Random Variables

Random Variable: general idea

A random variable (RV) is a variable whose outcome depends on a random experiment.

This terminology emphasizes that the outcome varies from observation to observation according to random variation that can be summarized by probabilities.

Discrete RV: the possible outcomes are a set of separate values, such as a variable with possible values 0, 1, 2, ...

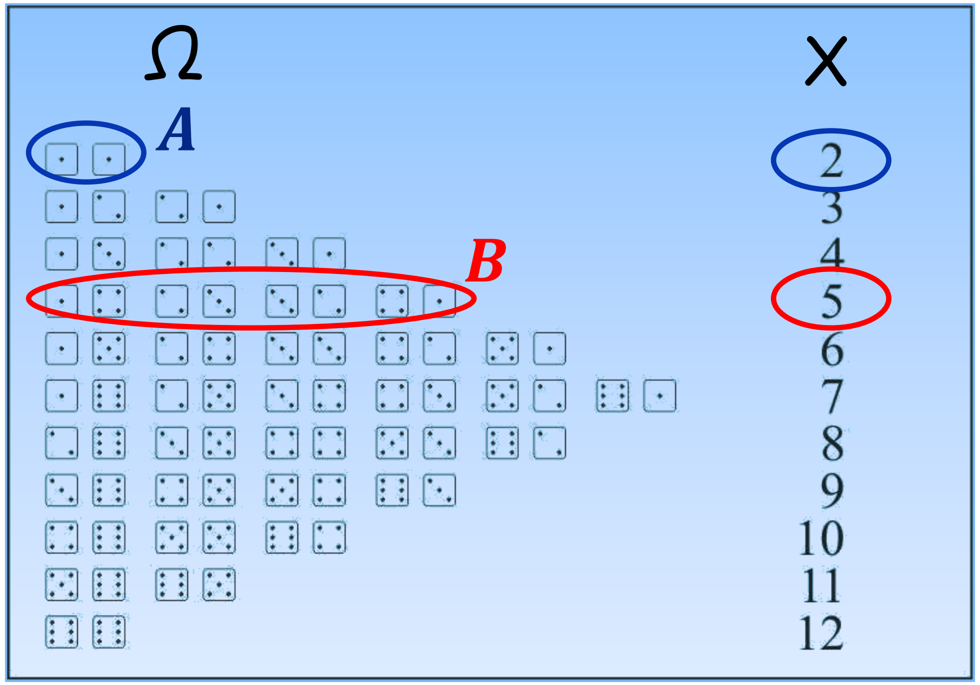
Continuous RV: the possible outcomes are an infinite continuum, such as all the real numbers

Example: discrete random variable

The values of a discrete random variable are discrete numbers.

For instance, the sum of the number obtained by rolling two dice and summing the numbers reported on their face.

Each outcome occurs with a certain probability given by the probability of the corresponding events in Ω



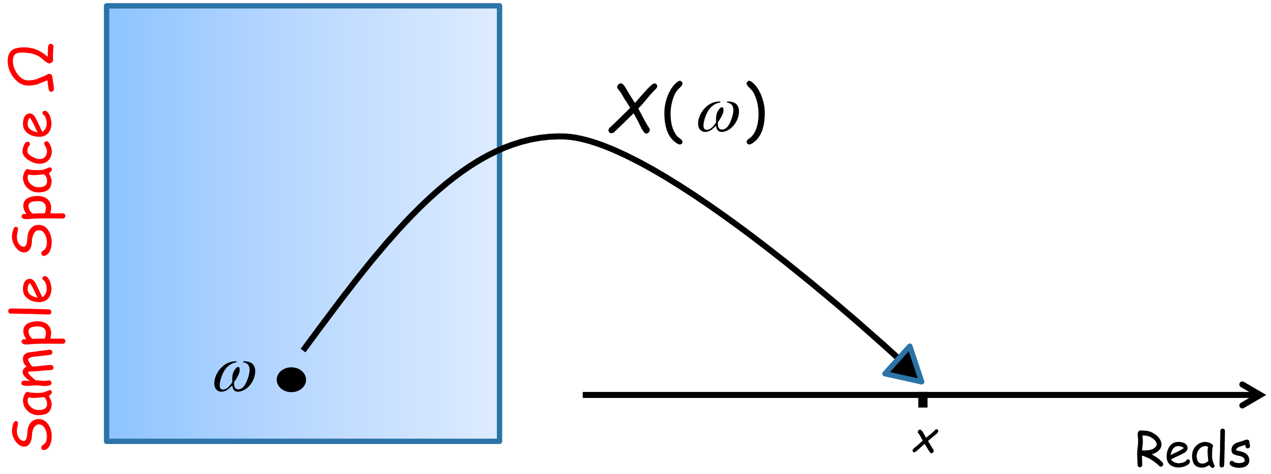
Hence $P(X=2) = P(\boldsymbol{A}) = \frac{1}{36}$

$$P(X=5) = P(\boldsymbol{B}) = \frac{4}{36} = \frac{1}{9}$$

Random variables: some formalism

Formally a RV is a function $X : \Omega \to W \subseteq \mathbb{R}$

i.e. the RV assigns to each elementary event ω a real number $X(\omega)$



Discrete random variable

- A discrete random variable X is a function that assigns to each elementary event of the sample space a discrete number.
- X takes its values according to a probability distribution $P(X = x)$

X	x_1	x_2	$\ldots$	x_k
$P(X = x)$	p_1	p_2	$\ldots$	p_k

- The set of values of X for which $P(X = x) > 0$ is called the **support** of X

Probability function

- The probability function (pf) is a rule that assigns to each value of X a probability i.e.

$$f(x) = P(X = x).$$

- This function must satisfy:

$$f(x) \geq 0, \text{ for any } x; \quad \sum_x f(x) = 1.$$

Example: discrete random variable - pf

A gambler takes a ball from an urn with 10 balls numbered from 1 to 10 and wins a given amount of money according to the following scheme (random experiment). Each ball has the same probability to be selected.

Outcomes	Win
1 - 4	50
5 - 7	100
8 - 9	200
10	400

Random variable

X	$f(x)$
50	0,4
100	0,3
200	0,2
400	0,1

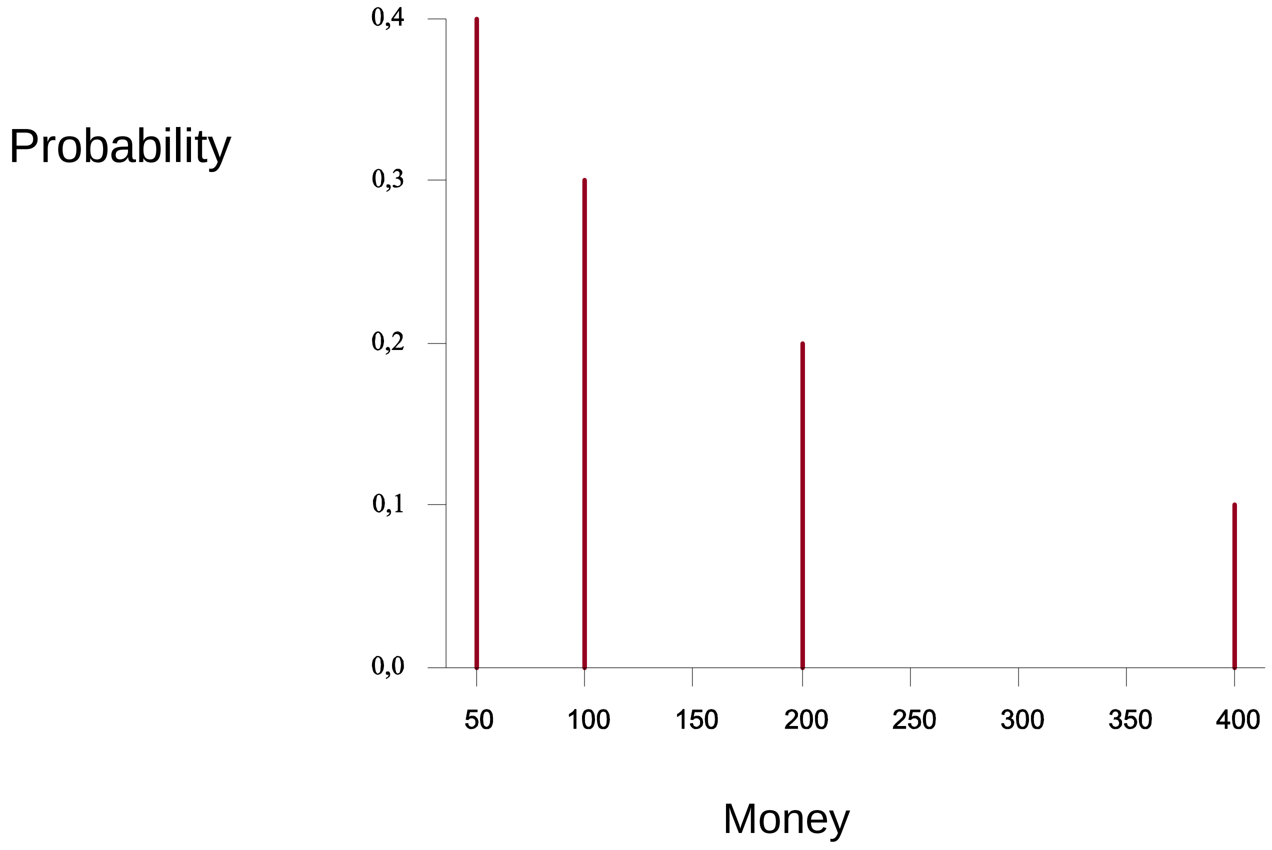
Note that

$$f(x) \geq 0, \; for\ all \;\; x,$$

$$\sum_{x} f(x) = f(50) + f(100) + f(200) + f(400) = 1\,.$$

Example: discrete random variable - pf

The probability function is graphically represented by a bar diagram



Continuous random variables

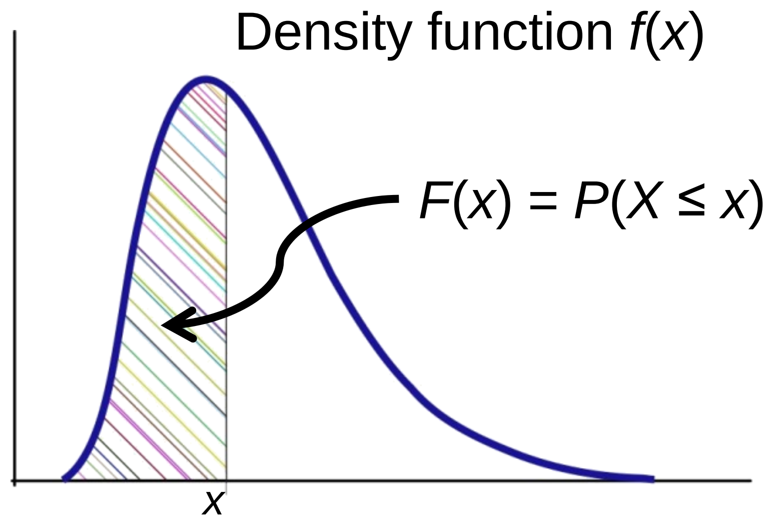
Continuous RV: the possible outcomes are an infinite continuum.

X is a continuous random variable if there exists a function, $f(x)$, called the **probability density function (pdf)** of X**,** such that for any value x, $\boldsymbol{F(x) = P(X \leq x)}$ is the area below the graph of $f(x)$ on the left of x i.e. we can probilize any interval of numbers.

$F(x)$ is known as the **cumulative density function** (**cdf**).

The cdf is another way to identify a random variable.

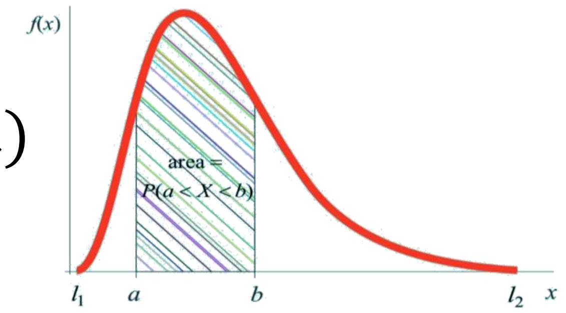
The cdf assigns to each value x of X the cumulative probability that X assumes a value smaller than or equal to x.



Continuous random variables

- The probability that X takes a value in a given real interval (*a*, *b*) is the area below the graph of *f* in the interval (*a*, *b*)

$$P(a < X \leq)$$
$$= F(b) - F(a)$$



- A density function must satisfy the following conditions

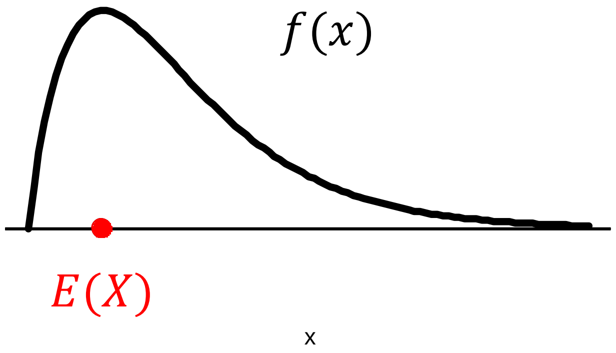
$$f(x) \geq 0, \forall x \; ; \qquad \int_{-\infty}^{+\infty} f(x)dx = 1$$

Expected value and variance of a RV

The expected value or mean of a random variable is a measure of centrality of X.

Technically:

- if X is discrete $\quad \mathrm{E}(X) = \sum_x x f(x)$
- if X is continuous $\mathrm{E}(X) = \int_{-\infty}^{+\infty} x f(x) dx$.



The variance of a random variable is a measure of variability of X around its mean.

Technically:

- if X is discrete $\quad \mathrm{Var}(X) = \sum_x (x-\mu)^2 f(x)$,
- if X is continuous $\mathrm{Var}(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx$.

$\sqrt{Var(X)}$ is the **standard deviation** of X

Example: discrete random variable: mean and variance

Given the probability distribution of the gambler win

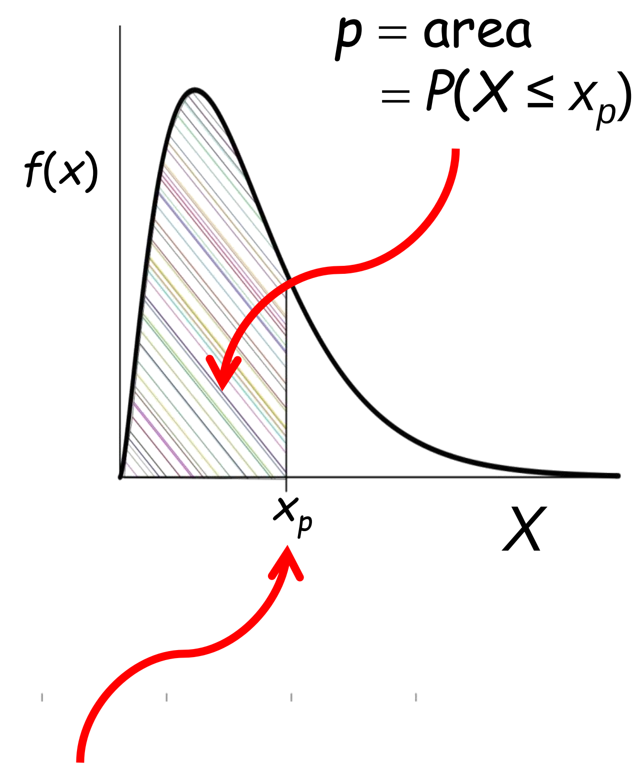
X	$f(x)$
50	0,4
100	0,3
200	0,2
400	0,1

- **mean:** $$\mathrm{E}(X) = \sum_x x f(x) = 50 \cdot 0{,}4 + 100 \cdot 0{,}3 + 200 \cdot 0{,}2 + 400 \cdot 0{,}1 = 130.$$
- **variance:** $$Var(X) = \sum_x (x-\mu)^2 f(x) = (50-130)^2 0{,}4 + (100-130)^2 0{,}3$$ $$+(200-130)^2 0{,}2 + (400-130)^2 0{,}1 = 11.100$$
- **standard deviation:** $$\sigma = \sqrt{11.100} = 105{,}36.$$

Quantiles of a RV

Given a certain probability p, the **quantile of order *p*** of X is the value x_p of the RV such that the probability of the variable being less than or equal to that value equals the given probability p .

More formally: x_p is s.t. $p = P(X \leq x_p) = F(x_p)$



quantile: value x_p in the support of X

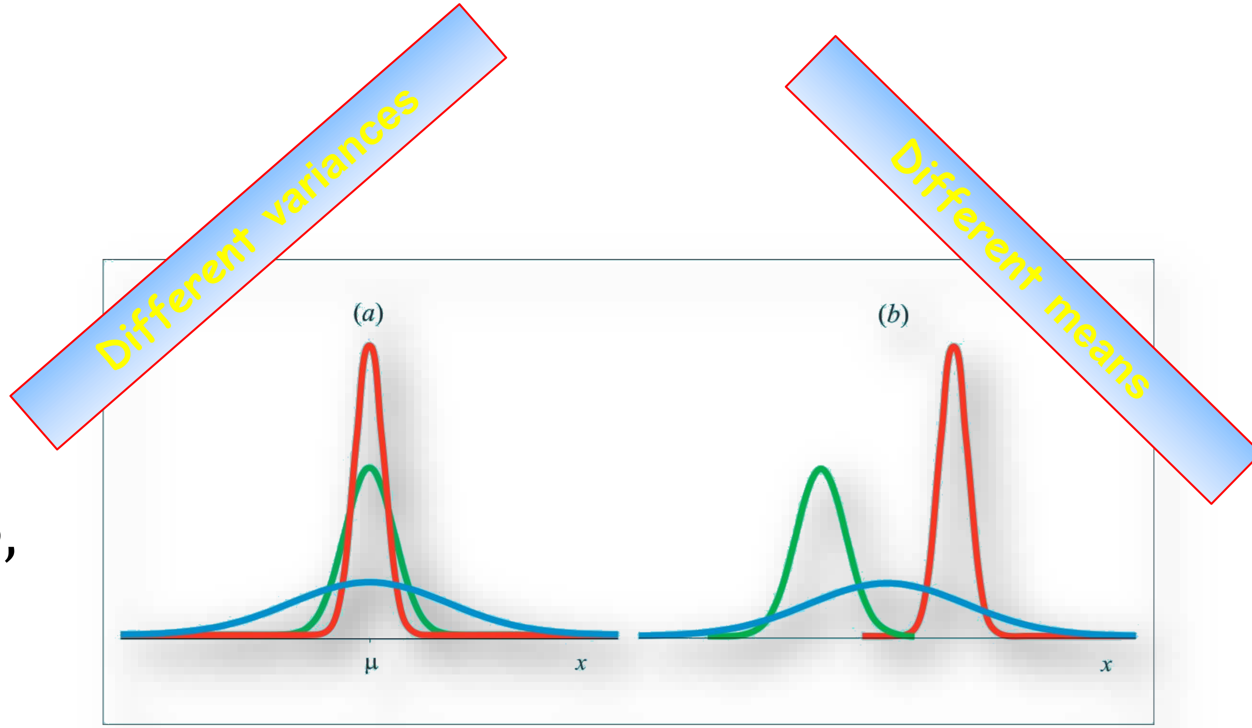
The median is the quantile of order 0.5

The Normal RV

The normal or Gaussian RV is defined by the pdf

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < +\infty,$$

$\pi = 3.14159...$



where $E(X) = \mu \in (-\infty, +\infty)$ and $\mathrm{Var}(X) = \sigma^2 > 0$

The pdf is symmetric about μ (which is also the median and the mode of X)

Standardised RVs

The **standardized RV** is obtained by the following transformation

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

where $\mathrm{E}(X) = \mu$ *and* $\mathrm{Var}(X) = \sigma^2$

A standardized RV always has a 0 mean and a unit variance.

Standardised Normal RV

If $X \sim N(\mu, \sigma^2)$ *(this reads " X is normally distributed with mean* μ *and variance* σ^2) then

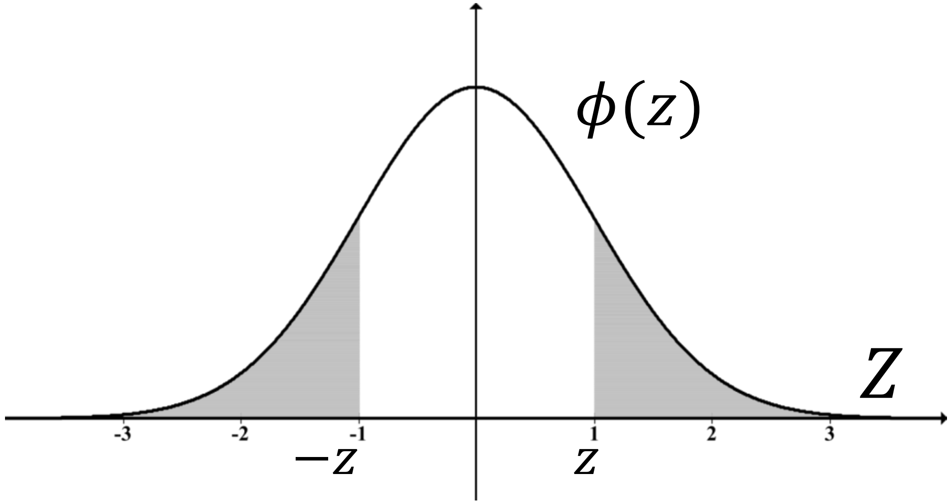
$$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

i.e. Z is the **standardised normal** with $E(Z) = 0$ and $Var(Z) = 1$.

From the symmetry of Z it follows:

$$\Phi(-z) = 1-\Phi(z),$$

where $\Phi(z) = P(Z \leq z)$ is the cdf of Z and $\phi(z)$ the pdf.



Normal RV and Standardised Normal RV

The following result holds true: $F(x) = P(X \leq x) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ for all $x \in \mathbb{R}$

Therefore, $z_p = \frac{x_p-\mu}{\sigma}$ and $x_p = \mu + \sigma z_p$ for any $p \in (0,1)$

Note: quantiles and cumulative probabilities of the standardized Normal are tabulated, hence this relationship allows one to calculate the probability associated to any real interval according to any Gaussian density.

Example: if X~N(8,15) calculate $P(6{,}5 < X \leq 8{,}3)$

$$P(6{,}5 < X < 8{,}3) = P(X \leq 8{,}4) - P(X \leq 6.5) = P\left(\frac{X-\mu}{\sigma} < \frac{8{,}3-8}{\sqrt{15}}\right) - P\left(\frac{X-\mu}{\sigma} < \frac{6{,}5-8}{\sqrt{15}}\right) = \Phi(0{,}08) - \Phi(-0{,}39)$$

From the Normal tables

$$= 0{,}5319 - (1 - 0{,}6517) = 0{,}1836.$$