Statistics

MATH2089

THE UNIVERSITY OF NEW SOUTH WALES

Semester 1, 2018 – Lecture 4

This lecture

4. Random variables

Additional reading: Sections 5.4, 1.3 and 3.6 in the textbook

4. Random variables

Introduction

Often, we are not interested in all of the details of an experiment but only in some numerical quantities determined by the outcome.

Example 1: tossing two dice when playing a board game

$S = \{(1,1),(1,2),\ldots,(6,5),(6,6)\}$

... but often only the **sum** of the points matters

$\rightarrow$ each possible outcome ω is characterised by a real number

Example 2: buying 2 electronic items

each of which may be either defective or acceptable

$S = \{(d,d),(d,a),(a,d),(a,a)\}$

... but we might only be interested in the **number of acceptable items** obtained in the purchase

$\rightarrow$ again, each possible outcome ω is characterised by a real number

It is often much more natural to directly think in terms of the numerical quantity of interest, called a **random variable**.

Random variable: definition

Definition

A random variable is a real-valued function defined over the sample space:

$$X : S \to \mathbb{R}$$
$$\omega \to X(\omega)$$

Usually*, a random variable is denoted by an uppercase letter.

Define S_X the domain of variation of X, that is the set of possible values taken by X.

Example 1: tossing two dice when playing a board game

$X =$ sum of the points, $\quad S_X = \{2, 3, 4, \ldots, 12\}$

Example 2: buying 2 electronic items

$X =$ number of acceptable items, $\quad S_X = \{0, 1, 2\}$

*except in your textbook

Events defined by random variables

For any fixed real value $x \in S_X$, assertions like "$X = x$" or "$X \leq x$" correspond to a set of possible outcomes

$$(X = x) = \{\omega \in S : X(\omega) = x\}$$
$$(X \leq x) = \{\omega \in S : X(\omega) \leq x\}$$

$\rightarrow$ they are events ! $\rightarrow$ meaningful to talk about their probability

Example 1 (ctd.) - If the dice are fair

$(X = 2) = \{(1, 1)\}$ $\rightarrow \mathbb{P}(X = 2) = 1/36$
$(X \geq 11) = \{(5, 6), (6, 5), (6, 6)\}$ $\rightarrow \mathbb{P}(X \geq 11) = 3/36 = 1/12$

The usual properties of probabilities apply, e.g.

- $\mathbb{P}(X \in S_X) = 1$
- $\mathbb{P}((X = x_1) \cup (X = x_2)) = \mathbb{P}(X = x_1) + \mathbb{P}(X = x_2)$ (if $x_1 \neq x_2$)
- $\mathbb{P}(X < x) = 1 - \mathbb{P}(X \geq x)$ ('$X < x$' is the complement of '$X \geq x$')

Notes

Note 1

It is important not to confuse:

- X, the name of the random variable
- $X(\omega)$, the numerical value taken by the random variable at some sample point ω
- x, a generic numerical value

Note 2

Most interesting problems can be stated, often naturally, in terms of random variables.

$\rightarrow$ Many inessential details about the sample space can be left unspecified, and one can still solve the problem

$\rightarrow$ Often more helpful to think of random variables simply as variables whose values are likely to lie within certain ranges of the real number line

Cumulative distribution function

A random variable is often described by its **cumulative distribution function** (cdf) (or just distribution).

Definition

The cdf of the random variable X is defined for any real number x, by

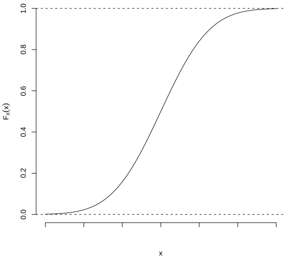
$$F(x) = \mathbb{P}(X \leq x)$$

All probability questions about X can be answered in terms of its distribution. We will denote $X \sim F$ (read 'X follows the distribution F').

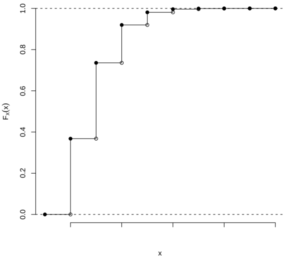
Some properties:

- For any $a \leq b$, $\mathbb{P}(a < X \leq b) = F(b) - F(a)$
- F is a nondecreasing function
- $\lim_{x \to +\infty} F(x) = F(+\infty) = 1$
- $\lim_{x \to -\infty} F(x) = F(-\infty) = 0$

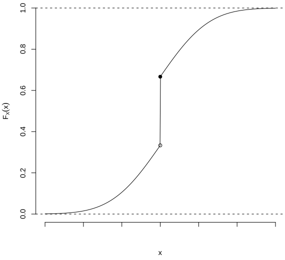
Cumulative distribution functions



Continuous distribution
$\rightarrow$ continuous r.v.



Discrete distribution
$\rightarrow$ discrete r.v.



Hybrid distribution
$\rightarrow$ hybrid r.v.

Note: hybrid distributions will not be introduced in this course.

Discrete random variables

Definition

A random variable is said to be discrete if it can only assume a finite (or at most countably infinite) number of values.

Suppose that those values are $S_X = \{x_1, x_2, \ldots\}$.

Definition

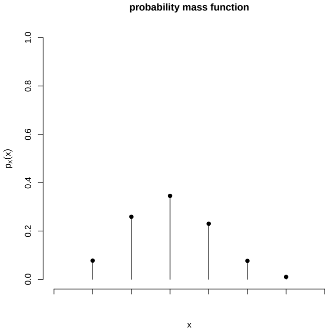
The probability mass function (pmf) of a discrete random variable X is defined for any real number x, by

$$p(x) = \mathbb{P}(X = x)$$

$\rightarrow p_X(x) > 0$ for $x = x_1, x_2, \ldots$, and $p_X(x) = 0$ for any other value of x

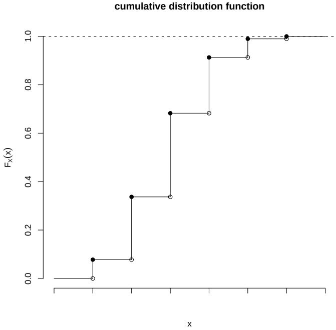
Obviously:

$$\mathbb{P}(X \in S_X) = \mathbb{P}((X = x_1) \cup (X = x_2) \cup \ldots) = \sum_{x \in S_X} p(x) = 1$$



Probability mass function:

- "spikes" at $x_1, x_2, \ldots$
- height of spike at x_i = $p(x_i)$



Cumulative distribution function:

- $F(x) = \sum_{i:x_i \leq x} p(x_i)$
- step function
- jumps at $x_1, x_2, \ldots$
- magnitude of jump at x_i = $p(x_i)$

Discrete random variables: examples

Examples of discrete random variables include:

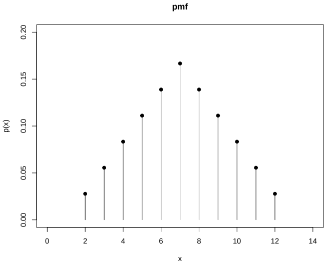
number of scratches on a surface, number of defective parts among 1000 tested, number of transmitted bits received in error, the sum of the points when tossing 2 dice, ...

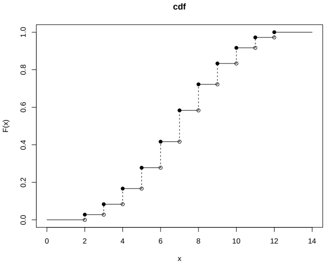
$\Rightarrow$ discrete random variables generally arise when we count things

Example: tossing 2 dice

$X =$ sum of the points; represent $p(x)$ and $F(x)$

Check that $p(x) = (6 - |7 - x|)/36$ for $x \in S_X = \{2, 3, 4, \ldots, 12\}$





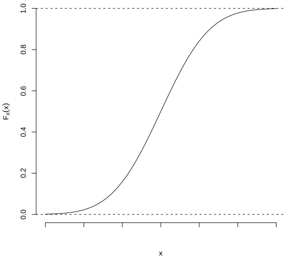
Bernoulli random variable

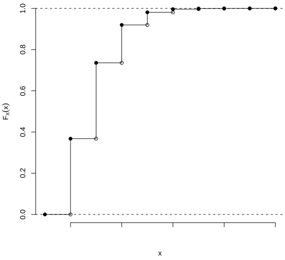
- Named after the Swiss scientist Jakob Bernoulli (1654-1705).
- That is the simplest random variable
- It can only assume 2 values, $S_X = \{0, 1\}$
- Its pmf is given by

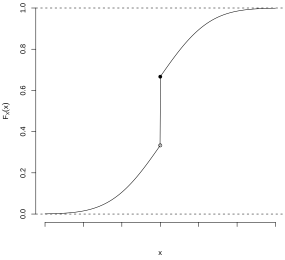
$$p(1) = \pi$$
$$p(0) = 1 - \pi$$

  for some value π
- It is often used to characterise the occurrence/non-occurrence of a given event, or the presence/absence of a given feature

Cumulative distribution functions







Continuous distribution
$\rightarrow$ continuous r.v.

Discrete distribution
$\rightarrow$ discrete r.v.

Hybrid distribution
$\rightarrow$ hybrid r.v.

Continuous random variables

As opposed to a discrete r.v., a continuous random variable X is expected to take on an uncountable number of values. S_X is therefore an uncountable set of real numbers (like an interval), and can even be $\mathbb{R}$ itself.

Definition

A random variable X is said to be continuous if there exists a nonnegative function $f(x)$ defined for all real $x \in \mathbb{R}$ such that for any set B of real numbers,

$$\mathbb{P}(X \in B) = \int_B f(x)dx$$

Consequence: $\mathbb{P}(X = x) = 0$ for any x !

$\rightarrow$ The probability mass function is useless

$\rightarrow$ The probability density function (pdf) $f(x)$ will play the central role

Continuous random variables: remark

Note 1: the fact that $\mathbb{P}(X = x) = 0$ for any x should not be disturbing

$\rightarrow$ coherent when dealing with measurements,

E.g. if we report a temperature of 74.8 degrees centigrade, owing to the limits of our ability to measure (accuracy of measuring devices), we really mean that the temperature lies "close to" 74.8, for instance between 74.75 and 74.85 degrees

Note 2: when we say that there is a zero probability that a random variable X will take on any value x, this does not mean that it is impossible that X will take on the value x!

In the continuous case, zero probability does not imply logical impossibility

$\rightarrow$ this should not be disturbing either, as we are always interested in probabilities connected with intervals and not with isolated points

Probability density function: properties

- $F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(y)dy$, that is

$$\boxed{f(x) = \frac{dF(x)}{dx} = F'(x)}$$

(wherever F is differentiable)

- $\boxed{f(x) \geq 0} \quad \forall x \in \mathbb{R} \qquad (F(x) \text{ is nondecreasing})$
- $\mathbb{P}(a \leq X \leq b) = \int_{a}^{b} f(x)dx = F(b) - F(a)$
- $\boxed{\int_{-\infty}^{+\infty} f(x)dx = 1}$
- For a small ε, $\mathbb{P}(x - \varepsilon/2 \leq X \leq x + \varepsilon/2) = \int_{x-\varepsilon/2}^{x+\varepsilon/2} f(y)dy \simeq \varepsilon f(x)$

Note: as $\mathbb{P}(X = x) = 0$, $\mathbb{P}(X < x) = \mathbb{P}(X \leq x)$ (for a continuous r.v.)

Continuous random variables: examples

Examples of continuous random variables include: electrical current, length, pressure, temperature, time, voltage, weight, speed of a car, amount of alcohol in a person's blood, efficiency of solar collector, strength of a new alloy, ...

→ Continuous random variables generally arise when we measure things

Example

Let X denote the current measured in a thin copper wire (in mA). Assume that the pdf of X is

$$f(x) = \begin{cases} C(4x - 2x^2) & \text{if } 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

What is the value of C ? Find $\mathbb{P}(X > 1.8)$

We must have $\int_{-\infty}^{+\infty} f(x)\,dx = 1$, so $C\int_0^2 (4x - 2x^2)\,dx = C \times \frac{8}{3} = 1$, that is $C = 3/8$

Then, $\mathbb{P}(X > 1.8) = \int_{1.8}^{+\infty} f(x)\,dx = 3/8 \times \int_{1.8}^{2} (4x - 2x^2)\,dx = 0.028.$

Discrete vs. Continuous random variables

Discrete r.v. | **Continuous r.v.**

Domain of variation

Discrete: $S_X = \{x_1, x_2, \dots\}$

Continuous: $S_X = [\alpha, \beta] \subseteq \mathbb{R}$

Probability mass function (pmf)

Discrete:

$p(x) = \mathbb{P}(X = x) \geq 0$ for all $x \in \mathbb{R}$

- $p(x) > 0$ if and only if $x \in S_X$
- $\sum_{x \in S_X} p(x) = 1$

Continuous: useless: $p(x) \equiv 0$

Probability density function (pdf)

Discrete: does not exist

Continuous:

$f(x) = F'(x) \geq 0$ for all $x \in \mathbb{R}$

- $f(x) > 0$ if and only if $x \in S_X$
- $\int_{x \in S_X} f(x) = 1$

Note the similarity between the conditions for pmf and pdf.

Parameters of a distribution

Fact

Some quantities characterise a random variable more usefully (although incompletely) than the whole cumulative distribution function.

$\rightarrow$ The focus is on certain general properties of the distribution of the r.v.

The two most important such quantities are:

- the expectation (or mean) and
- the variance

of a random variable

Often, we talk about the expectation or the variance of a distribution, understood as the expectation or the variance of a random variable having that distribution.

Expectation

The **expectation** or the **mean** of a random variable X, denoted $\mathbb{E}(X)$ or μ, is defined by

Discrete r.v.

$$\mu = \mathbb{E}(X) = \sum_{x \in S_X} x\, p(x)$$

Continuous r.v.

$$\mu = \mathbb{E}(X) = \int_{S_X} x\, f(x) dx$$

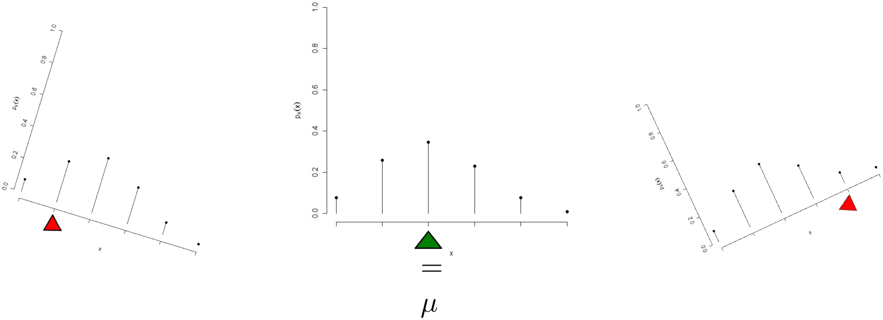
$\Rightarrow$ $\mathbb{E}(X)$ is a weighted average of the possible values of X, each value being weighted by the probability that X assumes it

Note: $\mathbb{E}(X)$ has the same units as X.

Expectation

Expectation = expected value, mean value, average value of X
= "central" value, around which X is distributed
= "centre of gravity" of the distribution

In the discrete case:



$\rightarrow$ localisation parameter

Expectation: examples

Example 1

What is the expectation of the outcome when a fair die is rolled?

$X =$ outcome, $S_X = \{1, 2, 3, 4, 5, 6\}$ with $p(x) = 1/6$ for any $x \in S_X$

$$\begin{aligned}\mu = \mathbb{E}(X) &= 1 \times 1/6 + 2 \times 1/6 + 3 \times 1/6 + 4 \times 1/6 + 5 \times 1/6 + 6 \times 1/6 \\ &= 3.5\end{aligned}$$

$\to \mu$ need not be a possible outcome !
$\to \mu$ is not the most likely outcome (this is called the mode)

Example 2

What is the expected sum when two fair dice are rolled?

X = sum of the two dice,
$S_X = \{2, 3, \ldots, 12\}$ with
$p(x) = (6 - |7 - x|)/36$ for any $x \in S_X$

$\rightarrow \mu = \mathbb{E}(X) = 2 \times 1/36 + 3 \times 2/36 + \ldots + 12 \times 1/36 = 7$

Example 3: Bernoulli r.v. (see Slide 13)

What is the expectation of a Bernoulli r.v.?

$\mathbb{E}(X) = 0 \times (1 - \pi) + 1 \times \pi = \pi$

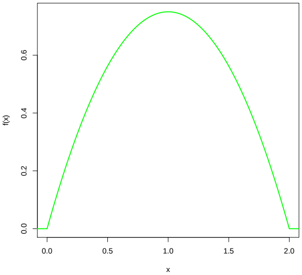
Expectation: examples

Example 4

Find the mean value of the copper current measurement X for Example on Slide 18, that is, with

$$f(x) = \begin{cases} \frac{3}{8}(4x - 2x^2) & \text{if } 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

The density is



By symmetry, it can be directly concluded that $\mu = 1$ mA

It can also be easily checked that

$$\mu = \mathbb{E}(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx$$
$$= \frac{3}{8}\int_0^2 x\,(4x - 2x^2)\, dx$$
$$= 1$$

Expectation of a function of a random variable

Sometimes we are not interested in the expected value of X, but in the expected value of a function of X, say $g(X)$.

There is actually no need for explicitly deriving the distribution of $g(X)$. Indeed, it can be shown

If X is a discrete r.v.

$$\mathbb{E}(g(X)) = \sum_{x \in S_X} g(x)\, p(x)$$

If X is a continuous r.v.

$$\mathbb{E}(g(X)) = \int_{S_X} g(x)\, f(x)\, dx$$

In particular, for 2 constants a and b:

Linear transformation

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$$

With $a = 0 \to \mathbb{E}(b) = b$ (“degenerate” random variable)

Variance of a random variable

Definition

The **variance** of a random variable X, usually denoted by $\mathbb{V}\text{ar}(X)$ or σ^2, is defined by

$$\mathbb{V}\text{ar}(X) = \mathbb{E}\left((X-\mu)^2\right)$$

Clearly, $\boxed{\mathbb{V}\text{ar}(X) \geq 0}$

If X is a discrete r.v.

$$\sigma^2 = \mathbb{V}\text{ar}(X) = \sum_{x \in S_X} (x-\mu)^2 p(x)$$

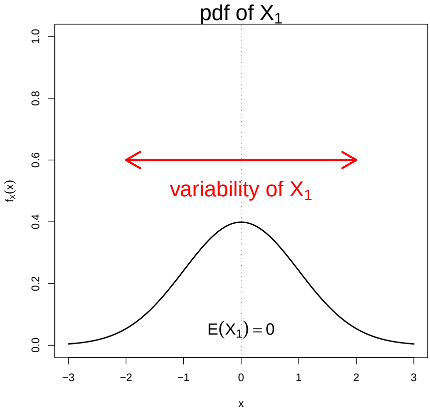
If X is a continuous r.v.

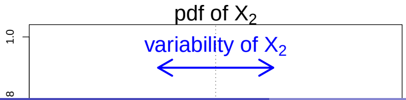
$$\sigma^2 = \mathbb{V}\text{ar}(X) = \int_{S_X} (x-\mu)^2 f(x) dx$$

- $\rightarrow$ Expected square of the deviation of X from its expected value
- $\rightarrow$ The variance quantifies the dispersion of the possible values of X around the "central" value μ, that is, the variability of X

Variance: illustration

Two random variables X_1 and X_2, with $\mathbb{E}(X_1) = \mathbb{E}(X_2)$





Variance: notes

Note 1

An alternative formula for $\mathbb{V}\text{ar}(X)$ is the following:

$$\sigma^2 = \mathbb{V}\text{ar}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \mathbb{E}(X^2) - \mu^2$$

Proof: ... $\square$

$\Rightarrow$ In practice, this is often the easiest way to compute $\mathbb{V}\text{ar}(X)$, using

$$\mathbb{E}(X^2) = \sum_{x \in S_X} x^2 p(x) \quad \text{or} \quad \mathbb{E}(X^2) = \int_{S_X} x^2 f(x) dx$$

Note 2

The variance σ^2 is not in the same units as X, which may make interpretation difficult.

$\Rightarrow$ often, we adjust for this by taking the square root of σ^2

This is called the **standard deviation** σ of X: $\sigma = \sqrt{\sigma^2} = \sqrt{\mathbb{V}\text{ar}(X)}$

Variance: linear transformation

A useful identity is that, for any constants a and b, we have

Linear transformation

$$\mathbb{V}\mathrm{ar}(aX + b) = a^2\, \mathbb{V}\mathrm{ar}(X)$$

Take $a = 1$, it follows that for any b, $\boxed{\mathbb{V}\mathrm{ar}(X + b) = \mathbb{V}\mathrm{ar}(X)}$

$\rightarrow$ variance not affected by translation

Take $a = 0$, if follows that for any b, $\boxed{\mathbb{V}\mathrm{ar}(b) = 0}$

("degenerate" random variable)

Variance : examples

Example 1

What is the variance of the number of points shown when a fair die is rolled?

X = outcome, $S_X = \{1,2,3,4,5,6\}$ with $p(x) = 1/6$ for any $x \in S_X$

$$\begin{aligned}\mathbb{E}(X^2) &= 1^2 \times 1/6 + 2^2 \times 1/6 + 3^2 \times 1/6 + 4^2 \times 1/6 + 5^2 \times 1/6 + 6^2 \times 1/6 \\ &= 91/6\end{aligned}$$

We know that $\mu = 3.5$ (Slide 23), so that

$$\sigma^2 = \mathbb{E}(X^2) - \mu^2 = 91/6 - 3.5^2 \simeq 2.92$$

The standard deviation is $\sigma = \sqrt{2.92} \simeq 1.71$

Example 2

What is the variance of the sum of the points when 2 fair dice are rolled ?

(Exercise) Check that $\sigma^2 \simeq 5.83$, $\sigma \simeq 2.41$.

Variance: examples

Example 3

What is the variance of a Bernoulli r.v.?

$\mathbb{E}(X^2) = 0^2 \times (1-\pi) + 1^2 \times \pi = \pi = \mathbb{E}(X)$

$\to \mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \pi - \pi^2 = \pi(1-\pi)$

Example 4

What is the variance of the copper current measurement X for Example on Slide 18, that is, with

$$f(x) = \begin{cases} \frac{3}{8}(4x - 2x^2) & \text{if } 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

We have $\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\,dx = \frac{3}{8}\int_0^2 x^2(4x - 2x^2)\,dx = 1.2$

We know that $\mu = 1$ (Slide 24), so that $\sigma^2 = 1.2 - 1^2 = 0.2$ mA^2

$\to \sigma \simeq 0.45$ mA

Standardisation

Standardisation is a very useful linear transformation.

Suppose you have a random variable X with mean μ and variance σ^2. Then, the associated standardised random variable, often denoted Z, is given by

$$Z = \frac{X - \mu}{\sigma},$$

that is, $Z = \frac{1}{\sigma}X - \frac{\mu}{\sigma}$. Hence, using the linear transformations properties,

$$\mathbb{E}(Z) = \frac{1}{\sigma}\mathbb{E}(X) - \frac{\mu}{\sigma} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0$$

$$\mathbb{V}\text{ar}(Z) = \frac{1}{\sigma^2}\,\mathbb{V}\text{ar}(X) = \frac{\sigma^2}{\sigma^2} = 1$$

$\rightarrow$ A standardised random variable has always mean 0 and variance 1.

Note 1: Z is a dimensionless variable (no unit)
Note 2: A standardised value of X is sometimes called z-score

Joint distribution function

Often, probability statements concerning two random variables, say X and Y, defined on the same sample space are of interest:

$$\omega \to (X(\omega), Y(\omega)))$$

$\to$ These two variables are most certainly related

$\to$ They should be jointly analysed, in order to understand the degree of relationship between them

For instance, we may simultaneously measure the weight and hardness of a rock, the pressure and temperature of a gas, thickness and compressive strength of a piece of glass, etc.

Definition

The joint cumulative distribution function of X and Y is given by

$$F_{XY}(x, y) = \mathbb{P}(X \leq x, Y \leq y) \qquad \forall (x, y) \in \mathbb{R} \times \mathbb{R}$$

Note: $(X \leq x, Y \leq y)$ is the usual notation for $(X \leq x) \cap (Y \leq y)$.

Joint distribution: discrete case

If X and Y are both discrete, the joint probability mass function is defined by

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

The marginal pmf of X and Y can be obtained by

$$p_X(x) = \sum_{y \in S_Y} p_{XY}(x, y) \quad \text{and} \quad p_Y(y) = \sum_{x \in S_X} p_{XY}(x, y)$$

Joint distribution: continuous case

Definition

X and Y are said to be jointly continuous if there exists a function $f_{XY}(x,y) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$ such that for any sets A and B of real numbers

$$\mathbb{P}(X \in A, Y \in B) = \int_A \int_B f_{XY}(x,y)dy\,dx$$

The function $f_{XY}(x,y)$ is the joint probability density of X and Y.

The marginal densities follow from

$$\int_A f_X(x)dx = \mathbb{P}(X \in A) = \mathbb{P}(X \in A, Y \in S_Y) = \int_A \int_{S_Y} f_{XY}(x,y)dy\,dx$$

Thus,

$$f_X(x) = \int_{S_Y} f_{XY}(x,y)dy \quad \text{and} \quad f_Y(y) = \int_{S_X} f_{XY}(x,y)dx$$

Expectation of a function of two random variables

For any function $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, the expectation of $g(X, Y)$ is given by

$$\mathbb{E}(g(X,Y)) = \sum_{x \in S_X} \sum_{y \in S_Y} g(x,y) p_{XY}(x,y) \qquad \text{(discrete case)}$$

$$= \int_{S_X} \int_{S_Y} g(x,y) f_{XY}(x,y) dy\, dx \qquad \text{(continuous case)}$$

Expectation of a function of two random variables

For instance, in the continuous case,

$$\begin{aligned}
\mathbb{E}(aX+bY) &= \int_{S_X}\int_{S_Y}(ax+by)f_{XY}(x,y)dy\,dx \\
&= \int_{S_X}\int_{S_Y} ax\,f_{XY}(x,y)dy\,dx + \int_{S_X}\int_{S_Y} by\,f_{XY}(x,y)dy\,dx \\
&= a\int_{S_X} x\int_{S_Y} f_{XY}(x,y)dy\,dx + b\int_{S_Y} y\int_{S_X} f_{XY}(x,y)dx\,dy \\
&= a\int_{S_X} xf_X(x)dx + b\int_{S_Y} yf_Y(y)dy \\
&= a\mathbb{E}(X) + b\mathbb{E}(Y)
\end{aligned}$$

Example

What is the expected sum obtained when two fair dice are rolled?

Let X be the sum and X_i the value shown on the ith die. Then, $X = X_1 + X_2$, and

$$\mathbb{E}(X) = \mathbb{E}(X_1) + \mathbb{E}(X_2) = 2 \times 3.5 = 7$$

Independent random variables

Definition

The random variables X and Y are said to be independent if, for all $(x, y) \in \mathbb{R} \times \mathbb{R}$,

$$\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x) \times \mathbb{P}(Y \leq y)$$

In other words, X and Y are independent if all couples of events $(X \leq x)$ and $(Y \leq y)$ are independent.

Characterisation: For any $(x, y) \in \mathbb{R} \times \mathbb{R}$,

$$F_{XY}(x, y) = F_X(x) \times F_Y(y),$$

which reduces to

$$p_{XY}(x, y) = p_X(x) \times p_Y(y) \qquad \text{(discrete case)}$$

or

$$f_{XY}(x, y) = f_X(x) \times f_Y(y) \qquad \text{(continuous case)}$$

Independent random variables

Property

If X and Y are independent, then for any functions h and g,

$$\mathbb{E}(h(X)g(Y)) = \mathbb{E}(h(X)) \times \mathbb{E}(g(Y))$$

Proof (in the continuous case):

$$\begin{aligned}
\mathbb{E}(h(X)g(Y)) &= \iint_{S_X \times S_Y} h(x)g(y)f_{XY}(x,y)dy\,dx \\
&= \int_{S_X} \int_{S_Y} h(x)g(y)f_X(x)f_Y(y)dy\,dx \\
&= \int_{S_X} h(x)f_X(x)dx \times \int_{S_Y} g(y)f_Y(y)dy \\
&= \mathbb{E}(h(X)) \times \mathbb{E}(g(Y))
\end{aligned}$$

Covariance of two random variables

Definition

The covariance of two random variables X and Y is defined by

$$\mathbb{C}\mathrm{ov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$$

Properties:

- $\mathbb{C}\mathrm{ov}(X, Y) = \mathbb{C}\mathrm{ov}(Y, X)$
- $\mathbb{C}\mathrm{ov}(X, X) = \mathbb{V}\mathrm{ar}(X)$
- $\boxed{\mathbb{C}\mathrm{ov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}$
- $\mathbb{C}\mathrm{ov}(aX + b, cY + d) = ac\,\mathbb{C}\mathrm{ov}(X, Y)$
- $\mathbb{C}\mathrm{ov}(X_1 + X_2, Y_1 + Y_2)$
  $= \mathbb{C}\mathrm{ov}(X_1, Y_1) + \mathbb{C}\mathrm{ov}(X_1, Y_2) + \mathbb{C}\mathrm{ov}(X_2, Y_1) + \mathbb{C}\mathrm{ov}(X_2, Y_2)$

Note: unit of $\mathbb{C}\mathrm{ov}(X, Y)$ = unit of $X \times$ unit of Y

Covariance: interpretation

Suppose X and Y are two Bernoulli random variables.

Then, XY is also a Bernoulli random variable which takes the value 1 if and only if $X = 1$ and $Y = 1$. It follows:

$$\mathbb{C}\text{ov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{P}(X = 1, Y = 1) - \mathbb{P}(X = 1)\mathbb{P}(Y = 1)$$

Then,

$$\begin{aligned}
\mathbb{C}\text{ov}(X, Y) > 0 &\Leftrightarrow \mathbb{P}(X = 1, Y = 1) > \mathbb{P}(X = 1)\mathbb{P}(Y = 1) \\
&\Leftrightarrow \frac{\mathbb{P}(X = 1, Y = 1)}{\mathbb{P}(X = 1)} > \mathbb{P}(Y = 1) \\
&\Leftrightarrow \mathbb{P}(Y = 1 | X = 1) > \mathbb{P}(Y = 1)
\end{aligned}$$

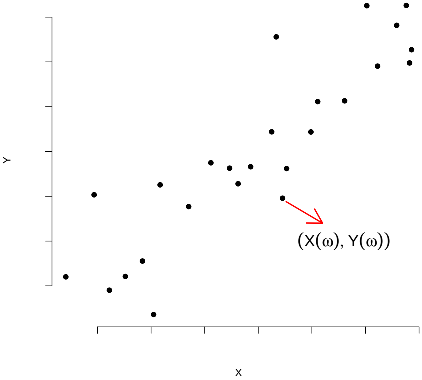
$\rightarrow$ The outcome $X = 1$ makes it more likely that $Y = 1$

$\rightarrow$ Y tends to increase when X does, and vice-versa

This result holds for any r.v. X and Y (not only Bernoulli r.v.).

Covariance: interpretation

- $\mathbb{C}\mathrm{ov}(X, Y) > 0 \to X$ and Y tend to increase or decrease together
-
-



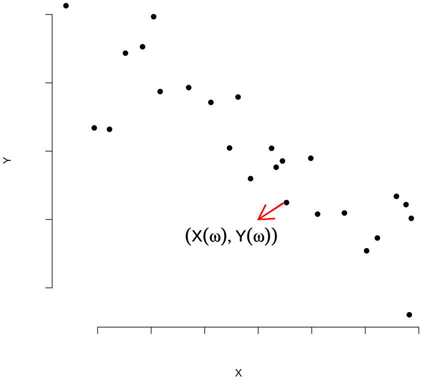
Fact

$$X \text{ and } Y \text{ independent } \Rightarrow \mathbb{C}\mathrm{ov}(X, Y) = 0$$

$$\mathbb{C}\mathrm{ov}(X, Y) = 0 \not\Rightarrow X \text{ and } Y \text{ independent}$$

Covariance: interpretation

-
- $\mathbb{C}\text{ov}(X, Y) < 0 \rightarrow X$ tends to increase as Y decreases and vice-versa
-

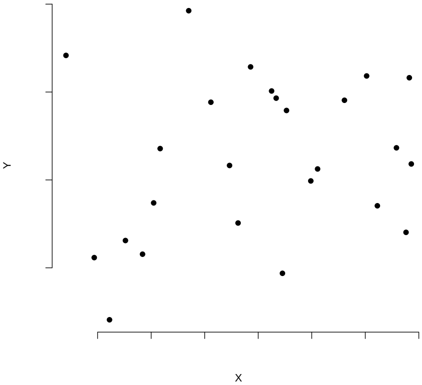


Fact

$$X \text{ and } Y \text{ independent } \Rightarrow \mathbb{C}\text{ov}(X, Y) = 0$$
$$\mathbb{C}\text{ov}(X, Y) = 0 \nRightarrow X \text{ and } Y \text{ independent}$$

Covariance: interpretation

-
-
- $\mathbb{C}\mathrm{ov}(X, Y) = 0 \rightarrow$ no linear association between X and Y



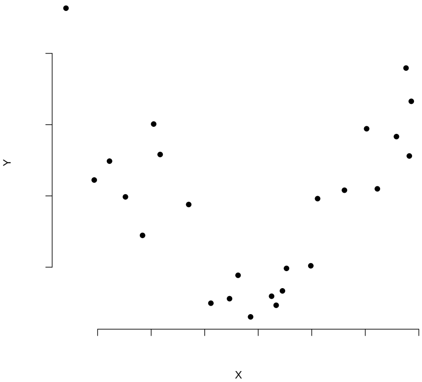
Fact

$$X \text{ and } Y \text{ independent } \Rightarrow \mathbb{C}\mathrm{ov}(X, Y) = 0$$

$$\mathbb{C}\mathrm{ov}(X, Y) = 0 \not\Rightarrow X \text{ and } Y \text{ independent}$$

Covariance: interpretation

-
-
- $\mathbb{C}\text{ov}(X, Y) = 0 \rightarrow$ no linear association between X and Y (doesn't mean there is no association!)



Fact

$$X \text{ and } Y \text{ independent } \Rightarrow \mathbb{C}\text{ov}(X, Y) = 0$$
$$\mathbb{C}\text{ov}(X, Y) = 0 \nRightarrow X \text{ and } Y \text{ independent}$$

Covariance: examples

Example

Let the pmf of a r.v. X be $p_X(1) = p_X(-1) = 1/2$ and $Y = X^2$. Find $\mathbb{C}\text{ov}(X, Y)$

Variance of a sum of random variables

From the properties of the covariance, it follows:

$$\begin{aligned}\mathbb{V}\mathrm{ar}\,(aX+bY) &= \mathbb{C}\mathrm{ov}\,(aX+bY, aX+bY) \\ &= \mathbb{C}\mathrm{ov}(aX, aX) + \mathbb{C}\mathrm{ov}(aX, bY) \\ &\qquad + \mathbb{C}\mathrm{ov}(bY, aX) + \mathbb{C}\mathrm{ov}(bY, bY) \\ &= \mathbb{V}\mathrm{ar}(aX) + \mathbb{V}\mathrm{ar}(bY) + 2\,\mathbb{C}\mathrm{ov}(aX, bY) \\ &= a^2\,\mathbb{V}\mathrm{ar}(X) + b^2\,\mathbb{V}\mathrm{ar}(Y) + 2ab\,\mathbb{C}\mathrm{ov}(X, Y)\end{aligned}$$

Now, if X and Y are independent random variables,

$$\mathbb{V}\mathrm{ar}\,(aX+bY) = a^2\,\mathbb{V}\mathrm{ar}(X) + b^2\,\mathbb{V}\mathrm{ar}(Y)$$

For instance, if X and Y are independent,

$$\mathbb{V}\mathrm{ar}(X+Y) = \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y)$$

$$\mathbb{V}\mathrm{ar}(X-Y) = \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y)$$

Example

Example

We have two scales for measuring small weights in a laboratory. Assume the true weight of an item is 2g. Both scales give readings which have mean 2g and variance $0.05g^2$. Compare using only one scale and using both scales then averaging the two measures in terms of the accuracy.

The first measure X has $\mathbb{E}(X) = 2$ and $\mathbb{V}\text{ar}(X) = 0.05$. Now, denote the second measure Y, independent of X, with $\mathbb{E}(Y) = 2$ and $\mathbb{V}\text{ar}(Y) = 0.05$. Then, take $W = \frac{X+Y}{2}$. We have

$$\mathbb{E}(W) = \frac{1}{2}\mathbb{E}(X) + \frac{1}{2}\mathbb{E}(Y) = \frac{2}{2} + \frac{2}{2} = 2 \text{ (g)}$$

and

$$\mathbb{V}\text{ar}(W) = \frac{1}{4} \times (\mathbb{V}\text{ar}(X) + \mathbb{V}\text{ar}(Y)) = \frac{1}{4} \times (0.05 + 0.05) = 0.025 \text{ (g}^2\text{)}$$

$\to$ Averaging 2 measures reduces the variance by 2.

Correlation

The covariance of two r.v. is important as an indicator of the relationship between them.

However, it heavily depends on units of X and Y (difficult interpretation, not scale-invariant).

$\rightarrow$ The correlation coefficient ρ is often used instead.

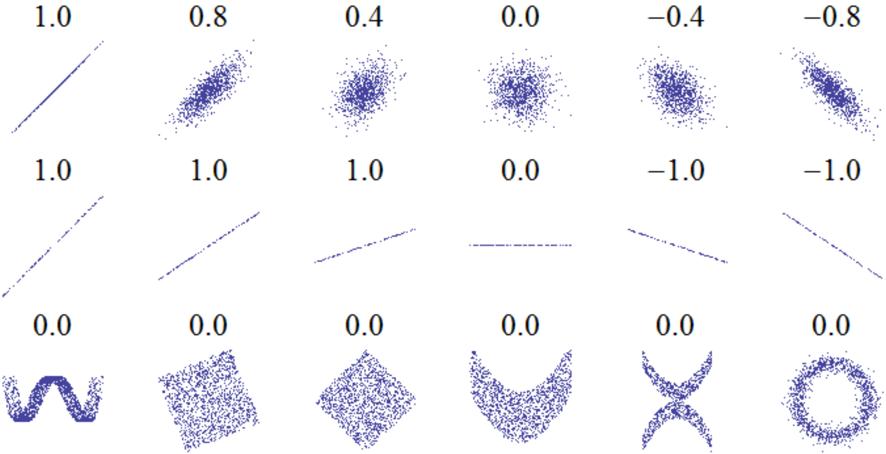
It is the covariance between the standardised versions of X and Y, or, explicitly,

$$\rho = \frac{\mathbb{C}\text{ov}(X, Y)}{\sqrt{\mathbb{V}\text{ar}(X)\,\mathbb{V}\text{ar}(Y)}}$$

Properties:

- ρ is dimensionless (no unit)
- ρ always has a value between -1 and 1.
- Positive (and negative) ρ means positive (and negative) linear relationship between X and Y
- The closer $|\rho|$ is to 1, the stronger is the linear relationship

Correlation examples



Objectives

Now you should be able to:

- understand the differences between discrete and continuous r.v. ☐
- for discrete r.v., determine probabilities from pmf and the reverse ☐
- for continuous r.v., determine probabilities from pdf and the reverse ☐
- for discrete r.v., determine probabilities from cdf and cdf from pmf and the reverse ☐
- for continuous r.v., determine probabilities from cdf and cdf from pdf and the reverse ☐
- calculate means and variances for both discrete and continuous random variables ☐
- use joint pmf and joint pdf to calculate probabilities ☐
- calculate and interpret covariances and correlations between two random variables ☐

Recommended exercises

$\rightarrow$ Q25 p.220, Q27 p.221, Q29&30 p.221, Q69 p.57, Q40&43 p.152, Q42 p.152, Q41 p.223, Q65 p.239 (2nd edition)

$\rightarrow$ Q27 p.225, Q29 p.225, Q31&32 p.225, Q71 p.59, Q42&45 p.157, Q44 p.157, Q43 p.227, Q67 p.243 (3rd edition)