c) = V(X)$

$V(cX) = c^2V(X)$

= V[(0.30)(Sales)]
= $(0.30)^2$ V(Sales)

- = $(0.30)^2(4000)^2$
= 1 440 000 $(\$)^2$

- $\sigma_{\text{Profit}} = \sqrt{1\ 440\ 000} = \$1\ 200$

Joint distributions

The **bivariate (or joint) distribution** is used when the relationship between two random variables is studied.

The joint probability that X assumes the value x, and Y assumes the value y is denoted

$$p(x,y) = P(X=x \text{ and } Y = y)$$

Discrete Bivariate Distributions

The joint probability function satisfies the following conditions:

1. $0 \leq p(x, y) \leq 1$
2. $\sum_{all\ x} \sum_{all\ y} p(x, y) = 1$ for all pairs (x, y).

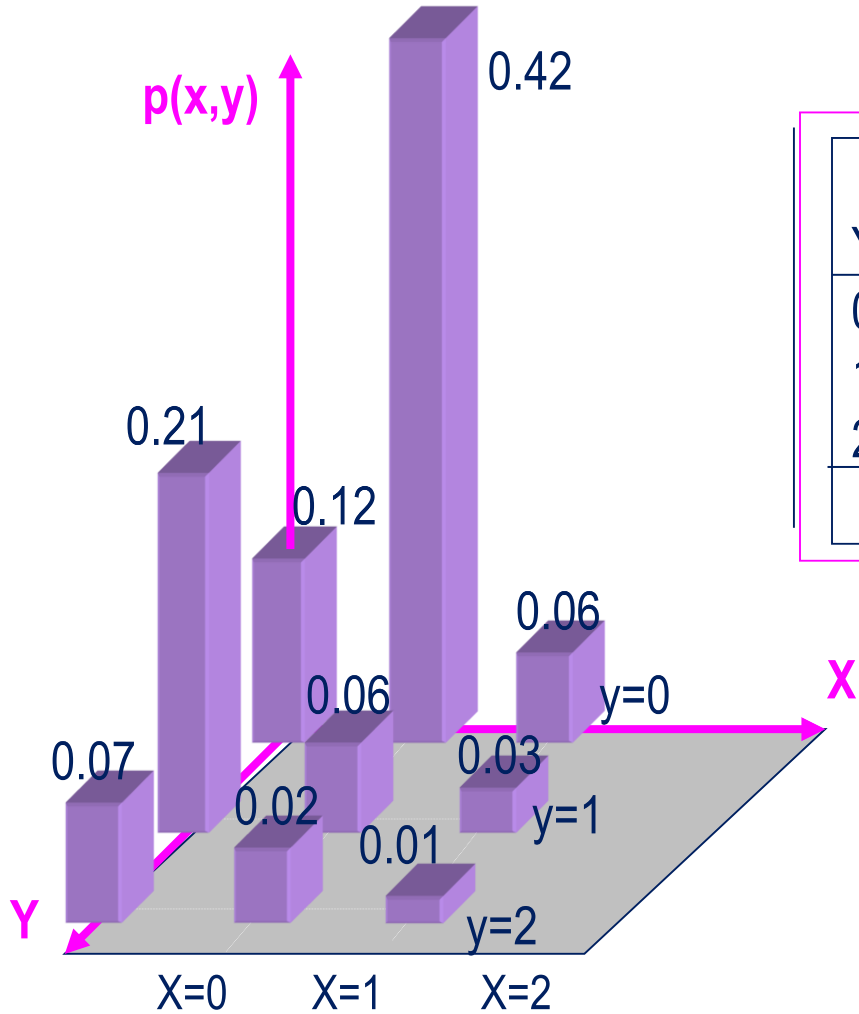
Example 6

Xavier and Yvette are two real estate agents. Let X and Y denote the number of houses that Xavier and Yvette, respectively, will sell in a month. An analysis of their past monthly performances has the following joint probabilities.

The bivariate probability distribution of (x,y)

	X		
Y	0	1	2
0	.12	.42	.06
1	.21	.06	.03
2	.07	.02	.01

Example 6: Solution



	X		
Y	0	1	2
0	.12	.42	.06
1	.21	.06	.03
2	.07	.02	.01

Marginal Probabilities

As before, we can calculate the ***marginal probabilities*** by summing across rows and down columns to determine the probabilities of X and Y individually:

		x			
		0	1	2	
y	0	0.12	0.42	0.06	0.6
	1	0.21	0.06	0.03	0.3
	2	0.07	0.02	0.01	0.1
		0.4	0.5	0.1	1.00

x	P(x)
0	0.4
1	0.5
2	0.1

y	P(y)
0	0.6
1	0.3
2	0.1

E.g the probability that Xavier sells 1 house = P(X=1) =0.50

Conditional Probability

$$P(X = x \mid Y = y) = \frac{P(X = x \text{ and } Y = y)}{P(Y = y)}$$

Example 7 – continued

	X			
Y	0	1	2	p(y)
0	.12	.42	.06	.60
1	.21	.06	.03	.30
2	.07	.02	.01	.10
p(x)	.40	.50	.10	1.00

Example 7…

$$P(X = x \mid Y = y) = \frac{P(X = x \text{ and } Y = y)}{P(Y = y)}$$

$$\left.\begin{aligned}
P(X = 0 \mid Y = 1) &= \frac{P(X = 0 \text{ and } Y = 1)}{P(Y = 1)} = \frac{.21}{.30} = .7 \\
P(X = 1 \mid Y = 1) &= \frac{P(X = 1 \text{ and } Y = 1)}{P(Y = 1)} = \frac{.06}{.30} = .2 \\
P(X = 2 \mid Y = 1) &= \frac{P(X = 2 \text{ and } Y = 1)}{P(Y = 1)} = \frac{.03}{.30} = .1
\end{aligned}\right\}$$

The sum is equal to 1.0

Conditions for Independence

Two random variables are said to be independent when

$$P(X=x|Y=y) = P(X=x) \quad \text{or} \quad P(Y=y|X=x) = P(Y=y).$$

This leads to the following relationship for independent variables

$$P(X=x \text{ and } Y=y) = P(X=x) \cdot P(Y=y)$$

Example 7…

$P(X=0|Y=1) = 0.7$
But $P(X=0)=0.4$

The condition for independence is not satisfied.

The variables X and Y are **not independent**.

Sum of Two Variables

The probability distribution of X + Y is determined by

- Determining all the possible values that X+Y can assume.
- For every possible value C of X + Y, adding the probabilities of all the combinations of X and Y for which X + Y = C.

Example 7…

Find the probability distribution of the total number of houses sold per week by Xavier and Yvette.

Solution

- X + Y is the total number of houses sold.
- X + Y can have the values 0, 1, 2, 3, 4.

	X			
Y	0	1	2	p(y)
0	.12	.42	.06	.60
1	.21	.06	.03	.30
2	.07	.02	.01	.10
p(x)	.40	.50	.10	1.00

The Probability Distribution of X+Y

$P(X+Y=0) = P(X=0 \text{ and } Y=0) = 0.12$

$P(X+Y=1) = P(X=0 \text{ and } Y=1) + P(X=1 \text{ and } Y=0) = 0.21 + 0.42 = 0.63$

$P(X+Y=2) = P(X=0 \text{ and } Y=2) + P(X=1 \text{ and } Y=1) + P(X=2 \text{ and } Y=0)$
$= 0.07 + 0.06 + 0.06 = .19$

	X			
Y	0	1	2	p(y)
0	.12	.42	.06	.60
1	.21	.06	.03	.30
2	.07	.02	.01	.10
p(x)	.40	.50	.10	1.00

The probabilities P(X+Y)=3 and P(X+Y) =4 are calculated the same way. The distribution follows.

The Probability Distribution of X+Y

The distribution of X + Y

x + y	0	1	2	3	4
p(x+y)	0.12	0.63	0.19	0.05	0.01

The Expected Value and Variance of X + Y

The distribution of X + Y

x + y	0	1	2	3	4
p(x+y)	0.12	0.63	0.19	0.05	0.01

The expected value and variance of X + Y can be calculated from the distribution of X + Y.

$$E(X+Y)=0(0.12)+1(0.63)+2(0.19)+3(0.05)+4(0.01)$$
$$=1.2$$
$$V(X+Y)=(0-1.2)^2(0.12)+(1-1.2)^2(0.63)+\ldots+(4-1.2)^2(0.01)$$
$$=0.56$$

The Expected Value and Variance of X + Y

The following relationship can assist in calculating E(X+Y)

- $E(X+Y) = E(X) + E(Y)$;

WARNING! It is **NOT** true that $V(X+Y) = V(X) + V(Y)$

This is true only when X and Y are independent.
We will study the general case later on. Preview:

$$V(X+Y) = V(X) + V(Y) + 2COV(X,Y)$$

Portfolio diversification and asset allocation

Consider an investor who forms a portfolio, consisting of only two stocks, by investing \$4000 in one stock and \$6000 in a second stock. Suppose that the returns from these two assets are independent. The results after 1 year are:

One-year results

Stock	Initial investment	Value of investment after one year	Rate of return on investment
1	\$4 000	\$5 000	$R_1 = \frac{(5000-4000)}{4000} = 0.25\ (25\%)$
2	\$6 000	\$5 400	$R_2 = \frac{(5400-6000)}{6000} = -0.10\ (-10\%)$
Total	\$10 000	\$10 400	$R_p = \frac{(10400-10000)}{10000} = 0.04\ (4\%)$

Portfolio diversification and asset allocation…

One-year results

Stock	Initial investment	Value of investment after one year	Rate of return on investment
1	\$4 000	\$5 000	$R_1 = \frac{(5000-4000)}{4000} = 0.25\ (25\%)$
2	\$6 000	\$5 400	$R_2 = \frac{(5400-6000)}{6000} = -0.10\ (-10\%)$
Total	\$10 000	\$10 400	$R_p = \frac{(10400-10000)}{10000} = 0.04\ (4\%)$

OR

$$w_1 = \frac{4000}{10000} = 0.4, \quad w_2 = \frac{6000}{10000} = 0.6$$

$$R_p = w_1R_1 + w_2R_2 = (0.4)(0.25) + (0.6)(-0.10) = 0.04$$

Portfolio diversification and asset allocation…

Mean and variance of a portfolio of two stocks, under the independence assumption

$$E(R_p) = w_1 E(R_1) + w_2 E(R_2)$$
$$V(R_p) = w_1^2 V(R_1) + w_2^2 V(R_2)$$

where

- w_1 and w_2 are the proportions or weights of investments 1 and 2,
- $E(R_1)$ and $E(R_2)$ are their expected values,

Example 8

An investor has decided to form a portfolio by putting 25% of his money into McDonald's stock and 75% into Cisco Systems stock. The investor assumes that the expected returns will be 8% and 15%, respectively, and that the standard deviations will be 12% and 22%, respectively. Suppose that the returns from these two assets are independent

a. Find the expected return on the portfolio.

b. Compute the standard deviation of the returns

Example 8: Solution

a. The expected values of the two stocks are

$$E(R_1) = 0.08 \quad \text{and } E(R_2) = 0.15$$

The weights are $w_1 = 0.25$ and $w_2 = 0.75$.

Thus,

$$\begin{aligned} E(R_2) &= w_1E(R_1) + w_2E(R_2) \\ &= 0.25(0.08) + 0.75(0.15) \\ &= 0.1325 \end{aligned}$$

Example 8: Solution

b. The standard deviations are $\sigma_1 = .12$ and $\sigma_2 = .22$.

Thus,

$$V(R_p) = w_1^2\sigma_1^2 + w_2^2\sigma_2^2$$
$$= (.25^2)(.12^2) + (.75^2)(.22^2)$$

These answers would be different in the more interesting and realistic case of assets whose returns move together (i.e., when removing the independence assumption).

We will study this more interesting and realistic case later