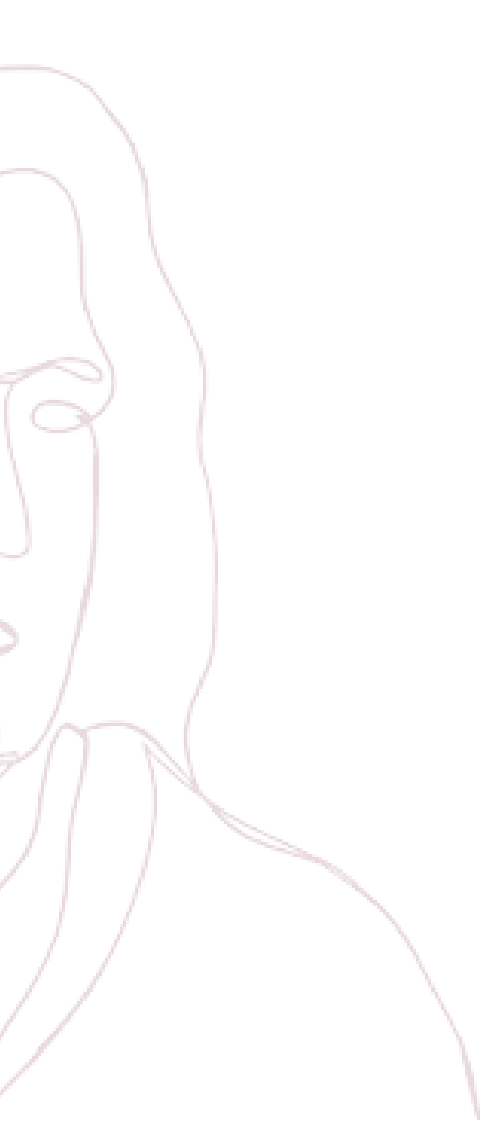




Random Variables

Statistics I

Disclaimer

These slides are a free translation and adaptation from the slide deck for Estatística I by Prof. Sandra Custódio and Prof. Teresa Ferreira from the Lisbon Accounting and Business School, Polytechnic University of Lisbon.

(Single) Random Variables

Random Variables

A random variable is a function that will allow us to quantify (transform into a number) each outcome.

Random Variable

A **random variable (r.v.)** X is a function $f : \Omega \rightarrow \Omega_X \subset \mathbb{R}$. Ω_X is known as the support of the r.v. X .

$$\omega \in \Omega \xrightarrow{X} X(\omega) \in \Omega_X \subset \mathbb{R}$$

$X(\omega)$ is the *image* under X of the outcome ω

Summarizing, a r.v. is a function that associates a real number to each outcome from Ω .

Random Variables

- $\{X = x\}$, $\{X \leq x\}$, $\{X > x\}$ are events
- $\{X = x\}$ happens when, from our experiment, we obtain ω such that $X(\omega) = x$. $\{X = x\} = \{\omega \in \Omega | X(\omega) = x\}$
- The probability of $X \leq x$ for example, is then

$$P(X \leq x) = P(\{\omega \in \Omega | X(\omega) \leq x\})$$

Types of random variables

Discrete Random Variable

X is a discrete r.v. when:

- The support of X , Ω_X , is finite or countable infinite.
- $P(\Omega_X) = 1$

In this case, $\Omega_X = \{x_1, x_2, \dots, x_n\}$ with $n \in \mathbb{N}$ if Ω_X is finite, and $\Omega_X = \{x_1, x_2, \dots, x_n, \dots\}$ if it is countable infinite.

Probability density function (*pdf*)

Let X be a discrete r.v. The *pdf* of X is a function $f_X : \mathbb{R} \rightarrow \mathbb{R}$ such that:

$$f_X(x) = \begin{cases} P(X = x) & , \text{ if } x \in \Omega_X \\ 0 & , \text{ if } x \in \mathbb{R} \setminus \Omega_X \end{cases}$$

Naturally, by construction the *pdf* satisfies the following properties:

$$f_X(x) \geq 0 \quad \forall x \in \mathbb{R}$$

$$\sum_{x_i \in \Omega_X} P(X = x_i) = 1$$

Probability density function (*pdf*)

The *pdf* gives the probability in a single *point*. The total probability is distributed among single points, x_i . A reasonable representation of a *pdf* of a discrete r.v. could be:

x	x_1	x_2	...	x_n	...
$f(x)$	p_1	p_2	...	p_n	...

Where $p_i = P(X = x_i)$

Example

Consider the discrete r.v. X with the following *pdf*:

x	0	1	2	3	4
$f(x)$	0.05	a	0.35	0.25	0.05

We could define:

- Support of X : $\Omega_X = \{0, 1, 2, 3, 4\}$
- $\sum_{x \in \Omega_X} f(x) = 1 \Rightarrow 0.05 + a + 0.35 + 0.25 + 0.05 = 1$ that is $a =$
- 0.3

Example

Our table is now:

x	0	1	2	3	4
$f(x)$	0.05	0.3	0.35	0.25	0.05

What is $P(X = 2|X \leq 3)$?

$$P(X = 2|X \leq 3) = \frac{P(X = 2 \cap X \leq 3)}{P(X \leq 3)} = \frac{P(X = 2)}{P(X \leq 3)}$$

$$= \frac{f(2)}{f(0) + \dots + f(3)} = \frac{0.35}{0.95} = 0.368$$

Continuous random variables

X is a continuous r.v. if:

- The support of X , Ω_X is uncountable infinite.
- $P(\Omega_X) = 1$
- $P(X = x) = 0 \ \forall x \in \mathbb{R}$.

Probability density function (*pdf*)

Let X a continuous r.v.

There is a function $f_X : \mathbb{R} \rightarrow \mathbb{R}$, the *pdf* of X such that:

- $f_X(x) \geq 0, \forall x \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f_X(x)dx = 1$

Technically, from Measure Theory, we need an *absolutely continuous* r.v. to ensure the existence of a *pdf*. These issues are beyond the scope of this course. Just know that when we say *continuous* r.v. we mean *absolutely continuous* r.v.

Probability density function (*pdf*)

Note that this *pdf* allows to compute the probability of events $x \in (a, b]$:

$$P(a < X \leq b) = \int_a^b f_X(x) dx$$

Observe that if you would do $X = a$ you would get the integral from a to a , which makes $dx = 0$ and therefore the integral (and the probability) becomes 0.

Example

Let X be a continuous r.v. with the following *pdf*:

$$f(x) = \begin{cases} \theta x^2 & , 0 \leq x < 1 \\ 0 & , \mathbb{R} \setminus [0, 1) \end{cases}$$

- Support for X : $\Omega_X = [0, 1)$

- $\int_{-\infty}^{\infty} f(x) dx = 1 \Leftrightarrow \int_0^1 \theta x^2 dx = \left[\theta \frac{x^3}{3} \right]_0^1$

- $\theta \frac{1}{3} - \theta \frac{0}{3} = 1 \Leftrightarrow \theta = 3$

Cumulative distribution function (*cdf*)

Let X be a r.v. The **distribution function** $F_X : \mathbb{R} \rightarrow [0, 1]$, defined as:

$$F_X(x) = P(X \leq x)$$

F_X is unique.

Cumulative distribution function (*cdf*)

With a discrete r.v.

x	0	1	2	3	4
$f(x)$	0.05	0.3	0.35	0.25	0.05

$$F(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ 0.05 & 0 \leq x < 1 \\ 0.05 + 0.3 = 0.35 & 1 \leq x < 2 \\ 0.35 + 0.35 = .7 & 2 \leq x < 3 \\ 0.7 + 0.25 = .95 & 3 \leq x < 4 \\ 1 & x \geq 4 \end{cases}$$

Cumulative distribution function (*cdf*)

Let's revisit our previous example:

$$P(X = 2 | X \leq 3) = \frac{P(X = 2)}{P(X \leq 3)} =$$

$$\frac{F(2) - F(2^-)}{F(3)} = \frac{0.7 - 0.35}{0.95} = 0.368$$

Cumulative distribution function (*cdf*)

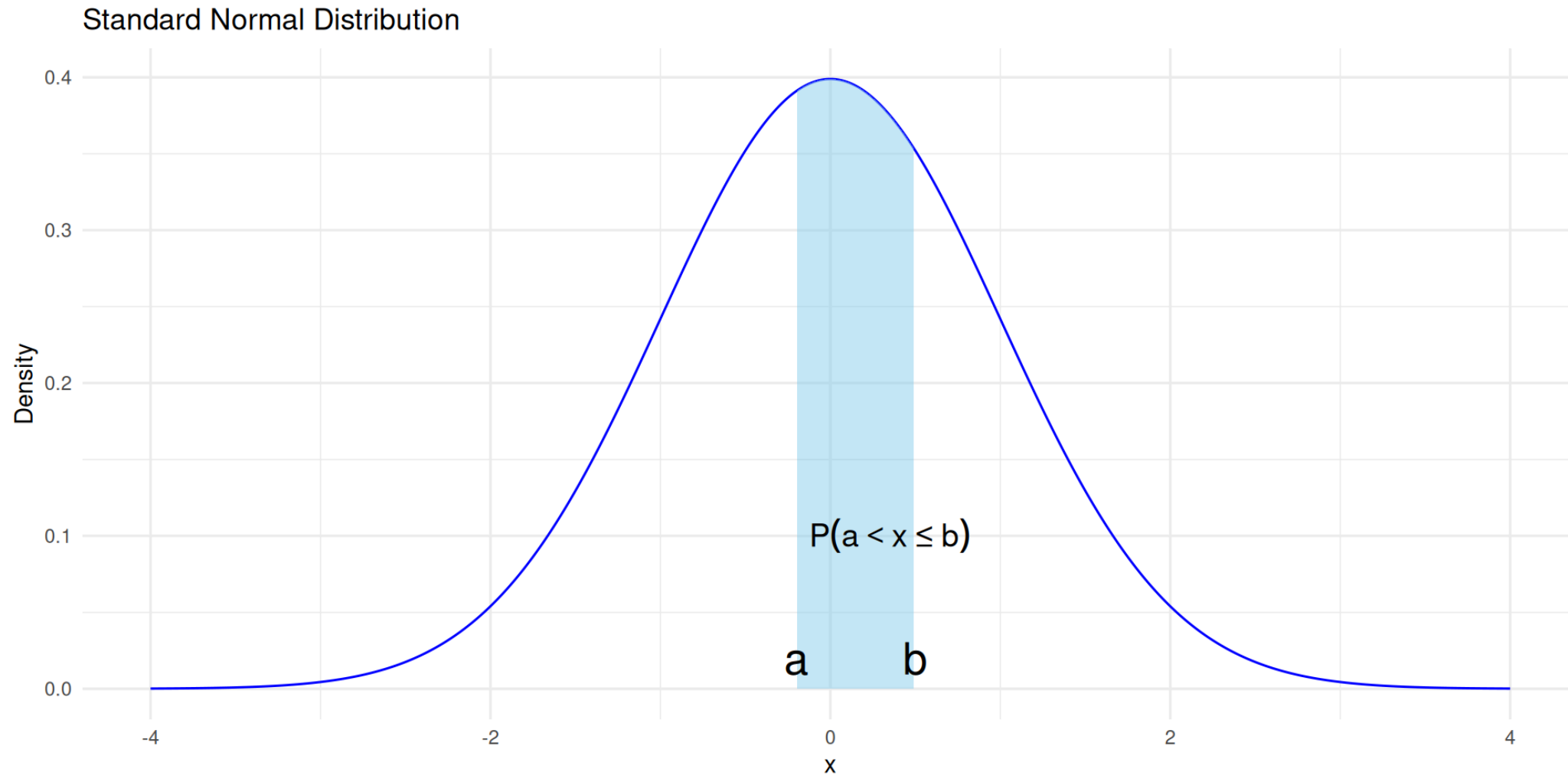
With a continuous r.v.:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(x)dx$$

The distribution function, F_X allows to compute the probability of $\{X \in (a, b]\}$

$$P(a < X \leq b) = \int_a^b f_X(x)dx = F_X(b) - F_X(a)$$

Example



Example

Consider the continuous r.v. defined previously, with the *pdf*:

$$f(x) = \begin{cases} 3x^2 & , 0 \leq x < 1 \\ 0 & , \mathbb{R} \setminus [0, 1) \end{cases}$$

Support for X : $\Omega_X = [0, 1)$

Distribution function (*cdf*):

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt = \begin{cases} 0 & , x < 0 \\ x^3 & , 0 \leq x < 1 \\ 1 & , x \geq 1 \end{cases}$$

Properties of the *cdf*

Nonetheless the r.v. is discrete or continuous, F_X has the following properties:

- $F_X : \mathbb{R} \rightarrow [0, 1]$
- F_X is continuous from the right: $\lim_{x \rightarrow a^+} F_X(x) = F(a)$
- F_X is monotone non-decreasing.
- $\lim_{x \rightarrow -\infty} F_X(x) = 0$
- $\lim_{x \rightarrow \infty} F_X(x) = 1$

Properties of the *cdf*

F_X for X r.v. discrete

- Single points, discontinuous support Ω_X
- It is continuous from the right
- $P(X < x) = F(X^-)$
- $P(X = x) = F(x) - F(x^-)$
- $P(a < X \leq b) = F(b) - F(a)$
- $P(a \leq X \leq b) = F(b) - F(a^-)$
- $P(a < X < b) = F(b^-) - F(a^-)$
- $P(a \leq X < b) = F(b^-) - F(a^-)$

Properties of the *cdf*

F_X for X r.v. continuous

- Continuous support Ω_X
- It is continuous in $\mathbb{R}$
- $P(X < x) = P(X \leq x) = F(x)$
- $P(a \square X \square b) = F(b) - F(a)$, [replace $\square$ for $<$ or $\leq$]

pdf and cdf

For a discrete r.v.

$$P(X = x) = F_X(x) - F_X(x^-)$$

$$+ \downarrow \uparrow -$$

$$F_X(x) = \sum_{x_i \leq x} P(X = x_i)$$

pdf and cdf

For a continuous r.v.

$$f_X(x) = \begin{cases} F'_X(x) & , x \in \mathbb{R} \text{ if } F'_X \text{ exists} \\ 0 & , \text{otherwise} \end{cases}$$

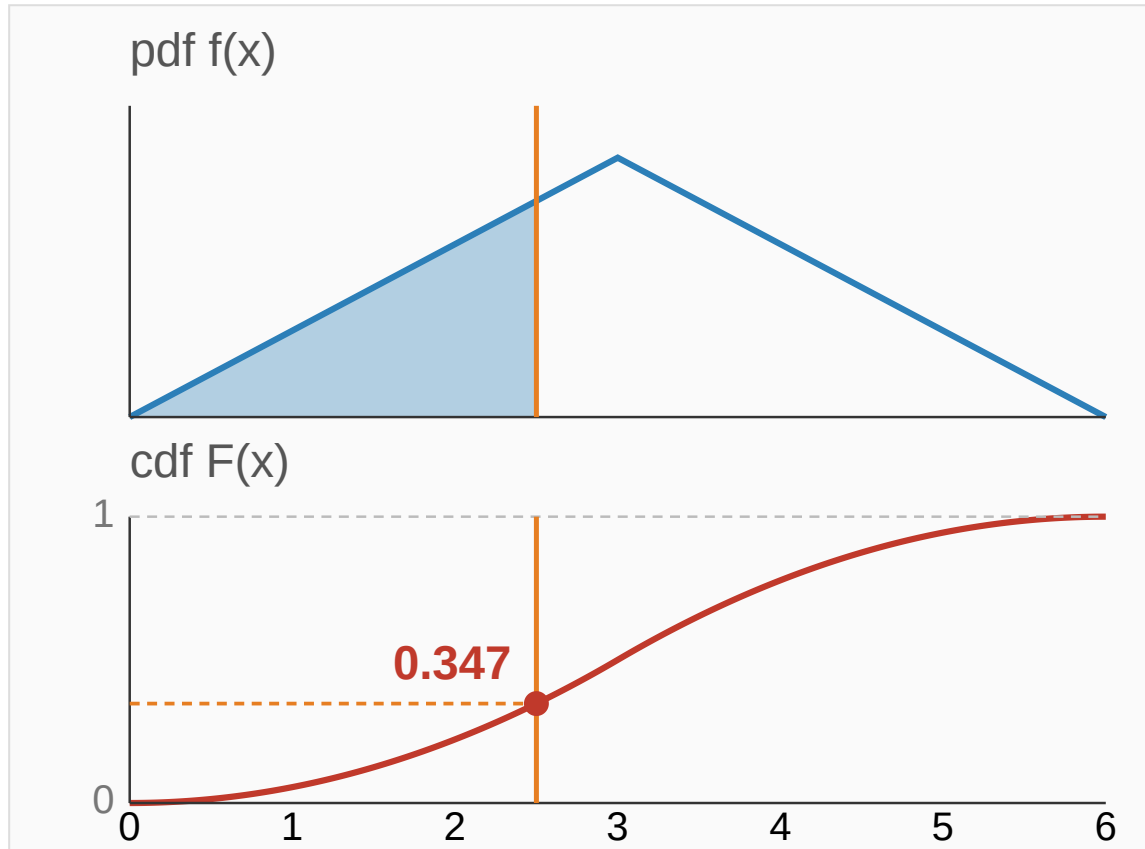
Derivative $\downarrow\uparrow$ *Primitive*

$$F_X(x) = \int_{-\infty}^x f_X(t) dt$$

The *pdf* of X , a continuous r.v. is not unique.

Reading one from the other

A continuous r.v. on $[0, 6]$. Move x : the **shaded area** under the *pdf* is exactly the **height** of the *cdf* at that point.



$x = 2.50$



shaded area = **0.347**

$F(2.50) = \mathbf{0.347}$

$P(X > 2.50) = 0.653$

$f(2.50) = 0.278$ (the slope)

What the chart showed

The *cdf* height **is** the *pdf* area. That is the whole content of $F_X(x) = \int_{-\infty}^x f_X(t)dt$.

Watch the steepness too: where the *pdf* is tall the *cdf* climbs fastest, and once the *pdf* returns to zero the *cdf* flattens at 1 and stays there. That is the other direction, $f_X(x) = F'_X(x)$.

The *pdf* height is **not** a probability: here it never exceeds $1/3$, and for other variables it can exceed 1 entirely. Only the **area** is a probability.

? Discrete and Continuous r.v. · Question 1

If X is a **continuous** random variable, then $P(X = a)$ equals:

A. $f_X(a)$

B. $F_X(a)$

C. 0

D. 1

☒ **C.** For a continuous r.v. the integral from a to a is zero. Probability lives on **intervals**, not on points.

? Discrete and Continuous r.v. · Question 2

The cumulative distribution function F_X is:

- A. decreasing
 - B. always continuous
 - C. non decreasing, going from 0 to 1
 - D. the same as the *pdf*
- ✓ **C.** $F_X(x) = P(X \leq x)$ can never decrease as x grows, and it goes from 0 to 1. For a discrete r.v. it is a step function, so not always continuous.



Discrete and Continuous r.v. · Question 3

Let X be continuous with $f(x) = 2x$ for $0 \leq x < 1$ and 0 otherwise.

Check that this is a valid *pdf*, find $F_X(x)$, then compute $P(X \leq 0.5)$ and $P(0.2 < X \leq 0.8)$.



Discrete and Continuous r.v. · Solution

Valid: $f(x) = 2x \geq 0$ on $[0, 1)$ and $\int_0^1 2x \, dx = [x^2]_0^1 = 1$ ✓

$F_X(x) = \int_0^x 2t \, dt = x^2$ for $0 \leq x < 1$ (and 0 below, 1 above).

$$P(X \leq 0.5) = 0.5^2 = 0.25$$

$$P(0.2 < X \leq 0.8) = F(0.8) - F(0.2) = 0.64 - 0.04 = 0.60$$

Statistical Moments of a Population

Describing a population

We could describe the range of X , a r.v., as a population, in the statistical sense, because it describes all the possible values it can take.

We can use numerical values to do so, which can represent dispersion or centrality of the data.

Expected Value or mean

The expected value or mean, is a location parameter for our r.v.

Definition

The **expected value** or **mean** of a random variable X is:

- $\mu_X = E[X] = \sum_{x \in \Omega_X} xP(X = x) < \infty$ if X is discrete
- $\mu_X = E[X] = \int_{-\infty}^{\infty} x f_X(x) dx < \infty$ if X is continuous.

Not all random variables have an expected value, it might be infinite.

Expected Value or mean

Let X, Y rvs, and $a, b \in \mathbb{R}$ scalars. Some properties of the mean

1. $E[a] = a$
2. $E[aX + bY] = aE[X] + bE[Y]$
3. If $Y = g(X)$, a r.v.:

- Discrete, then

$$E[Y] = \sum_{x \in \Omega_X} g(x)P(X = x) < \infty$$

- Continuous, then

$$E[Y] = \int_{-\infty}^{\infty} g(x)f_X(x)dx < \infty$$

Example discrete

Let X be a discrete r.v. as in the previous example:

x	0	1	2	3	4
$f(x)$	0.05	0.3	0.35	0.25	0.05

Let $g(X) = 2(X - 1)^2 + 3(X - 1) - 5$, find $E[g(X)]$.

Example discrete

$$g(X) = 2(X - 1)^2 + 3(X - 1) - 5$$

$$= 2(X^2 - 2X + 1) + 3X - 3 - 5$$

$$= 2X^2 - 4X + 2 + 3X - 8$$

$$= 2X^2 - X - 6$$

$$E[Y] = E[2X^2 - X - 6]$$

$$= 2E[X^2] - E[X] - 6$$

Example discrete

We only need to find $E[X]$ and $E[X^2]$ to obtain $E[g(X)]$.

$$E[X] = \sum xP(X = x)$$

$$= 0 \times .05 + 1 \times .3 + 2 \times .35 + 3 \times .25 + 4 \times .05 = 1.95$$

$$E[X^2] = \sum x^2P(X = x)$$

$$= 0 \times .05 + 1 \times .3 + 4 \times .35 + 9 \times .25 + 16 \times .05 = 4.75$$

$$E[g(X)] = 2 \times 4.75 - 1.95 - 6 = 1.55$$

Example continuous

Recall our example for continuous r.v.s. X :

$$f_X(x) = \begin{cases} 3x^2 & , 0 \leq x < 1 \\ 0 & , \mathbb{R} \setminus [0, 1) \end{cases}$$

Find $E[g(X)]$ when $g(X) = 2(X - 1)^2 + 3(X - 1) - 5$ We know already $g(X) = 2X^2 - X - 6$. Let's focus on $E[x]$ and $E[X^2]$.

Example continuous

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 x \times 3x^2 dx$$

$$= \int_0^1 3x^3 dx = \left[3 \frac{x^4}{4} \right]_0^1 = \frac{3}{4} = 0.75$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^1 x^2 \times 3x^2 dx$$

$$= \int_0^1 3x^4 dx = \left[3 \frac{x^5}{5} \right]_0^1 = \frac{3}{5} = 0.6$$

Example continuous

Finally,

$$E[g(x)] = 2 \times 0.6 - 0.75 - 6 = -5.55$$

p-quantile

The p-quantile, x_p , of a r.v. X is a location parameter, with fixed value.

p-quantile for discrete r.v.

x_p is the value for $x \in \Omega_X$ such that:

1. $P(X \leq x) \geq p$
2. $P(X \geq x) \geq 1 - p$

Or what is the same, $x \in \Omega_X$ such that $F_X(x^-) \leq p \leq F_X(x)$

p-quantile for continuous r.v.

x_p is an $x \in \Omega_X$ such that $F_X(x) = p$

Let's apply this for the examples we just used for the expected value.

Example - Discrete

Find the median (0.5-quantile) for X

$$F(x) = P(X \leq x) = \begin{cases} 0 & , x \leq 0 \\ 0.05 & , 0 \leq x < 1 \\ 0.35 & , 1 \leq x < 2 \\ 0.7 & , 2 \leq x < 3 \\ 0.95 & , 3 \leq x < 4 \\ 1 & , x \geq 4 \end{cases}$$

For example $F(2^-) = 0.35 \leq 0.5 \leq 0.7 = F(2)$ and therefore

$X_{0.5} = Me = 2$. Given that $E[X] = 1.95 < Me(X) = 2$ the distribution is slightly negatively (or left) skewed.

Example - Continuous

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ x^3 & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Let's find x such that $F(x) = 0.5$

$$F(x) = 0.5 \Leftrightarrow x^3 = 0.5 \Leftrightarrow x = \sqrt[3]{0.5} \approx 0.7937$$

And therefore, $x_{0.5} = Me = .7937$

Variance

Let X be a r.v. The **variance** of X , if it exists, is defined as:

$$V[X] = E \left[(X - E[X])^2 \right]$$

It can be shown, very easily, with some algebraic manipulation that

$$V[X] = E[X^2] - (E[X])^2$$

Variance

Remember that $E[X] \equiv \mu_X$

1. For a discrete r.v.:

$$V[X] = \sum_{x \in \Omega_X} (x - \mu_X)^2 P(X = x)$$

2. For a continuous r.v.:

$$V[X] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$$

Usually we write $V[X]$ as σ_X^2 .

Variance

Some properties for the variance:

1. $V[a] = 0$ for $a \in \mathbb{R}$
2. $V[aX + b] = a^2 V[X]$ for $a, b \in \mathbb{R}$ and X a r.v.
3. If X and Y are independent r.v. with finite variance, then

$$V[X \pm Y] = V[X] + V[Y]$$

Standard deviation σ_X

If σ_X^2 is the variance of X , then the standard deviation is known as:

$$\sigma_X = \sqrt{V[X]}$$

One characteristic of the standard deviation is that its units are the same as those of the random variable.

Coefficient of variation

While the variance and standard deviation allow us to measure the dispersion of the data, we might want to have it relative to the mean (a $\sigma_X = 1$ can be a lot for X taking relatively low values, but negligible if we are talking in millions!)

For that we use the coefficient of variation:

$$C.V._X = \frac{\sigma_X}{\mu_X} \times 100$$

Coefficient of variation

Some properties of the CV_X

1. It is not defined for $\mu_X = 0$
2. Lower values for CV_X means less dispersion around μ_X , therefore, more precision.
3. CV_X could be a measure for risk, if we are looking at returns for some asset.
4. CV_X is the relative weight of deviations from the mean, over the mean itself.

Example

Let's compute σ^2 , σ , and CV for our previous examples:

$$1. V[X] = E[X^2] - \mu_X^2 = 4.75 - 1.95^2 = 0.9475$$

$$2. \sigma = \sqrt{V[X]} = \sqrt{0.9475} = 0.97$$

$$3. CV = \frac{\sigma}{\mu} \times 100 = \frac{0.97}{1.95} \times 100 = 49.7\%$$

Values below 50% for CV allow us to see μ as representative for the data. The lower, the closer the data to μ and therefore the more representative it is.

Example

For the continuous r.v. case:

$$1. V[X] = E[X^2] - \mu_X^2 = 0.6 - 0.75^2 = 0.0375$$

$$2. \sigma = \sqrt{V[X]} = \sqrt{0.0375} = 0.1936$$

$$3. CV = \frac{\sigma}{\mu} \times 100 = \frac{0.1936}{0.75} \times 100 = 25.81\% < 50\%$$

? Moments of a Population · Question 1

For constants a and b , $V[aX + b]$ equals:

A. $a^2V[X]$

B. $aV[X] + b$

C. $a^2V[X] + b^2$

D. $aV[X]$

✓ **A.** Adding b shifts the whole distribution without changing its spread, so b drops out. Scaling by a scales the **squared** deviations by a^2 .

? Moments of a Population · Question 2

The coefficient of variation is useful because it:

A. is a pure number, measuring dispersion **relative** to the mean

B. keeps the units of X

C. is always below 1

D. equals the variance

 **A.** A σ of 1 is huge if the mean is 2, and negligible if the mean is a million. Dividing by the mean makes the comparison fair.



Moments of a Population · Question 3

The number of faulty items X in a batch has $f(0) = 0.5$, $f(1) = 0.3$, $f(2) = 0.15$, $f(3) = 0.05$.

Compute $E[X]$, $V[X]$, σ_X and the coefficient of variation.



Moments of a Population · Solution

$$E[X] = 0(0.5) + 1(0.3) + 2(0.15) + 3(0.05) = 0.75$$

$$E[X^2] = 0(0.5) + 1(0.3) + 4(0.15) + 9(0.05) = 1.35$$

$$V[X] = 1.35 - 0.75^2 = 0.7875 \cdot \sigma_X = \sqrt{0.7875} \approx 0.887$$

$CV = \frac{0.887}{0.75} \times 100 \approx 118\%$. Well above 50%, so the mean is **not** a representative summary of this distribution.

Random Pairs

Random pair

When running an experiment, it could be interesting to study the relationship between *two numeric features associated to each of the outcomes*.

Random pair

A random pair (X, Y) is a function $f_{X,Y} : \Omega \rightarrow (\Omega_X, \Omega_Y) \subset \mathbb{R}^2$. (Ω_X, Ω_Y) is known as the support of the random pair (X, Y) .

$$\omega \in \Omega \xrightarrow{(X,Y)} (X(\omega), Y(\omega)) \in (\Omega_X, \Omega_Y) \subset \mathbb{R}^2$$

$X(\omega)$ is the image, under X of outcome ω , and $Y(\omega)$ the image under Y of the same outcome.

Discrete random pair

A random pair (X, Y) is discrete when:

- The support (X, Y) , (Ω_X, Ω_Y) , is a finite or countable infinite of pairs.
- $P(\Omega_X, \Omega_Y) = 1$

Joint density function $f_{X,Y}$

Let (X, Y) a discrete random pair, the joint density function $f_{X,Y}(x, y)$ is a function $f_{X,Y} : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined as:

$$f_{X,Y}(x, y) = \begin{cases} P(X = x, Y = y) & , (x, y) \in (\Omega_X, \Omega_Y) \\ 0 & , (x, y) \in \mathbb{R}^2 \setminus (\Omega_X, \Omega_Y) \end{cases}$$

Joint density function $f_{X,Y}$

$f_{X,Y}$ satisfies the following properties:

1. $f_{X,Y}(x, y) \geq 0 \forall (x, y) \in \mathbb{R}^2$
2. $\sum_{x_i \in \Omega_X} \sum_{y_j \in \Omega_Y} P(X = x_i, Y = y_j) = 1 \quad \forall i, j = 1, 2, \dots$

A possible notation for $P(X = x_i, Y = y_j)$ is $p_{i,j}$

Joint density function $f_{X,Y}$

	y_1	y_2	$\dots$	y_j	$\dots$	
x_1	p_{11}	p_{12}	$\dots$	p_{1j}	$\dots$	$\sum_{j=1}^{\infty} p_{1j}$
x_2	p_{21}	p_{22}	$\dots$	p_{2j}	$\dots$	$\sum_{j=1}^{\infty} p_{2j}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
x_i	p_{i1}	p_{i2}	$\dots$	p_{ij}	$\dots$	$\sum_{j=1}^{\infty} p_{ij}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
	$\sum_{i=1}^{\infty} p_{i1}$	$\sum_{i=1}^{\infty} p_{i2}$	$\dots$	$\sum_{i=1}^{\infty} p_{ij}$	$\dots$	1

Marginal probability function

Given a random pair (X, Y) , the **marginal probability function** of X and Y is respectively:

- $f_X(x_i) = P(X = x_i) =$

$$\sum_{j=1}^{\infty} P(X = x_i, Y = y_j) = \sum_{j=1}^{\infty} p_{ij}$$

- $f_Y(y_j) = P(Y = y_j) =$

$$\sum_{i=1}^{\infty} P(X = x_i, Y = y_j) = \sum_{i=1}^{\infty} p_{ij}$$

For $i = 1, 2, \dots$ and $j = 1, 2, \dots$. Note that these functions have one dimension only.

Example

At SuperStore 🏬²⁴, three trained employees are qualified to operate the checkout counters, restock products on the shelves, and perform some administrative tasks. SuperStore has three checkout counters, and at least one of them must always be operating.

At any given day and moment when SuperStore is open to customers, consider the following random variables:

- X N of employees in the checkout counters 🛒💳 .
- Y N of employees restocking products on the shelves 📦 .

Example

The r.v. X has $\Omega_X = \{1, 2, 3\}$ and the following *pdf*:

x	1	2	3
$f_X(x)$	0.17	0.8	0.03

Consider the following table for the joint probability of (X, Y)

$X \setminus Y$	0	1	2	
1	a	$2b$	b	
2	0.1	c	0	
3	0.03	0	0	
				1

Example

1. If $P(X = 1, Y = 0) = 0.02$, find b and c

- Directly we get $a = 0.02$ as it is $P(X = 1, Y = 0)$
- Note that the top row is $f_Y(y)$ and the left column is $f_X(x)$
- Fill directly $f_X(x)$ with 0.17, 0.8, and 0.03.
- Fill $f_Y(y)$ with the summation of each column

Example

$X \setminus Y$	0	1	2	$f_X(x)$
1	0.02	$2b$	b	0.17
2	0.1	c	0	0.8
3	0.03	0	0	0.03
$f_Y(y)$	0.15	$2b + c$	b	1

- From first row: $0.02 + 2b + b = 0.17 \Rightarrow b = \frac{0.15}{3} = 0.05$
- From second row: $0.1 + c = 0.8 \Rightarrow c = 0.7$

Example

2. $P(X = 2|Y \geq 1)$ is approximately...?

- $$P(X = 2|Y \geq 1) = \frac{P(X = 2, Y \geq 1)}{P(Y \geq 1)}$$
- $$= \frac{P(X = 2, Y = 1) + P(X = 2, Y = 2)}{P(Y = 1) + P(Y = 2)}$$
- $$\frac{0.7 + 0}{0.8 + 0.05} = \frac{0.7}{0.85} \approx 0.8235$$

Independence of random variables (revisited)

Let (X, Y) a discrete random pair, X, Y are independent if, and only if:

$$P(X = x, Y = y) = P(X = x)P(Y = y) \quad \forall (x, y) \in \mathbb{R}^2$$

Equivalently: the joint *pdf* is the product of the marginal *pdfs*, and the joint *cdf* is the product of the marginal *cdfs*.

Example

3. ... (continued exercise) Are X and Y independent?

- $P(X = 2, Y = 2) = 0$
- $P(X = 2)P(Y = 2) = 0.8 \times 0.05 = 0.04$
- $0 \neq 0.04$
- X and Y are not independent.

Moments of a random pair

Definition

Let the discrete random pair (X, Y) have a joint *pdf* $P(X = x, Y = y)$ and a function $g : \mathbb{R}^2 \rightarrow \mathbb{R}$. The **expected value or mean** of $g(X, Y)$ is:

$$E[g(X, Y)] = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} g(x_i, y_j) P(X = x_i, Y = y_j)$$

If $g(x, y) = xy$, then $E[g(x, y)] = E[XY]$ and that equals

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} x_i y_j P(x = x_i, Y = y_j)$$

Moments of a random pair

Definition

Let the discrete random pair (X, Y) have a joint *pdf* $P(X = x, Y = y)$, and $\mu_X = E[X]$ and $\mu_Y = E[Y]$. The **covariance** between X and Y is:

$$\text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

Given that $E[(X - \mu_X)(Y - \mu_Y)]$ exists.

Note that this is equivalent to $\text{cov}(X, Y) = E[XY] - E[X]E[Y]$

Properties of the covariance

The covariance tries to capture how the two r.v. move together. If it is positive, it means that both tend to go in the same direction more often than not (both above or below their means at the same time). Being negative means that more often than not when one is above its mean, the other is below.

1. $\text{cov}(X, Y) = \text{cov}(Y, X)$
2. $\text{cov}(X, X) = V[X]$ and $\text{cov}(Y, Y) = V[Y]$ if $V[X]$ and $V[Y]$ exist.
3. $\text{cov}(a + bX, c + dY) = bd\text{cov}(X, Y)$ with $a, b, c, d \in \mathbb{R}$

Properties of the covariance

If X and Y are independent r.v. then $cov(X, Y) = 0$. Note that the opposite is not necessarily true, i.e. $cov(X, Y) = 0$ does not imply that X and Y are independent.

Another important identity with the covariance is the following:

$$V[X \pm Y] = V[X] + V[Y] \pm 2cov(X, Y)$$

Example

Knowing that $E[Y] = 0.9$, $cov(X, Y)$ is equal to? ...

- From the first table: $E[X] = 0.17 \times 1 + 0.8 \times 2 + 0.03 \times 3 = 1.86$

- $$E[XY] = \sum_x \sum_y xyP(X = x, Y = y) =$$

$$= 1 \times 0 \times 0.02 + 1 \times 1 \times 0.1 + 1 \times 2 \times 0.05 +$$

$$+ 3 \times 0 \times 0.1 + 2 \times 1 \times 0.7 + 2 \times 2 \times 0 +$$

$$+ 3 \times 0 \times 0.03 + 3 \times 1 \times 0 + 3 \times 2 \times 0 = 1.6$$

- $cov(X, Y) = E[XY] - E[X]E[Y] = 1.6 - 1.86 \times 0.9 = -0.074$

Correlation coefficient

A caveat of the covariance is that its units depends directly on the units of X and Y . The correlation coefficient allow us to express this relationship, between X and Y without being affected by the units in which these r.v. are measured.

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{V[X]V[Y]}} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Clearly $\rho \in [-1, 1]$. Note also that $|\rho| = 1$ if and only if $P(Y = a + bX) = 1$ with $a, b \in \mathbb{R}$. If X and Y are independent r.v. then $\rho = 0$.

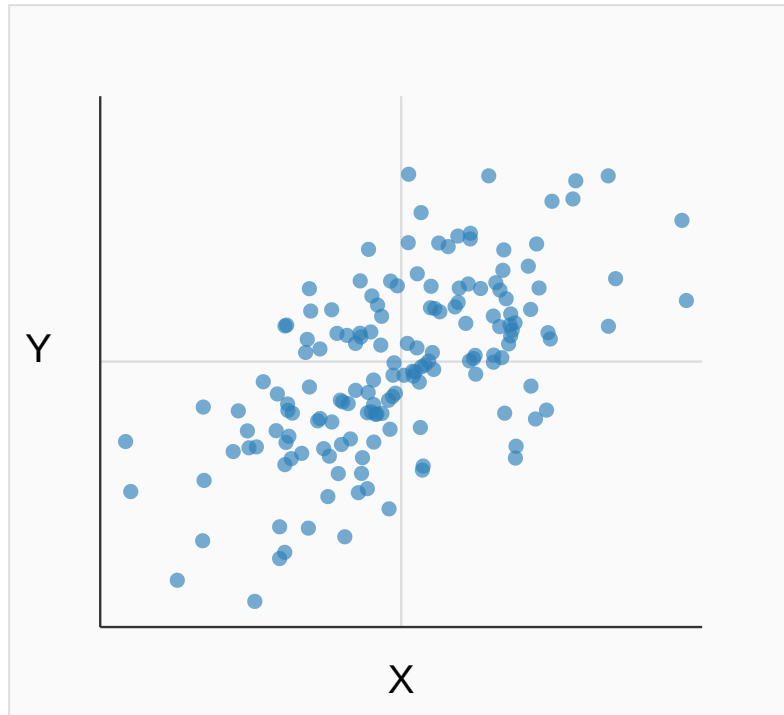
Correlation coefficient

Correlation coefficient	Correlation
$ \rho = 1$	Perfect
$0.8 \leq \rho < 1$	Strong
$0.5 \leq \rho < 0.8$	Moderate
$0.1 \leq \rho < 0.5$	Weak
$0 < \rho < 0.1$	Very weak
$\rho = 0$	None

Write *positive* or *negative* in front of correlation if $\rho > 0$ or $\rho < 0$ respectively.

What does ρ actually look like?

Same 160 observations every time. Only ρ changes. Match what you see against the table on the previous slide.



$\rho = 0.60$



positive moderate

$\text{cov}(X, Y) = 0.60$ (here $\sigma_X = \sigma_Y = 1$)

$\rho = 0.60$

ρ measures only the **linear** part of the relationship.

What the chart showed

At $\rho = 0$ the cloud is a shapeless blob; as $|\rho|$ grows it tightens onto a line, and at $|\rho| = 1$ every point sits exactly on it. The **sign** sets the tilt, the **magnitude** sets the tightness.

Because both variables here have $\sigma = 1$, covariance and correlation coincide. Rescale X into cents and the covariance changes by a factor of 100 while ρ does not move: that is exactly the unit problem ρ was built to solve.

$\rho = 0$ means **no linear relationship**, not “no relationship”. Points on a perfect parabola $Y = X^2$ have $\rho = 0$ and are entirely dependent.

Example

5. Based on the previous question, find the correlation coefficient between X and Y .

- From the marginal probability function we obtain $V[X]$ and $V[Y]$:

$$V[X] = 0.1804 \text{ and } V[Y] = 0.19$$

- Therefore,

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{-0.074}{\sqrt{0.1804} \sqrt{0.19}} = -0.3997$$

- We observe a weak negative linear correlation between X and Y .

? Random Pairs · Question 1

If X and Y are independent, then:

A. $\text{cov}(X, Y) = 1$

B. $V[X + Y] = V[X] - V[Y]$

C. $\rho_{XY} = 1$

D. $\text{cov}(X, Y) = 0$

✓ **D.** Independence implies $E[XY] = E[X]E[Y]$, so the covariance vanishes. Note also $V[X + Y] = V[X] + V[Y]$ in that case.

? Random Pairs · Question 2

“If $\text{cov}(X, Y) = 0$ then X and Y are independent.” This statement is:

- A. always true
 - B. false, zero covariance does not imply independence
 - C. true only for discrete r.v.
 - D. true only when $\rho = 1$
- ✓ **B.** Covariance only captures the **linear** association. Independence is strictly stronger, so the implication runs one way only.



Random Pairs · Question 3

The pair (X, Y) has $f_{X,Y}(0, 0) = 0.3$, $f_{X,Y}(0, 1) = 0.2$, $f_{X,Y}(1, 0) = 0.2$, $f_{X,Y}(1, 1) = 0.3$.

Find the marginals, $cov(X, Y)$ and ρ_{XY} . Are X and Y independent?



Random Pairs · Solution

Marginals: $P(X = 0) = P(X = 1) = 0.5$ and $P(Y = 0) = P(Y = 1) = 0.5$, so $E[X] = E[Y] = 0.5$ and $V[X] = V[Y] = 0.5 - 0.25 = 0.25$.

$E[XY] = 1 \times 1 \times 0.3 = 0.3$, so $cov(X, Y) = 0.3 - 0.25 = 0.05$ and $\rho_{XY} = \frac{0.05}{0.5 \times 0.5} = 0.2$.

Not independent: $f_{X,Y}(0, 0) = 0.3 \neq P(X = 0)P(Y = 0) = 0.25$.

Bibliography

- Figueiredo, F., Figueiredo, A., Ramos, A. & Teles, R. (2009). Estatística Descritiva e Probabilidades (2ª Edição). Escolar Editora.
- Murteira, B., Ribeiro, C.R., Silva, J.R. & Pimenta, C. (2007). Introdução à Estatística (2ª Edição). McGraw-Hill.
- Pestana, D. & Velosa, S.F. (2008). Introdução à Probabilidade e à Estatística (3ª Edição). Fundação Calouste Gulbenkian.
- Paulino, C.D. & Branco, J.A. (2005). Exercícios de Probabilidade e Estatística. Escolar Editora.