UNIVERSITY OF LINCOLN

SCHOOL OF MATHEMATICS AND PHYSICS

# MTH1005
# PROBABILITY AND STATISTICS

*Semester B*
*Lecture 7 (5/3/2024)*


***Danilo Roccatano***

**Office:** INB 3323
**Email:** droccatano@lincoln.ac.uk

# SUMMARY OF LAST WEEK

We have given the introduced

- the expectation of random variables,
- The variance of random variables,
- Definition of moments.

We will now look at the joint random variables.

# SUMMARY OF LAST WEEK

The **expectation value** gives the weighted average of the possible values of the random variable. The weighting is the likelihood that that value turns up. This is the mean of the random variable, often written as $\mu$ .

Given a random variable X with a probability mass function p(x) we de ne the expectation of X, written as E[X] as

$$\mathrm{E}[X] = \sum_{\text{all } x} x \cdot p(x)$$

The expectation of a discrete random variable X is just the arithmetic mean of the values it takes on.

The equivalent for a continuous random variable Z is

$$\mathrm{E}[Z] = \int_{\text{all } z} z \cdot f(z) \mathrm{d}z$$

# SUMMARY OF LAST WEEK

**Definition**

Let $g(X)$ be any function of a random variable $X$. Then

$$\mathrm{E}[g(X)] = \sum_{x \in R_X} g(x) \cdot p(x)$$

or for the continuous random variable $Z$

$$\mathrm{E}[g(Z)] = \int_{-\infty}^{\infty} g(z) \cdot f(z) \mathrm{d}z$$

---

if $X$ is a random variable, then

- $\mathrm{E}[a] = a$
- $\mathrm{E}[aX] = a\mathrm{E}[X]$
- $\mathrm{E}[g_1(X) + g_2(X)] = \mathrm{E}[g_1(X)] + \mathrm{E}[g_2(X)]$, where $g_1(X)$ and $g_2(X)$ are any functions of X.

these define the properties of a linear operator.

# SUMMARY OF LAST WEEK: VARIANCE OF A RANDOM VARIABLE

The mean describes where a distribution is centred - expected value.

The **variance** describes how widely the values of the distribution are spread around the mean.

**Definition**

If $X$ is a random variable with mean $\mu$, then the variance of $X$, denoted by $\mathrm{Var}(X)$ or $\sigma_X^2$, is defined by

$$\sigma_X^2 = \mathrm{Var}(X) = \mathrm{E}[(X - \mu)^2]$$

this can also be written

$$\sigma_X^2 = \mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$$

It is the expected squared distance of a value from the centre of the distribution.

# SUMMARY OF LAST WEEK: STANDARD DEVIATION

the standard deviation of a random variable X is the square root of its variance.

$$\sigma_X = \sqrt{\mathrm{Var}(X)} = \sqrt{\mathrm{E}[X^2] - (\mathrm{E}[X])^2} = \sqrt{\mathrm{E}[(X-\mu)^2]}$$

a major advantage of the standard deviation is that it has the same units, if any, as the variable itself.

It gives the expected root mean squared distance of a data point from the centre of the distribution.

The smaller the variance or standard deviation, the more well defined a distribution is.

# MOMENTS OF A RANDOM VARIABLE

The mean and variance describe a probability law for a random variable to some extent - the centre of the distribution

$$\mu_X = E[X]$$

and how far points stray from the centre $\sigma^2_X$.

The $k^{th}$ moment of a random variable X is defined as

$$m_k = E[X^k]$$

## MOMENTS OF A RANDOM VARIABLE: Shape of a distribution

Therefore

$$m_1 = E[X] = \mu_X$$

And the second moment is

$$m_2 = E[X^2] = \sigma_X^2 + \mu_X^2$$

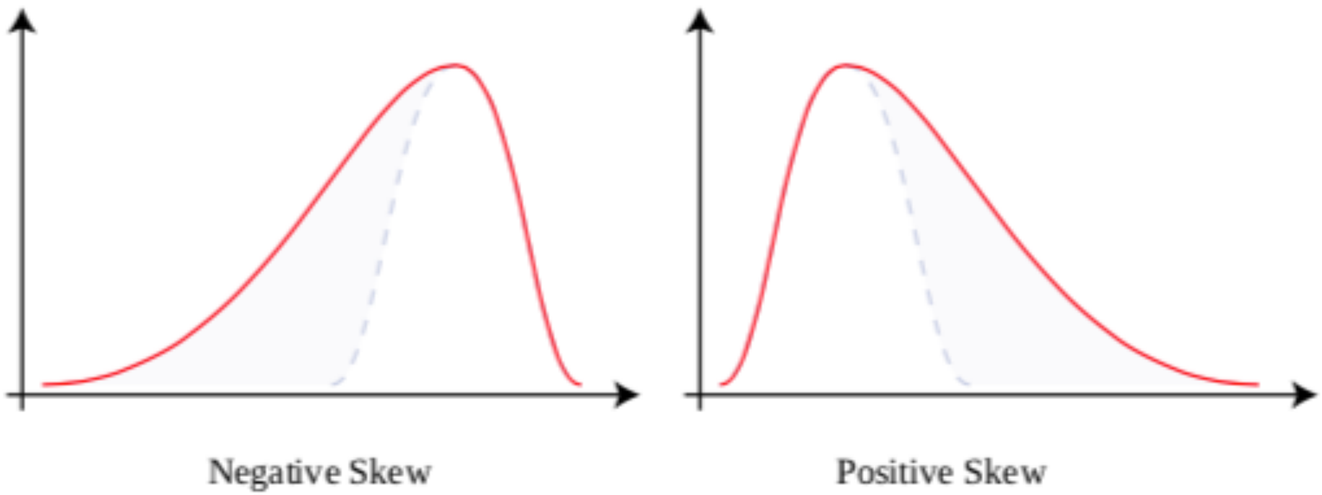
so the mean and variance can be defined in terms of the first two moments of the distribution.

# MOMENTS OF A RANDOM VARIABLE: Shape of a distribution

Higher moments de ne skewness ($m_3$ ) →

- how wonky a distribution is (zero for a symmetric distribution)

- and other details of the shape of the distribution.




A full set of moments fully defines a distribution (though sometimes not all moments will be well defined).

# JOINT RANDOM VARIABLES

Joint random variables refer to a concept in probability theory where we consider the behavior and outcomes of multiple random variables together.

This involves studying how these variables co-vary or interact with each other within a given probability distribution.

By analyzing joint random variables, we gain insights into the relationship between different aspects of a system or phenomenon, allowing us to make more informed predictions and decisions in various fields such as statistics, economics, and engineering.

# JOINT PROBABILITY MASS FUNCTION

## DEFINITION

Let $X$ and $Y$ be two discrete rv's defined on the sample space $\mathcal{S}$ of an experiment. The **joint probability mass function** $p(x, y)$ is defined for each pair of numbers $(x, y)$ by

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

It must be the case that $p(x, y) \geq 0$ and $\sum_x \sum_y p(x, y) = 1$.

Now let $A$ be any particular set consisting of pairs of $(x, y)$ values (e.g., $A = \{(x, y): x + y = 5\}$ or $\{(x, y): \max(x, y) \leq 3\}$). Then the probability $P[(X, Y) \in A]$ that the random pair $(X, Y)$ lies in the set $A$ is obtained by summing the joint pmf over pairs in $A$:

$$P[(X, Y) \in A] = \sum_{(x, y) \in A} \sum p(x, y)$$

# JOINT RANDOM VARIABLES

We define two random variables, X and Y. Consider the joint probability mass function of X and Y, denoted as

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

for discrete random variables,

It's important to note that the capitalized X represents the random variable itself, while lowercase x represents a specific value it can take on.

In the discrete case, **how many values can this function take?** Let's examine an example.

# JOINT RANDOM VARIABLES: AN EXAMPLE

Given a population of 41 students, we set up two random variables:

- X=0, if a student was doing pure maths, and X = 1 if a student was doing any other course code.
- Y = 0 if a student's surname did not begin with M, Y = 1 if a student's surname began with M.

There are 4 joint probabilities for X equal 0 or 1 and Y is 0 or 1.

So, we need to define $p(0, 0)$, $p(0, 1)$, $p(1, 0)$, $p(1, 1)$

# JOINT RANDOM VARIABLES: AN EXAMPLE

Counting the students for the different cases, we can build the following Table with probabilities that both of the variables have specific values.
The joint PMF can be described using a tabular representation

| | $X = 0$ | $X = 1$ | sum |
|---|---|---|---|
| $Y = 0$ | $\frac{22}{41}$ | $\frac{16}{41}$ | $\frac{38}{41}$ |
| $Y = 1$ | $\frac{3}{41}$ | $\frac{0}{41}$ | $\frac{3}{41}$ |
| sum | $\frac{25}{41}$ | $\frac{16}{41}$ | $\frac{41}{41}$ |

The marginal values give us the individual probability mass functions of X , $p_X$ (x) and Y, $p_Y$ (y), respectively.

## JOINT RANDOM VARIABLES: MARGINAL PMF

- Once the joint pmf of the two variables $X$ and $Y$ is available, it is in principle straightforward to obtain the distribution of just one of these variables.

  As an example, let $X$ and $Y$ be the number of statistics and mathematics courses, respectively, currently being taken by a randomly selected statistics major.

- Suppose that we wish the distribution of $X$, and that when $X = 2$, the only possible values of $Y$ are 0, 1, and 2.

# JOINT RANDOM VARIABLES: MARGINAL PMF

- Then
  - $p_X(2) = P(X = 2) = P[(X, Y) = (2, 0) \text{ or } (2, 1) \text{ or } (2, 2)]$

$$= p(2, 0) + p(2, 1) + p(2, 2)$$

- That is, the joint pmf is summed over all pairs of the form (2, *y*). More generally, for any possible value *x* of *X*, the probability $p_X(x)$ results from holding *x* fixed and summing the joint pmf $p(x, y)$ over all *y* for which the pair $(x, y)$ has positive probability mass.

  The same strategy applies to obtaining the distribution of *Y* by itself.

# MEANINGS OF MARGINAL PMF

The marginal values along rows and columns give us the individual probability mass functions of X , $p_Y$ (x) and of Y , $p_X$ (y), respectively.

We can see this property of the marginal probability value (column or row sums) as each entry in the table is

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

We then have using, the law of total probability

$$P(A) = \sum_{i=1}^{n} P(A \cap E_i) = \sum_{i=1}^{n} P(A \mid E_i) P(E_i)$$

# MEANINGS OF MARGINAL PMF

we can immediately apply this result to our example for the variable X, where the $E_i$ are the values of *y,* obtaining

$$P\{X = x\} = \sum_{\text{all } y} P(X = x \cap Y = y) = \sum_{\text{all } y} P(X = x \mid Y = y) P(Y = y).$$

and equally, for the Y variable, considering $E_i$ as the values of *x:*

$$P\{Y = y\} = \sum_{\text{all } x} P(Y = y \cap X = x) = \sum_{\text{all } x} P(Y = y \mid X = x) P(X = x).$$

# MARGINAL PMF

## Definition

The **marginal probability mass function of $X$**, denoted by $p_X(x)$, is given by

$$p_X(x) = \sum_{y:\, p(x,y)>0} p(x, y) \quad \text{for each possible value } x$$

Similarly, the **marginal probability mass function of $Y$** is

$$p_Y(y) = \sum_{x:\, p(x,y)>0} p(x, y) \quad \text{for each possible value } y.$$

# JOINT RANDOM VARIABLES: ANOTHER EXAMPLE

Anyone who purchases an insurance policy for a home or automobile must specify a deductible amount, the amount of loss to be absorbed by the policyholder before the insurance company begins paying out.

Suppose that a particular company offers auto deductible amounts of \$100, \$500, and \$1000, and homeowner deductible amounts of \$500, \$1000, and \$2000.

Consider randomly selecting someone who has both auto and homeowner insurance with this company and let

*X =“the amount of the auto policy deductible”*

and

*Y = “the amount of the homeowner policy deductible”.*

# JOINT RANDOM VARIABLES: ANOTHER EXAMPLE

- The joint pmf of these two variables appears in the accompanying *joint probability table*:

| | $p(x, y)$ | $y$ = 500 | 1000 | 5000 |
|---|---|---|---|---|
| $x$ | 100 | .30 | .05 | 0 |
| | 500 | .15 | .20 | .05 |
| | 1000 | .10 | .10 | .05 |

- According to this joint pmf, there are nine possible $(X, Y)$ pairs: (100, 500), (100, 1000), … , and finally (1000, 5000).
- The probability of (100, 500) is

$$p(100, 500) = P(X = 100, Y = 500) = .30.$$

- Clearly $p(x, y) \geq 0$, and it is easily confirmed that the sum of the nine displayed probabilities is 1.

# JOINT RANDOM VARIABLES: ANOTHER EXAMPLE

- The probability $P(X = Y)$ is computed by summing $p(x, y)$ over the two $(x, y)$ pairs for which the two deductible amounts are identical:

$$P(X = Y) = p(500, 500) + p(1000, 1000) = .15 + .10 = .25$$

| | $p(x, y)$ | $y$: 500 | 1000 | 5000 |
|---|---|---|---|---|
| | 100 | .30 | .05 | 0 |
| $x$ | 500 | .15 | .20 | .05 |
| | 1000 | .10 | .10 | .05 |

- Similarly, the probability that the auto deductible amount is at least \$500 is the sum of all probabilities corresponding to $(x, y)$ pairs for which $x \geq 500$; this is the sum of the probabilities in the bottom two rows of the joint probability table:

$$P(X \geq 500) = .15 + .20 + .05 + .10 + .10 + .05 = .65$$

| | $p(x, y)$ | $y$: 500 | 1000 | 5000 |
|---|---|---|---|---|
| | 100 | .30 | .05 | 0 |
| $x$ | 500 | .15 | .20 | .05 |
| | 1000 | .10 | .10 | .05 |

# JOINT RANDOM VARIABLES: ANOTHER EXAMPLE

## MARGINAL PMF

The possible *X* values are *x* = 100, 500, and *x* = 1000, so computing row totals in the joint probability table yields

$$p_X(100) = p(100, 500) + p(100, 1000) + p(100, 5000) = .30 + .05 + 0 = .35$$

$$p_X(500) = .15 + .20 + .05 = .40, \; p_X(1000) = 1 - (.35 + .40) = .25$$

The marginal pmf of $X$ is then

$$p_X(x) = \begin{cases} .35 & x = 100 \\ .40 & x = 500 \\ .25 & x = 1000 \\ 0 & \text{otherwise} \end{cases}$$

# JOINT RANDOM VARIABLES: ANOTHER EXAMPLE

- Similarly, the marginal pmf of *X* is then

$$p_X(x) = \begin{cases} .35 & x = 100 \\ .40 & x = 500 \\ .25 & x = 1000 \\ 0 & \text{otherwise} \end{cases}$$

- From this pmf, $P(X \geq 500) = .40 + .25 = .65$, which we already calculated in Example 5.1. Similarly, the marginal pmf of *Y* is obtained from the column totals as

$$p_Y(y) = \begin{cases} .55 & y = 500 \\ .35 & y = 1000 \\ .10 & y = 5000 \\ 0 & \text{otherwise} \end{cases}$$

# EXPECTATION OF DISCRETE JOINT RANDOM VARIABLES.

# EXPECTATION OF DISCRETE JOINT RANDOM VARIABLES.

The joint **probability mass function** contains all the information about the variables it describes.

We've just seen that we can retrieve the individual probability mass/density functions by **summing** over the other variables.

If we have a general function of the variables, **g(X, Y)**, say, then we can obtain the expectation value of that function as

$$\mathrm{E}[g(X,Y)] = \sum_y \sum_x g(x,y)p(x,y)$$

for discrete random variables and

# PROBABILITY OF CONTINUOUS JOINT RANDOM VARIABLES

## PROBABILITY OF CONTINUOUS JOINT RANDOM VARIABLES.

- The probability that the observed value of a continuous rv $X$ lies in a one-dimensional set $A$ (such as an interval) is obtained by integrating the pdf $f(x)$ over the set $A$.

- Similarly, the probability that the pair $(X, Y)$ of continuous rv's falls in a two-dimensional set $A$ (such as a rectangle) is obtained by integrating a function called the *joint density function*.

# CONTINUOUS JOINT RANDOM VARIABLES.

## DEFINITION

Let $X$ and $Y$ be continuous rv's. A **joint probability density function** $f(x, y)$ for these two variables is a function satisfying $f(x, y) \geq 0$ and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$. Then for any two-dimensional set $A$

$$P[(X, Y) \in A] = \iint_A f(x, y)\, dx\, dy$$

In particular, if $A$ is the two-dimensional rectangle $\{(x, y): a \leq x \leq b, c \leq y \leq d\}$, then

$$P[(X,Y) \in A] = P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y)\, dy\, dx$$

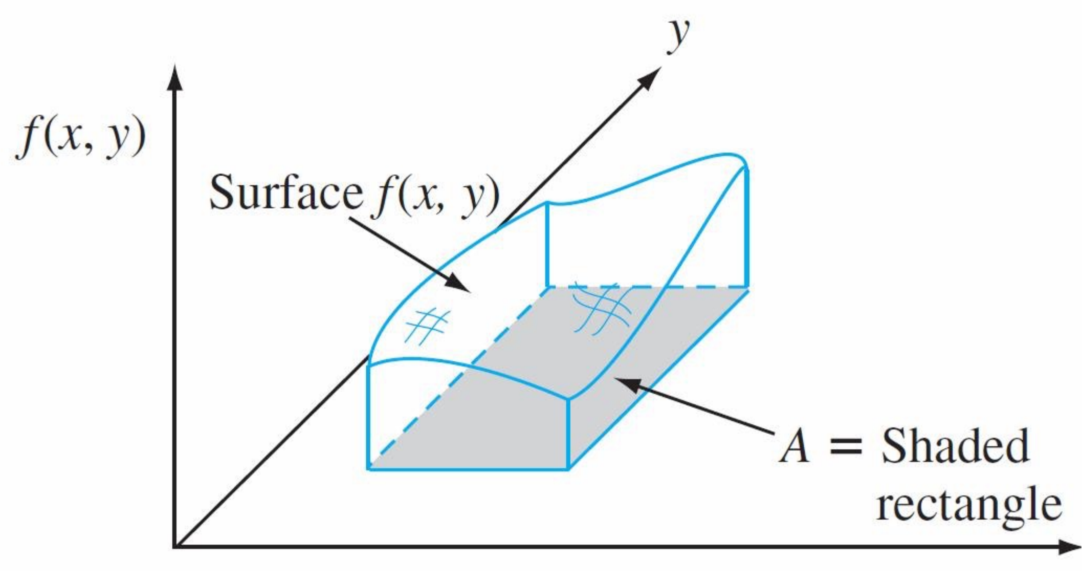
## PROBABILITY OF CONTINUOUS JOINT RANDOM VARIABLES.

- We can think of $f(x, y)$ as specifying a surface at height $f(x, y)$ above the point $(x, y)$ in a three-dimensional coordinate system.

  Then $P[(X, Y) \in A]$ is the volume underneath this surface and above the region $A$, analogous to the area under a curve in the case of a single rv.

## PROBABILITY OF CONTINUOUS JOINT RANDOM VARIABLES.

- This is illustrated in the following figure.




$P[(X, Y) \in A]$ = volume under density surface above $A$

**Figure 5.1**

# PROBABILITIES OF CONTINUOUS JOINT RANDOM VARIABLES: EXAMPLE

- A bank operates both a drive-up facility and a walk-up window. On a randomly selected day, let

$X$ = "*the proportion of time that the drive-up facility is in use*" (at least one customer is being served or waiting to be served) and
$Y$ = "*the proportion of time that the walk-up window is in use*".

- Then the set of possible values for ($X$, $Y$) is the rectangle

$$D = \{(x, y): 0 \leq x \leq 1, 0 \leq y \leq 1\}.$$

# PROBABILITIES OF CONTINUOUS JOINT RANDOM VARIABLES: EXAMPLE

Suppose the joint pdf of (*X*, *Y*) is given by

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = \int_0^1 \int_0^1 \frac{6}{5}(x + y^2)\, dx\, dy$$

To verify that this is a legitimate pdf, note that *f* (*x*, *y*) ≥ 0 and

$$= \int_0^1 \int_0^1 \frac{6}{5} x\, dx\, dy + \int_0^1 \int_0^1 \frac{6}{5} y^2\, dx\, dy$$

# PROBABILITIES OF CONTINUOUS JOINT RANDOM VARIABLES: EXAMPLE

cont'd

$$= \int_0^1 \frac{6}{5} x \, dx + \int_0^1 \frac{6}{5} y^2 \, dy = \frac{6}{10} + \frac{6}{15} = 1$$

$$= \frac{6}{10} + \frac{6}{15}$$

$$= 1$$

The probability that neither facility is busy more than

$$P\left(0 \le X \le \frac{1}{4}, 0 \le Y \le \frac{1}{4}\right) = \int_0^{1/4} \int_0^{1/4} \frac{6}{5} (x + y^2) \, dx \, dy$$

$$= \frac{6}{5} \int_0^{1/4} \int_0^{1/4} x \, dx \, dy + \frac{6}{5} \int_0^{1/4} \int_0^{1/4} y^2 \, dx \, dy$$

# PROBABILITIES OF CONTINUOUS JOINT RANDOM VARIABLES: EXAMPLE

cont'd

$$= \frac{6}{20} \cdot \frac{x^2}{2}\bigg|_{x=0}^{x=1/4} + \frac{6}{20} \cdot \frac{y^3}{3}\bigg|_{y=0}^{y=1/4}$$

$$= \frac{7}{640}$$

$$= .0109$$

# MARGINAL PROBABILITIES OF CONTINUOUS JOINT RANDOM VARIABLES

- The marginal pdf of each variable can be obtained in a manner analogous to what we did in the case of two discrete variables.

  The marginal pdf of $X$ at the value $x$ results from holding $x$ fixed in the pair $(x, y)$ and *integrating* the joint pdf over $y$. Integrating the joint pdf with respect to $x$ gives the marginal pdf of $Y$.

# MARGINAL PROBABILITIES OF CONTINUOUS JOINT RANDOM VARIABLES

## DEFINITION

The **marginal probability density functions** of $X$ and $Y$, denoted by $f_X(x)$ and $f_Y(y)$, respectively, are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \qquad \text{for } -\infty < x < \infty$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \qquad \text{for } -\infty < y < \infty$$

# MARGINAL PROBABILITIES OF CONTINUOUS JOINT RANDOM VARIABLES: EXAMPLE

- The marginal pdf of *X*, which gives the probability distribution of busy time for the drive-up facility without reference to the walk-up window, is

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,dy = \int_0^1 \frac{6}{5}(x + y^2)\,dy = \frac{6}{5}x + \frac{2}{5}$$

- for $0 \le x \le 1$ and 0 otherwise. The marginal pdf of *Y* is

$$f_Y(y) = \begin{cases} \frac{6}{5}y^2 + \frac{3}{5} & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

## MARGINAL PROBABILITIES OF CONTINUOUS JOINT RANDOM VARIABLES: EXAMPLE

- Then

$$P(.25 \leq Y \leq .75) = \int_{.25}^{.75} f_Y(y)\,dy = \frac{37}{80} = .4625$$

# EXPECTATION OF CONTINUOUS JOINT RANDOM VARIABLES.

The joint **probability density function** contains all the information about the variables it describes.

$$P\{X = x, Y = y\} = f(x, y)dxdy$$

We've just seen that we can retrieve the individual probability density functions by **integrating** over the other variables.

If we have a general function of the variables, **g(X, Y)**, say, then we can obtain the expectation value of that function as

$$\mathrm{E}[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y$$

for continuous random variables

# EXPECTATION OF SUMS OF RANDOM VARIABLES

Using the last information,  we now come to a very important general result that will be the basis of a lot of the development we do later.

What we will seek to show is that

$$E[X_1 + \ \ldots + E_n] = E[X_1] + \ \ldots E[X_n]$$

the expectation of the sum of the random variables $X_1$ … $X_n$ is just the sum of their individual expectations.

# EXPECTATION OF SUMS OF RANDOM VARIABLES

Let us start with just two variables, X and Y . We can use the general expression for the expectation of a function of two random variables

$$
\begin{aligned}
\mathrm{E}[X+Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf(x,y)\mathrm{d}x\mathrm{d}y + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{\infty}xf_X(x)\mathrm{d}x + \int_{-\infty}^{\infty}yf_Y(y)\mathrm{d}y \\
&= \mathrm{E}[X] + \mathrm{E}[Y]
\end{aligned}
$$

At this point we can use proof by induction to show what we were after:

$$E[X_1 + \ \ldots + E_n] = E[X_1] + \ \ldots E[X_n]$$

# Example

In one of the practical, we assigned X as the 'number of heads' when two coins were flipped. We also have defined other teo random variables, $X_1$ ‘the number of heads on the first coin’, and $X_2$ ‘the number of heads on the second coin’, so that
$X = X_1 + X_2$.
We have calculated the expectation value of $X_1$ as

$$
\begin{aligned}
\mathrm{E}[X_1] = \mathrm{E}[X_2] &= \sum_{\text{all } x_1} x_1 \cdot p(x_1) \\
&= 0 \cdot p(0) + 1 \cdot p(1) \\
&= 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} \\
&= \frac{1}{2} = \mu_{X_1} \\
&= \mu_{X_2}
\end{aligned}
$$

We can see that as expected

$$\mathbf{E[X] = E[X_1\,] + E[X_2\,] = 1}$$

what about the variance of the variables?

# Example

$$
\begin{aligned}
\mathrm{E}[X_1^2] = \mathrm{E}[X_2^2] &= \sum_{\text{all } x_1} x_1^2 \cdot p(x_1) \\
&= 0 \cdot p(0) + 1 \cdot p(1) \\
&= 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} \\
&= \frac{1}{2}
\end{aligned}
$$

So

$$\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = \frac{1}{2} - (\frac{1}{2})^2 = \frac{1}{4}$$

and we get in this case (because $X_1$ and $X_2$ are independent)

$$Var(X) = Var(X_1) + Var(X_2)$$

It should be clear that this would hold for tossing three coins, four coins … *n* coins, which again could be proved by induction.

# COVARIANCES JOINT RANDOM VARIABLES

Using the general expression of the expectation value for joint random variable (adjusted to our particular case)

$$\mathrm{E}[g(X,Y)] = \sum_y \sum_x g(x,y)p(x,y)$$

We can also write this expression as

$$\begin{aligned}\mathrm{Cov}(X_1, X_2) &= \mathrm{E}\Big[(X_1 - \mathrm{E}[X_1]) \cdot (X_2 - \mathrm{E}[X_2])\Big] \\ &= \sum_{\text{all } x_1} \sum_{\text{all } x_2} (x_1 - \mu_{X_1}) \cdot (x_2 - \mu_{X_2}) \cdot p(x_1, x_2)\end{aligned}$$

# USEFUL FORMULA

In a similar way to the variance the covariance can more easily be calculated using an alternative expression

$$\begin{aligned}\operatorname{Cov}(X,Y) &= \mathrm{E}\Big[(X-\mathrm{E}[X])\cdot(Y-\mathrm{E}[Y])\Big]\\ &= \mathrm{E}\Big[XY-\mathrm{E}[X]Y-X\mathrm{E}[Y]+\mathrm{E}[X]\mathrm{E}[Y]\Big]\\ &= E[XY]-E[X]E[Y]-E[X]E[Y]+E[X]E[Y]\\ &= E[XY]-E[X]E[Y]\end{aligned}$$

# COVARIANCES JOINT RANDOM VARIABLES

So,

$$\begin{aligned}\mathrm{Cov}(X_1, X_2) &= \mathrm{E}\Big[(X_1 - \mathrm{E}[X_1]) \cdot (X_2 - \mathrm{E}[X_2])\Big] \\ &= \sum_{\text{all } x_1} \sum_{\text{all } x_2} (x_1 - \mu_{X_1}) \cdot (x_2 - \mu_{X_2}) \cdot p(x_1, x_2)\end{aligned}$$

In this case we can very reasonably expect that

$$\mathbf{p(x_1,x_2) = p(x_1)p(x_2)}$$

(remember the definition of independent probabilities).

# EXAMPLE

Calculate $E[X_1 X_2]$ and $Cov(X_1, X_2)$ for our two coins example.

$$
\begin{aligned}
\mathrm{E}[X_1 X_2] = \sum_{\text{all } x_1} \sum_{\text{all } x_2} (x_1) \cdot (x_2) \cdot p(x_1, x_2) = \\
\sum_{\text{all } x_1} \sum_{\text{all } x_2} (x_1) \cdot (x_2) \cdot p(x_1) \cdot p(x_2) = \\
(0) \times (0) \times \frac{1}{2} \times \frac{1}{2} + \\
(1) \times (0) \times \frac{1}{2} \times \frac{1}{2} + \\
(0) \times (1) \times \frac{1}{2} \times \frac{1}{2} + \\
(1) \times (1) \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}
\end{aligned}
$$

## COVARIANCES JOINT RANDOM VARIABLES

We can now write out all the terms in the double sum:

$$\sum_{\text{all } x_1} \sum_{\text{all } x_2} (x_1 - \mu_{X_1}) \cdot (x_2 - \mu_{X_2}) \cdot p(x_1, x_2) =$$

$$(0 - \frac{1}{2}) \times (0 - \frac{1}{2}) \times \frac{1}{2} \times \frac{1}{2} +$$
$$(1 - \frac{1}{2}) \times (0 - \frac{1}{2}) \times \frac{1}{2} \times \frac{1}{2} +$$
$$(0 - \frac{1}{2}) \times (1 - \frac{1}{2}) \times \frac{1}{2} \times \frac{1}{2} +$$
$$(1 - \frac{1}{2}) \times (1 - \frac{1}{2}) \times \frac{1}{2} \times \frac{1}{2} =$$
$$\frac{1}{16} - \frac{1}{16} - \frac{1}{16} + \frac{1}{16} = 0$$

this will be the case whenever the two variables are independent.

# EXAMPLE

looking back we see that this is indeed equal to $E[X_1]E[X_2]$, so again

$$\begin{aligned}\mathrm{Cov}(X_1, X_1) &= \mathrm{E}\Big[(X_1 - \mathrm{E}[X_1]) \cdot (X_2 - \mathrm{E}[X_2])\Big] \\ &= \mathrm{E}[X_1 X_2] - \mathrm{E}[X_1]\mathrm{E}[X_2] = 0\end{aligned}$$

# SUMMARY: COVARIANCES JOINT RANDOM VARIABLES

- Random variables whose covariance is zero are called uncorrelated ***Cov(X,Y)=0***

- In the case that that the two variables are independent the covariance is also zero as

$$\mathbf{p(x_1,x_2) = p(x_1)p(x_2)}$$

and

$$E[X_1X_2]=E[X_1]E[X_2]$$

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

**However, note that the Cov(X,Y) can be zero even if the two variable are dependent.**

# MEANING OF COVARIANCE

It indicates whether there is a relationship between two random variables. We just saw that in the case of two independent random variables the covariance will be zero.

Let's define two special random variables, X and Y. They indicate whether or not events A and B occur. They are sometimes called **indicator variables.**
In our language of random variables

if A occurs

$$X = \begin{cases} 1\, if\ A\ occurs \\ 0\quad otherwise \end{cases}$$

$$Y = \begin{cases} 1\, if\ B\ occurs \\ 0\quad otherwise \end{cases}$$

and we can then see that

$$XY = \begin{cases} 1\, if\ X = 1, Y = 1 \\ 0\qquad otherwise \end{cases}$$

# MEANING OF COVARIANCE

So if we look at the covariance of X and Y we get

$$\begin{aligned} Cov(X, Y) &= E[XY] - E[X]E[Y] \\ &= P(\{X = 1, Y = 1\}) - P(\{X = 1\})P(\{Y = 1\}) \end{aligned}$$

the last line comes from the definition of the variables, it is the proportion that the variable is in the 1 state.

All the other terms in the general expression are zero because we multiply by zero.

# MEANING OF COVARIANCE

What if the covariance was greater than 0?
What would that imply for the variables (and therefore the events A and B)?

$$\mathrm{Cov}(X, Y) > 0 \Longrightarrow P(\{X = 1, Y = 1\}) > P(\{X = 1\})P(\{Y = 1\})$$

$$\Longrightarrow \frac{P(\{X = 1, Y = 1\})}{P(\{X = 1\})} > P(\{Y = 1\})$$

Using the definition of a conditional probability

$$\Longrightarrow P(\{Y = 1 \mid X = 1\}) > P(\{Y = 1\})$$

This means that if the event A has happened (so X=1), it becomes more likely that event B (Y=1) will also happen.

Note that we could swap X and Y above, and it would imply that B became more likely when A has occurred too.

# MEANING OF COVARIANCE

It might be that the covariance is negative. Then we get

$$\mathrm{Cov}(X, Y) < 0 \Longrightarrow P(\{X = 1, Y = 1\}) < P(\{X = 1\})P(\{Y = 1\})$$

$$\Longrightarrow \frac{P(\{X = 1, Y = 1\})}{P(\{X = 1\})} < P(\{Y = 1\})$$

$$\Longrightarrow P(\{Y = 1 \mid X = 1\}) < P(\{Y = 1\})$$

So, in this case, the event A happening makes B less likely.

# CORRELATION

In general it can be shown that a positive **Cov(X,Y)** is an indication that Y increases when X does. A negative Cov(X,Y) is an indication that Y decreases when X increases.

What we want is a better indicator than the covariance.

The difficulty with the covariance is that it has dimensions, and its size is dependent on the definition of the variables. Instead, or in addition, we can use the correlation. definition

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

# CORRELATION

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

It can be shown that the correlation will always have a value between −1 and +1.

The significance of the correlation is similar to just discussed for the covariance:

- Positive correlation between two variables means that X increases as Y increases.
- Negative correlation means that X decreases as Y increases.