Continuous random variables and limit laws

Statistical Computing and Empirical Methods
Unit EMATM0061, Data Science MSc

Rihuan Ke
rihuan.ke@bristol.ac.uk

Teaching Block 1, 2024

University of BRISTOL

What we will cover today

We will introduce the concept of a continuous random variable

We will see continuous random variables can be understood via the probability density function

We will discuss expectation, variance, standard deviation, covariance and correlation in this context.

We introduce the important example of Gaussian random variables

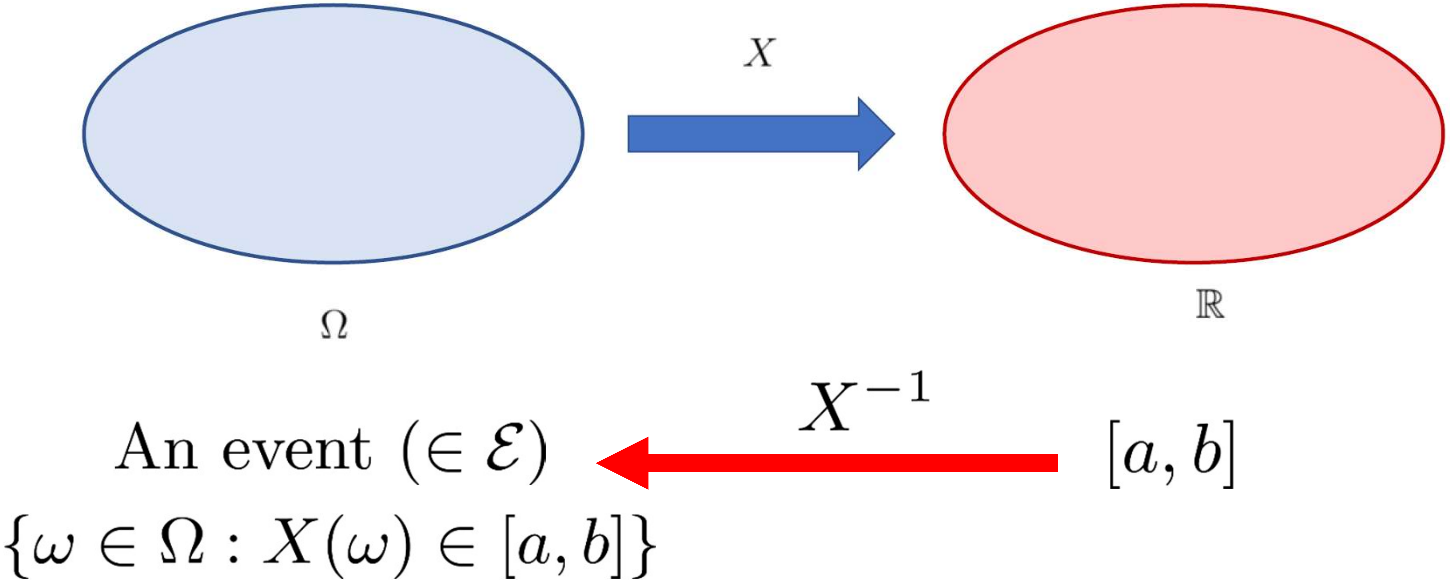
We will discuss the law of large numbers and the central limit theorem

Random variables

Random variables

Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$. A random variable is a mapping $X : \Omega \to \mathbb{R}$, such that

$$\text{for every } a, b \in \mathbb{R},\ \{\omega \in \Omega : X(\omega) \in [a, b]\} \text{ is an event in } \mathcal{E}$$



1. Continuous random variables

We often want to model stochastic quantities which can take on a continuum of possible values.

- The wing span of a penguin selected at random from a population of penguins on an island

- The time taken by an athlete to run the London marathon

- The level of rainfall in a given location on a particular day of the year.

Continuous random variables

Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$.

Recall that a random variable is a mapping $X : \Omega \to \mathbb{R}$ such that for every $a, b \in \mathbb{R}$, the set

$$\{\omega \in \Omega : X(\omega) \in [a, b]\} \text{ is an event in } \mathcal{E}.$$

A continuous random variable is a random variable X that can take on a continuum of values (rather than a discrete subset of $\mathbb{R}$).

Example: Uniform distribution.

A random variable X that takes any value in $[a, b]$ (for some $a < b$), and the values of X "spread evenly" across $[a, b]$.

Probability density function

Recall that a **discrete random variable** can be characterised by its **probability mass function** $x \to \mathbb{P}(X = x)$ for $x \in \mathbb{R}$.

However, for a continuous random variable, we have $\mathbb{P}(X = x) = 0$ for all $x \in \mathbb{R}$. So we can not map each x to a meaningful probability.

Instead of discussing the probability of X taking on any one single value, we can consider the "probability per unit length".

To understand continuous random variable $X : \Omega \to \mathbb{R}$, we require a probability density function f.

With the probability density function, we can compute the probability by integrating the values of f:

$$\mathbb{P}(X \in [c, d]) = \int_c^d f(x)dx.$$

Probability density function

Recall that a random variable is a mapping $X : \Omega \to \mathbb{R}$ such that for every $a, b \in \mathbb{R}$, the set

$$\{\omega \in \Omega : X(\omega) \in [a, b]\} \text{ is an event in } \mathcal{E}.$$

Probability density function and continuous random variable

A probability density function (p.d.f.) is a function $f : \mathbb{R} \to [0, \infty)$ such that

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

A continuous random variable is a random variable $X : \Omega \to \mathbb{R}$ with a probability density function $f_X : \mathbb{R} \to [0, \infty)$ such that for all $a, b \in \mathbb{R}$, we have

$$\mathbb{P}(X \in [a, b]) = \int_a^b f_X(x)dx.$$

If the above equality holds, then we say that $f_X : \mathbb{R} \to [0, \infty)$ is the p.d.f. of the random variable $X : \Omega \to \mathbb{R}$.

The associated distribution function is given by
$F_X(x) := \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(z)dz$ for all $x \in \mathbb{R}$.

Example

Random variables with uniform distributions.

For given $a, b \in \mathbb{R}$, define a function

$$f_X(x) := \frac{1}{b-a}\mathbb{1}_{[a,b]}(x) = \begin{cases} \frac{1}{b-a}, & \text{if } x \in [a,b] \\ 0. & \text{otherwise.} \end{cases}$$

f_X is a well-defined probability density function because $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

Let $X : \Omega \to \mathbb{R}$ be a random variable satisfying for any c, d:

$$\mathbb{P}(X \in [c,d]) = \int_c^d f_X(x)dx = \frac{1}{b-a} \cdot \text{length}([c,d] \cap [a,b]).$$

So X is a continuous random variable, and X is said to follow a uniform distribution on $[a, b]$.

Expectation for continuous random variables

Recall that a continuous random variable is a random variable $X : \Omega \to \mathbb{R}$ with a probability density function $f_X : \mathbb{R} \to [0, \infty)$ such that for all $a, b \in \mathbb{R}$, we have $\mathbb{P}(X \in [a, b]) = \int_a^b f_X(x)dx$.

Expectation

The expectation of a continuous random variable X is defined by

$$\mathbb{E}(X) := \int_{-\infty}^{\infty} x f_X(x)dx.$$

Example: Let X be a random variable with uniform distribution on $[a, b]$, then

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)dx = \int_a^b x \frac{1}{b-a} dx = \frac{b+a}{2}.$$

Expectation is linear

Expectation is linear in the following sense:

Given random variables $X_1, X_2, \cdots, X_K : \Omega \to \mathbb{R}$ and numbers $\alpha_1, \alpha_2, \cdots, \alpha_K \in \mathbb{R}$, we have

$$\mathbb{E}\left(\sum_{i=1}^{K} \alpha_i X_i\right) = \sum_{i=1}^{K} \alpha_i \mathbb{E}(X_i).$$

Variance and standard deviation

Recall that a continuous random variable is a random variable $X : \Omega \to \mathbb{R}$ with a probability density function $f_X : \mathbb{R} \to [0, \infty)$ such that for all $a, b \in \mathbb{R}$, we have $\mathbb{P}(X \in [a, b]) = \int_a^b f_X(x)dx$.

Recall that the expectation of a continuous random variable is defined by $\mathbb{E}(X) := \int_{-\infty}^{\infty} x f_X(x)dx$.

Variance and standard deviation

The variance of a continuous random variable is given by

$$\mathrm{Var}(X) := \mathbb{E}\big[(X - \mathbb{E}(X))^2\big] = \int_{-\infty}^{\infty} (X - \mathbb{E}(X))^2 f_X(x)dx.$$

In addition, the standard deviation of a random variable is defined by $\sigma(X) := \sqrt{\mathrm{Var}(X)}$.

The variance can also be computed by:

$$\mathrm{Var}(X) := \mathbb{E}\big[(X - \mathbb{E}(X))^2\big] = \mathbb{E}\big[X^2 - 2X \cdot \mathbb{E}(X) + (\mathbb{E}(X))^2\big] = \mathbb{E}\big(X^2\big) - [\mathbb{E}(X)]^2$$

Example

Recall that the variance of a continuous random variable is given by $\mathrm{Var}(X) := \mathbb{E}\big[(X - \mathbb{E}(X))^2\big] = \int_{-\infty}^{\infty} (X - \mathbb{E}(X))^2 f_X(x)dx.$

Recall that the probability density function of a uniform random variable X is given by

$$f_X(x) := \frac{1}{b-a}\mathbb{1}_{[a,b]}(x) = \begin{cases} \frac{1}{b-a}, & \text{if } x \in [a,b] \\ 0. & \text{otherwise.} \end{cases}$$

Example: Let X be a random variable with uniform distribution on $[a, b]$, then

$$\mathrm{Var}(X) = \mathbb{E}\big(X^2\big) - [\mathbb{E}(X)]^2 = \int_a^b \frac{x^2}{b-a}dx - \left(\frac{b+a}{2}\right)^2 = \frac{1}{12}(b-a)^2.$$

Covariance and correlation

Recall that the variance of a continuous random variable is given by
$\mathrm{Var}(X) := \mathbb{E}\big[(X - \mathbb{E}(X))^2\big] = \int_{-\infty}^{\infty} (X - \mathbb{E}(X))^2 f_X(x)dx.$

Covariance and correlation

Given a pair of random variables $X, Y : \Omega \to \mathbb{R}$, we define the covariance and correlation respectively by

$$\mathrm{Cov}(X,Y) := \mathbb{E}[(X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))] = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y),$$

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

$\mathrm{Cov}(X,X) = \mathrm{Var}(X).$

$\mathrm{Corr}(X,X) = 1.$

If X, Y are indpendent, then $\mathrm{Cov}(X,Y) = \mathrm{Corr}(X,Y) = 0.$

Variance of linear combination of random variables

Recall that the variance of a continuous random variable is given by $\mathrm{Var}(X) := \mathbb{E}\left[(X - \mathbb{E}(X))^2\right] = \int_{-\infty}^{\infty} (X - \mathbb{E}(X))^2 f_X(x)dx$.

Recall that given a pair of random variables $X, Y : \Omega \to \mathbb{R}$, we define the covariance and correlation respectively by
$\mathrm{Cov}(X,Y) := \mathbb{E}[(X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))] = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$, and
$\mathrm{Corr}(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$.

Lemma

Given random variables $X_1, \cdots, X_K : \Omega \to \mathbb{R}$ and $\alpha_1, \cdots, \alpha_K \in \mathbb{R}$, we have

$$\mathrm{Var}\left(\sum_{i=1}^{K} \alpha_i X_i\right) = \sum_{i=1}^{K} \alpha_i^2 \mathrm{Var}(X_i) + 2 \sum_{1 \le i < j \le k} \alpha_i \cdot \alpha_j \mathrm{Cov}(X_i, X_j).$$

In particular, if $X_1, \cdots, X_k$ are independent, then
$\mathrm{Var}\left(\sum_{i=1}^{K} \alpha_i X_i\right) = \sum_{i=1}^{K} \alpha_i^2 \mathrm{Var}(X_i)$.

2. Gaussian random variables

Gaussian random variables

A Gaussian random variable is a continuous random variable X with the probability density function $f_{\mu,\sigma} : \mathbb{R} \to [0, \infty)$ defined by

$$f_{\mu,\sigma}(x) := \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right) \text{ for all } x \in \mathbb{R}.$$

Here (μ, σ) are parameters.

A Gaussian random variable is often referred to as a normal random variable.

The distribution of a Gaussian random variable is often referred to as a Gaussian distribution or a normal distribution.

We often write $X \sim \mathcal{N}(\mu, \sigma^2)$ to mean X is Gaussian with parameters μ, σ.

If $\mu = 0$ and $\sigma = 1$: the standard Gaussian random variable and the standard Gaussian distribution.

Gaussian random variables

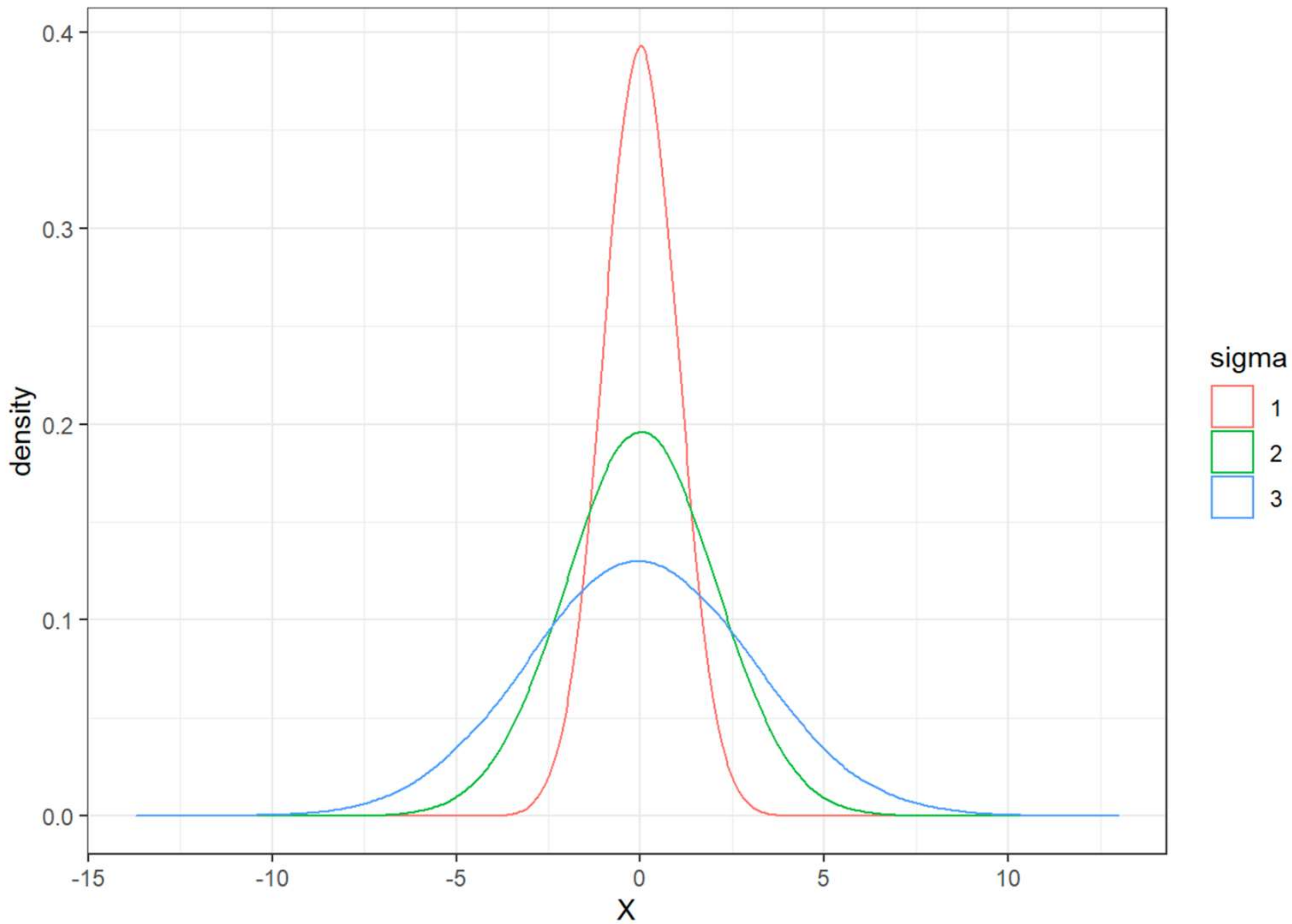
Recall that $f_{\mu,\sigma}(x) := \frac{1}{\sigma\sqrt{2\pi}}\exp\Big(-\frac{1}{2}\big(\frac{x-\mu}{\sigma}\big)^2\Big)$.

To make sure that $f_{\mu,\sigma}$ is a well-defined denisty we must have $\int_{-\infty}^{\infty} f_{\mu,\sigma}(x)dx = 1$.

To prove this, let's start with the Gaussian integral $\int_{-\infty}^{\infty} e^{-z^2}dz = \sqrt{\pi}$ and then apply a change of variables $z = \frac{x-\mu}{\sigma\sqrt{2}}$ with derivative $\frac{dz}{dx} = \frac{1}{\sigma\sqrt{2}}$,

$$\begin{aligned}
\int_{-\infty}^{\infty} f_{\mu,\sigma}(x)dx &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)dx \\
&= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp(-z^2)\left(\frac{dz}{dx}\right)^{-1} dz \\
&= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-z^2}\cdot\left(\sigma\sqrt{2}\right)dz = 1.
\end{aligned}$$

Gaussian random variables



$$X \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

Expectation of Gaussian random variables

Recall that the expectation of a continuous random variable is defined by $\mathbb{E}(X) := \int_{-\infty}^{\infty} x f_X(x)dx$.

Recall that $f_{\mu,\sigma}(x) := \frac{1}{\sigma\sqrt{2\pi}}\exp\Big(-\frac{1}{2}\big(\frac{x-\mu}{\sigma}\big)^2\Big)$.

Recall that $\int_{-\infty}^{\infty} f_{\mu,\sigma}(x)dx = 1$.

$$\begin{aligned}\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_{\mu,\sigma}(x)dx &= \int_{-\infty}^{\infty} \mu f_{\mu,\sigma}(x) + (x-\mu)f_{\mu,\sigma}(x)dx \\ &= \mu + \int_{-\infty}^{\infty} (x-\mu)f_{\mu,\sigma}(x)dx\end{aligned}$$

$$\begin{aligned}\int_{-\infty}^{\infty} (x-\mu)f_{\mu,\sigma}(x)dx &= \int_{-\infty}^{\infty} (x-\mu)\cdot\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx \\ &= \left[\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\cdot(-\sigma)\right]_{-\infty}^{\infty} = 0-0=0\end{aligned}$$

So the **expectation** of the Gaussian random variable is $\mathbb{E}(X) = \mu$.

The variance of Gaussian random variables

Recall that $f_{\mu,\sigma}(x) := \frac{1}{\sigma\sqrt{2\pi}}\exp\Big(-\frac{1}{2}\big(\frac{x-\mu}{\sigma}\big)^2\Big)$.

Recall that the **expectation** of the Gaussian random variable is $\mathbb{E}(X) = \mu$.

Recall that the variance of a continuous random variable is given by
$\mathrm{Var}(X) := \mathbb{E}\big[(X - \mathbb{E}(X))^2\big] = \int_{-\infty}^{\infty} (X - \mathbb{E}(X))^2 f_X(x)dx$.

We apply a change of variables $z = \frac{x-\mu}{\sigma\sqrt{2}}$ with derivative $\frac{dz}{dx} = \frac{1}{\sigma\sqrt{2}}$:

$$\begin{aligned}
\mathrm{Var}(X) = \mathbb{E}\big((X-\mu)^2\big) &= \int_{-\infty}^{\infty} (x-\mu)^2 f_{\mu,\sigma} dx \\
&= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\
&= \int_{-\infty}^{\infty} 2\sigma^2 \cdot z^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-z^2} \cdot \frac{dx}{dz} dz \\
&= \int_{-\infty}^{\infty} 2\sigma^2 \cdot z^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-z^2} \cdot (\sigma\sqrt{2}) dz = \sigma^2 \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} 2z^2 e^{-z^2} dz.
\end{aligned}$$

The variance of Gaussian random variables

Let $u := e^{-z^2}$ and $v := -z$, then $\frac{du}{dz} = -2ze^{-z^2}$ and $\frac{dv}{dz} = -1$.

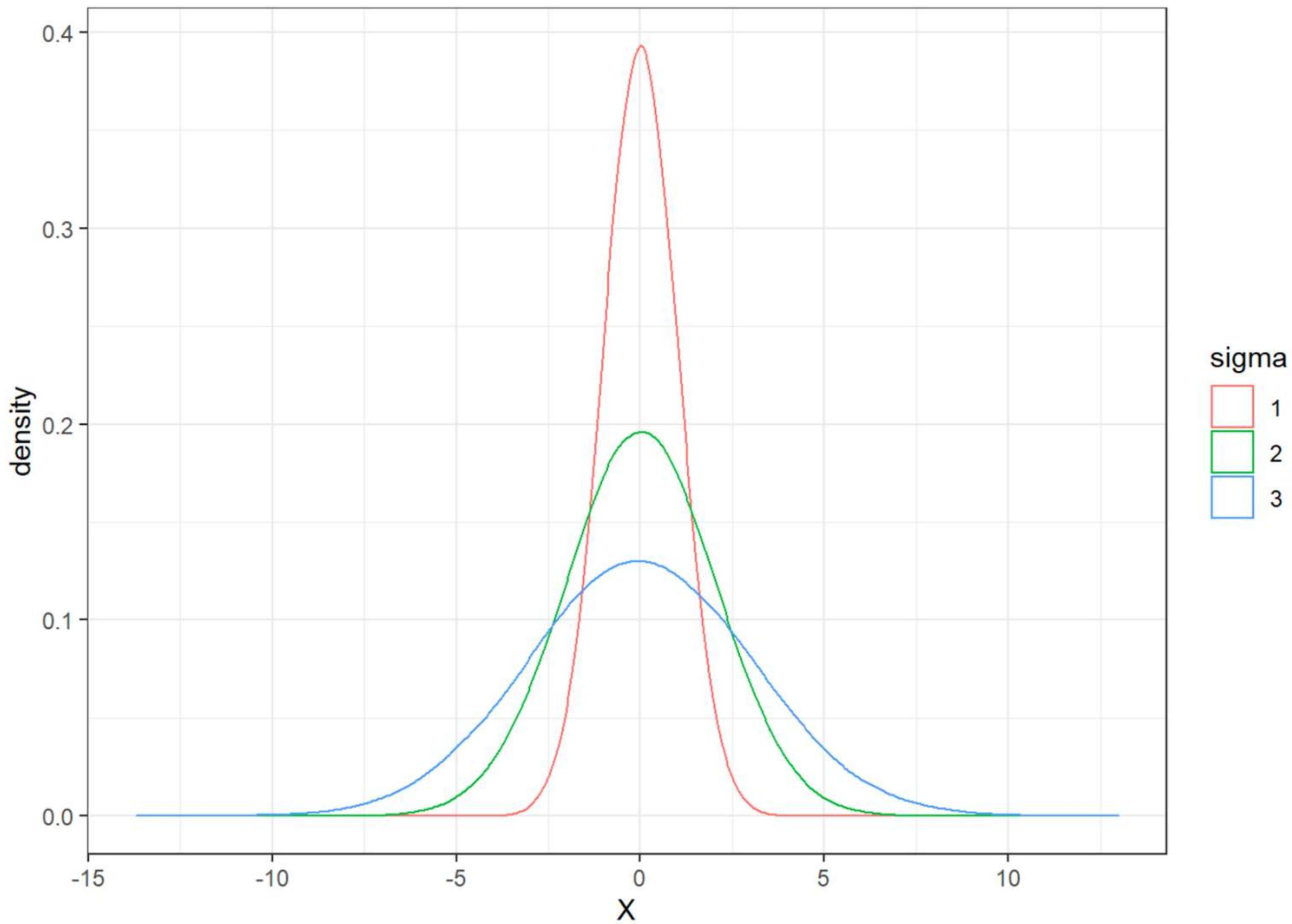
By integration by parts,

$$\begin{aligned}\int_{-\infty}^{\infty} 2z^2 e^{-z^2} dz = \int_{-\infty}^{\infty} v \frac{du}{dz} dz = [uv]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} u \frac{dv}{dz} dz \\ = [-z \cdot e^{-z^2}]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-z^2} dz \\ = 0 + \sqrt{\pi} = \sqrt{\pi}.\end{aligned}$$

Therefore, the variance of a Gaussian random variable is

$$\mathrm{Var}(X) = \sigma^2 \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} 2z^2 e^{-z^2} dz = \sigma^2.$$

Gaussian random variables

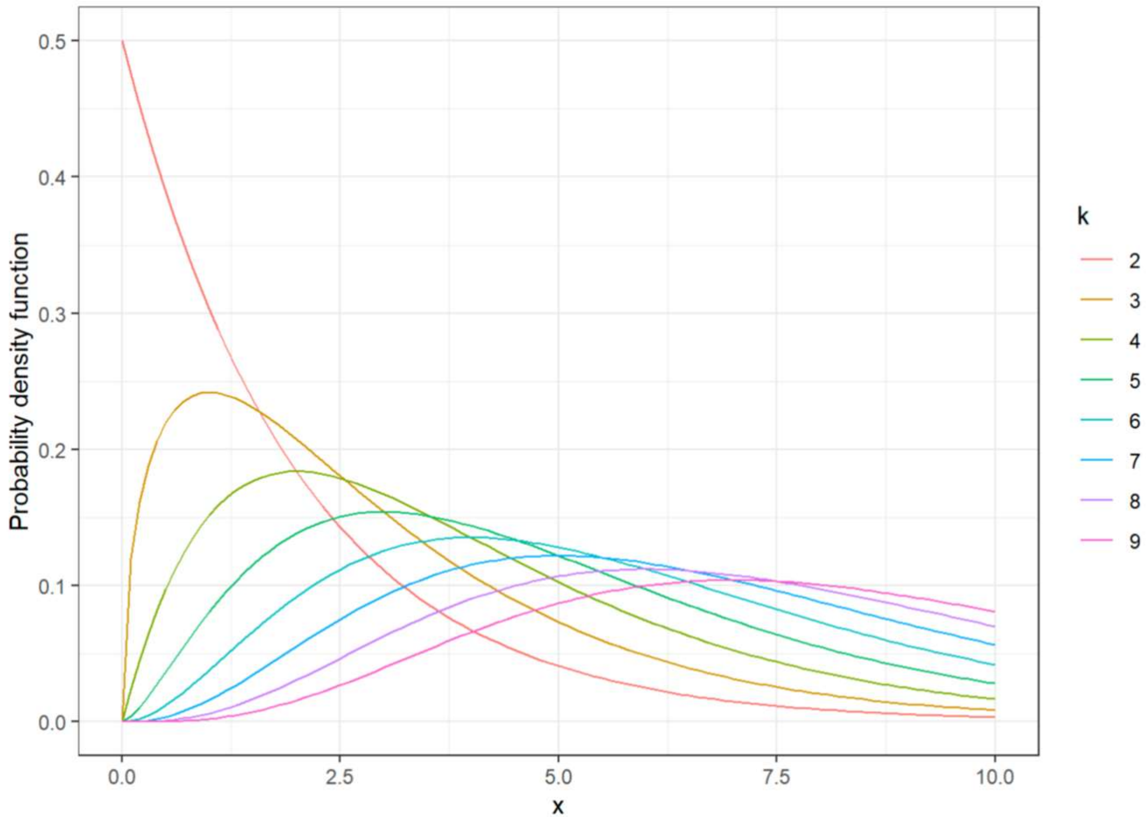


$$X \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

Chi-squared distribution

A random variable Q is said to be chi-squared with k degrees of freedom if $Q = \sum_{i=1}^{k} Z_i^2$ with independent $Z_1, Z_2, \cdots, Z_k \sim \mathcal{N}(0,1)$.

We write $Q \sim \chi^2(k)$.

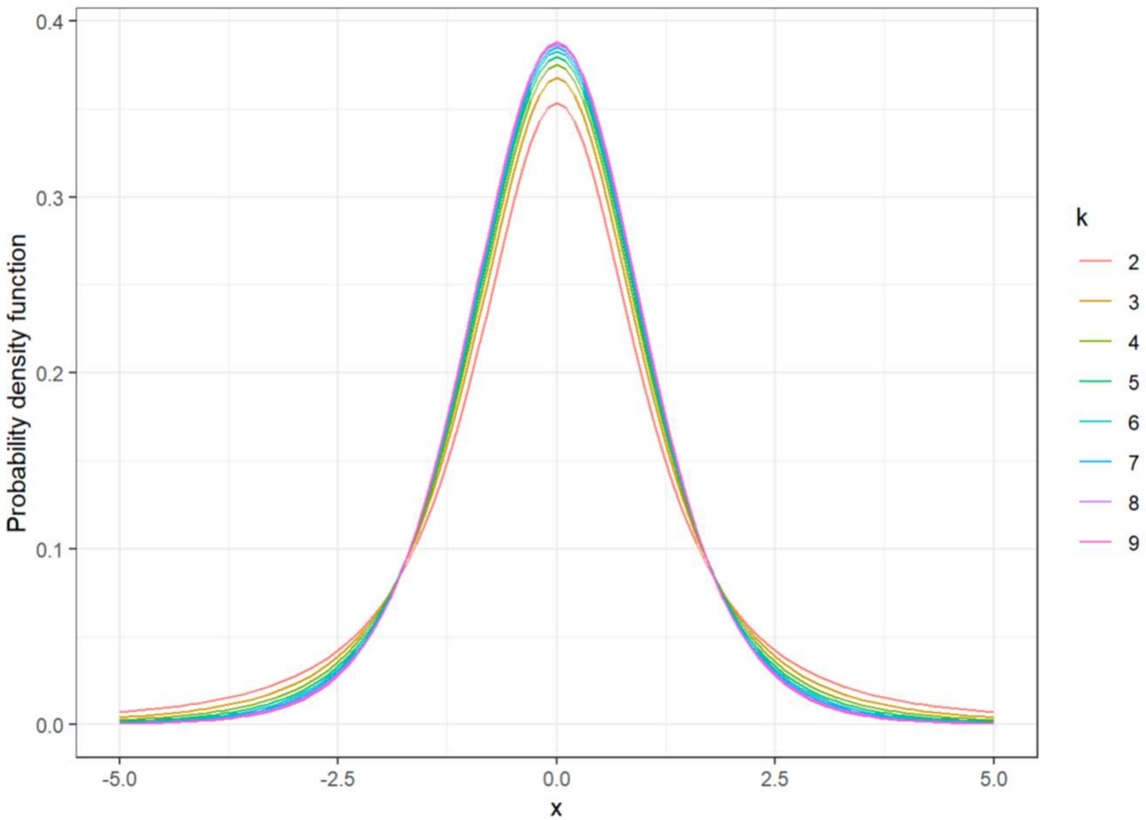


Expectation $\mathbb{E}(Q) = \sum_{i=1}^{k} \mathbb{E}(Z_i^2) = k.$

Student's t distribution

A random variable T is said to be t distributed with k degrees of freedom if $T = \frac{Z}{\sqrt{Q/k}}$ for two independent random variables $Z \sim \mathcal{N}(0,1)$ and $Q \sim \chi^2(k)$.

The distribution of T is called a Student's t distribution.



The cumulative distribution function of a continuous random variable

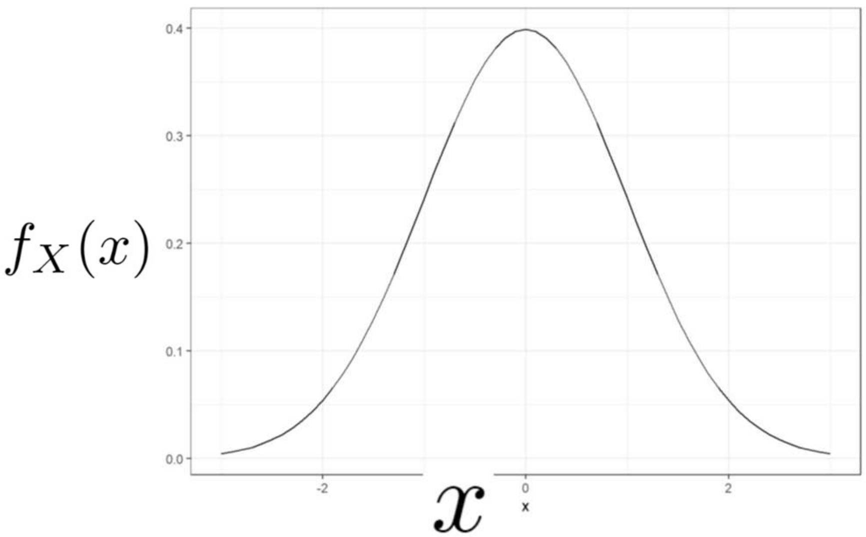
Recall that a continuous random variable is a random variable $X : \Omega \to \mathbb{R}$ with a probability density function $f_X : \mathbb{R} \to [0, \infty)$ such that for all $a, b \in \mathbb{R}$, we have $\mathbb{P}(X \in [a, b]) = \int_a^b f_X(x)dx$.

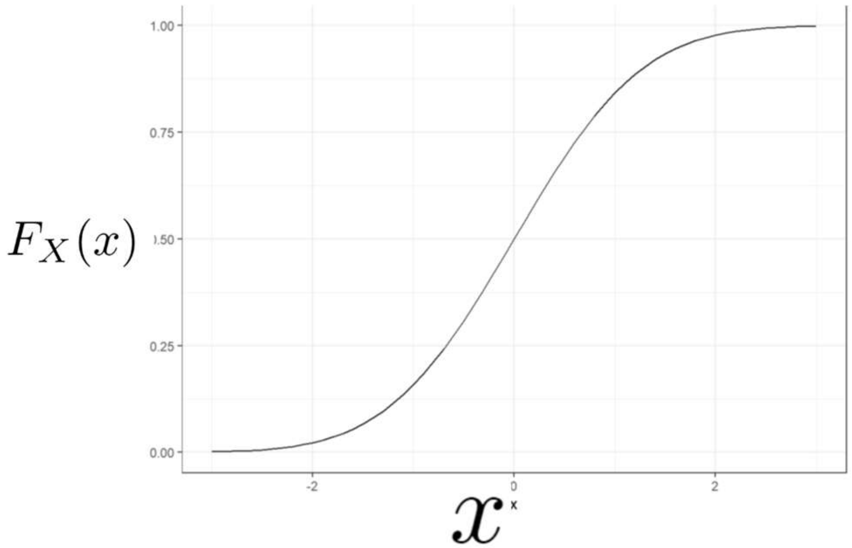
Cumulative distribution function

The cumulative distribution function of a continuous random variable is given by

$$F_X(x) := \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(z)dz.$$

For all $z \in \mathbb{R}$, we have $f_X(z) = \frac{dF_X(x)}{dx}\,|_z$





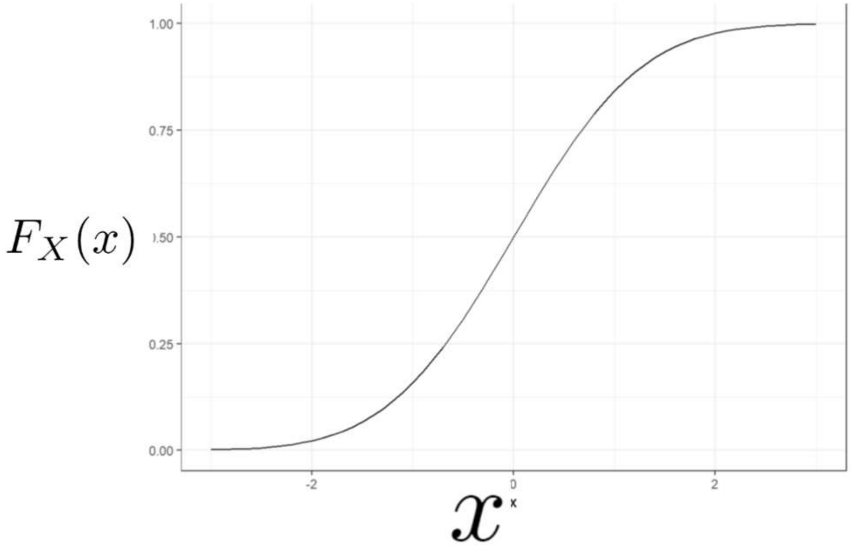
Quantile function

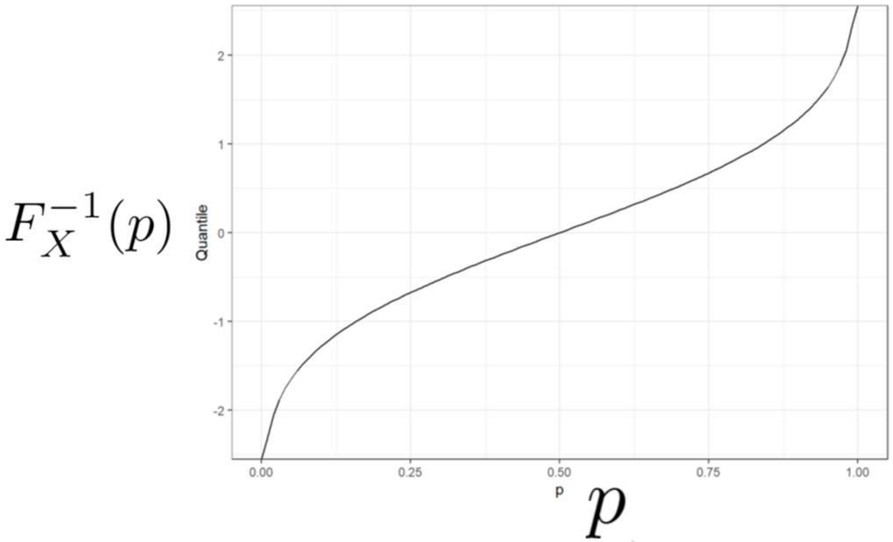
Recall that the cumulative distribution function of a continuous random variable is given by $F_X(x) := \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(z)dz$.

Quantile function

The quantile function $F_X^{-1} : [0,1] \to \mathbb{R}$ is defined by

$$F_X^{-1}(p) := \inf\{x \in \mathbb{R} : F_X(x) = \mathbb{P}(X \leq x) \geq p\} \text{ for } p \in [0,1].$$





Quantile function

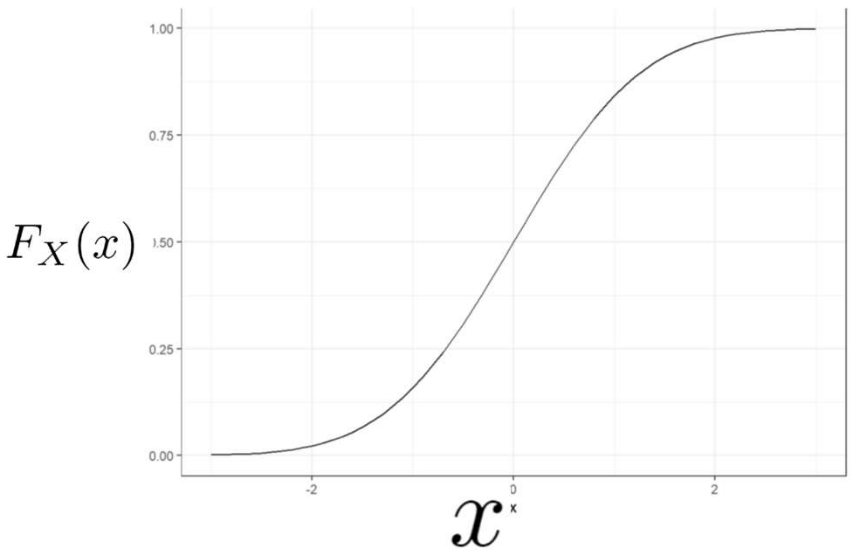
Quantile function

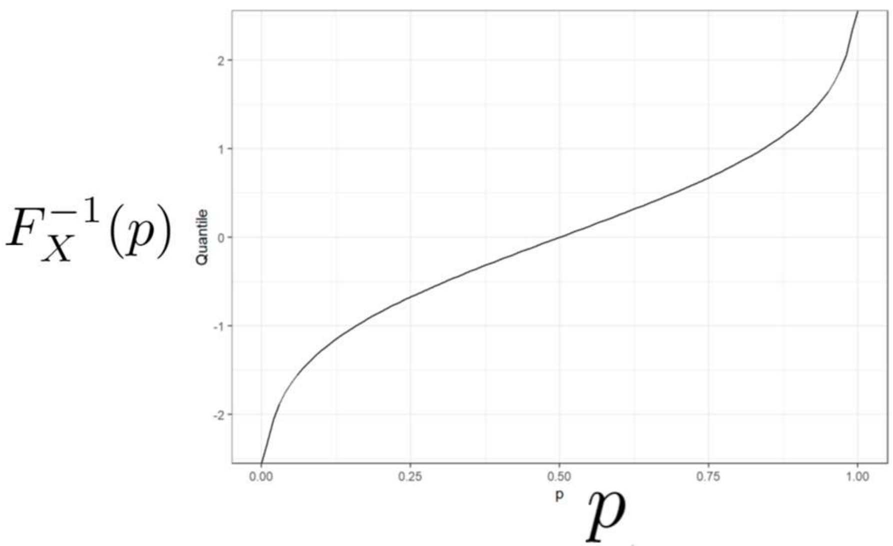
The quantile function $F_X^{-1} : [0,1] \to \mathbb{R}$ is defined by

$$F_X^{-1}(p) := \inf\{x \in \mathbb{R} : F_X(x) = \mathbb{P}(X \leq x) \geq p\} \text{ for } p \in [0,1].$$

We refer to the values $F_X^{-1}(p)$ for given p as population quantiles.

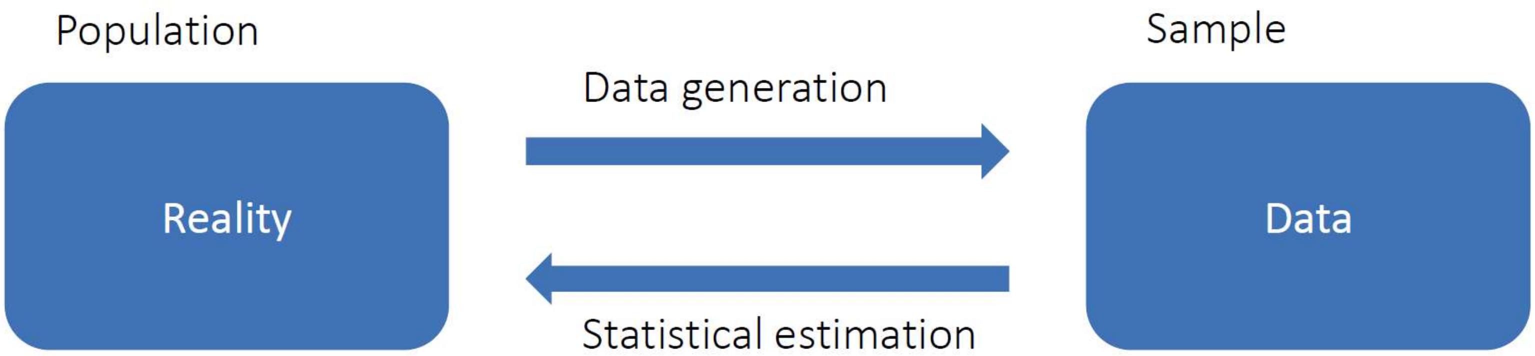
In particular, $F_X^{-1}(0.5)$ is the population median of the distribution of X.





The law of large numbers and the central limit theorem

Statistical estimation and probability



A distribution P_X for a random variable X

A sample of independent copies $X_1, \cdots, X_n \sim P_X$.

We model reality via the distribution P_X of a random variable $X : \Omega \to \mathbb{R}$.

We model our sample as a sequence of random variables $X_1, \cdots, X_n$ (independent copies of X).

Independent and identically distributed random variables

Recall that the distribution is $\mathbb{P}_X(S) = \mathbb{P}(X \in S)$ for a reasonable subset S of $\mathbb{R}$. Also the cumulative distribution function is $F_X(x) = \mathbb{P}(X \le x)$ for $x \in \mathbb{R}$.

i. i. d. random variables

We say that $X_1, \cdots, X_n$ are independent and identically distributed (i.i.d.) if

1. The sequence $X_1, \cdots, X_n$ consists of **mutually independent** random variables;
2. For all $x \in \mathbb{R}$, we have $F_{X_1}(x) = F_{X_2}(x) = \cdots = F_{X_n}(x)$.

Note that the condition 2. is equivalent to $P_{X_1} = P_{X_2} = \cdots = P_{X_n}$.

In particular, we refer to $X_1, \cdots, X_n$ as independent copies of X if $P_X = P_{X_1} = P_{X_2} = \cdots = P_{X_n}$.

The law of large numbers

Sequences of i.i.d. random variables are surprisingly well-behaved.

Given the roll of a single fair dice the outcome X is highly unpredictable.

However, suppose we take a very large number of independent dolls of a fair dice rolls $X_1, \cdots, X_n$, and then compute the sample average of the sequence $\frac{1}{n}\sum_{i=1}^n X_i$.

Then we expect the sample average $\frac{1}{n}\sum_{i=1}^n X_i$ to be close to $\mathbb{E}(X) = \frac{7}{2}$ for large n.

This is an example of a law of large numbers.

Law of large numbers: the sample mean approaches the population mean in some sense when the sample size is large.

Sample mean of dice rolls

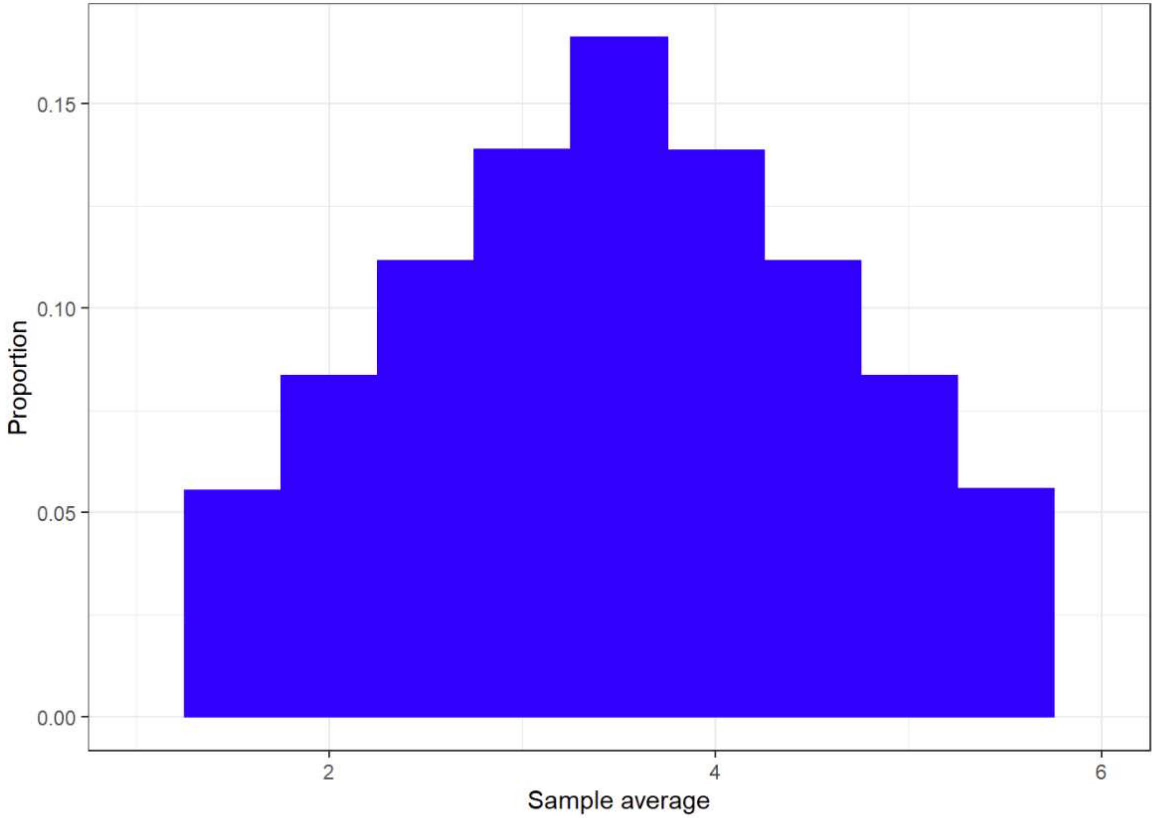
The sample average of the sequence $\frac{1}{n}\sum_{i=1}^{n} X_i$ of independent dice rolls.

```
num_trials<-1000000 # set the number of trials
set.seed(0) # set the random seed
sample_size<-2 # set the sample size

# simulate the sample average of a dice roll:
# add columns dice_sample (1...6) and sample_avg (sample average)
dice_sample_average_simulation <- data.frame(trial=1:num_trials) %>%
  mutate(dice_sample = map(.x=trial, ~sample(6,sample_size,replace=TRUE))) %>%
  mutate(sample_avg = map_dbl(.x=dice_sample, ~mean(.x)))

# plot a histogram to display the results
dice_sample_average_simulation %>%
  ggplot (aes(x=sample_avg) ) +
  geom_histogram(aes(y=..count../sum(..count..)),
                 binwidth=1/sample_size,fill="blue",color="blue") +
  theme_bw() + xlim(c(1,6)) +
  xlab("Sample average") + ylab("Proportion")
```

Sample mean of dice rolls

The sample average of the sequence $\frac{1}{n}\sum_{i=1}^{n} X_i$ of independent dice rolls.

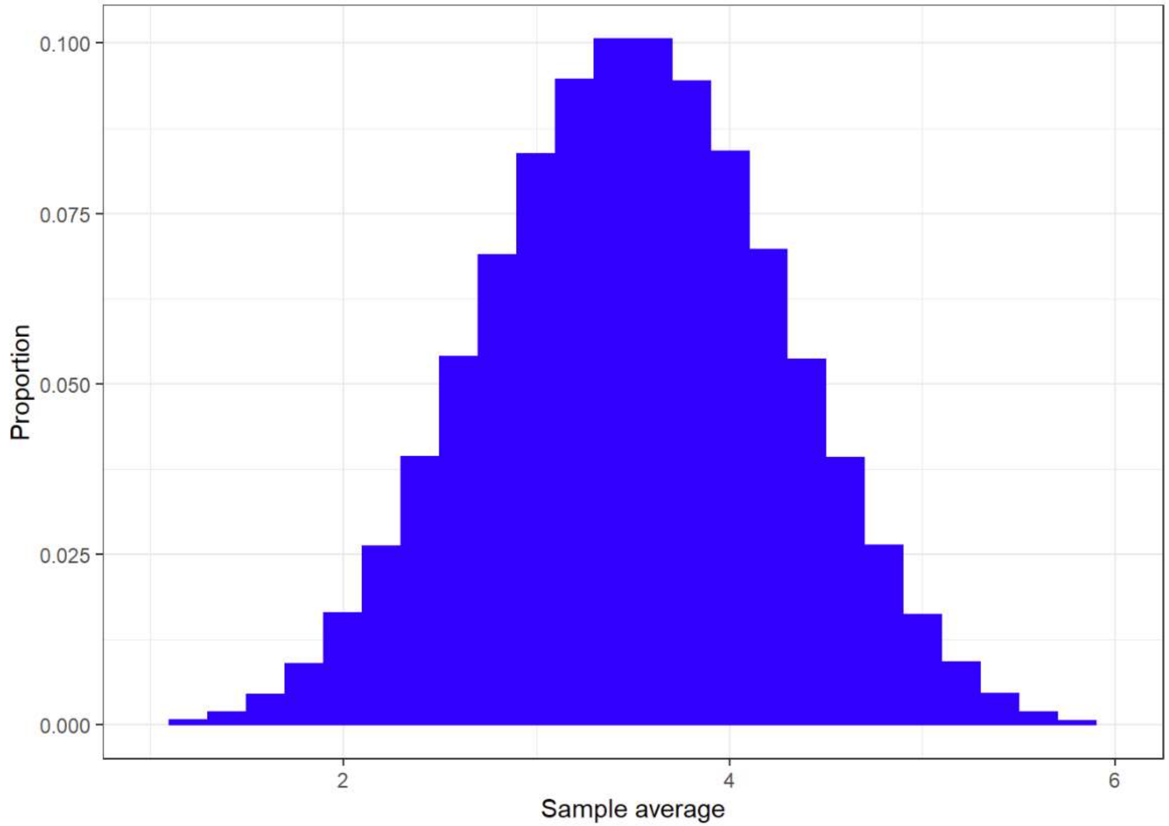
Sample size $n = 2$



Sample mean of dice rolls

The sample average of the sequence $\frac{1}{n}\sum_{i=1}^{n} X_i$ of independent dice rolls.

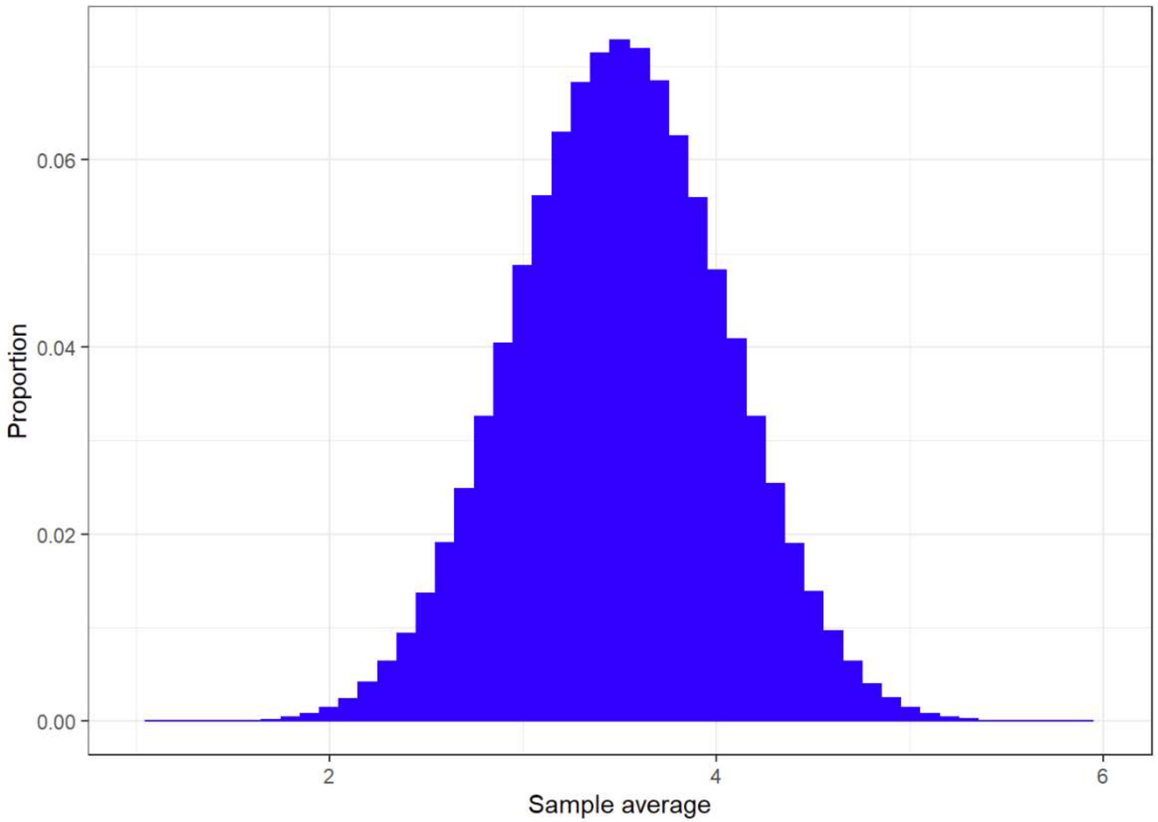
Sample size $n = 5$



Sample mean of dice rolls

The sample average of the sequence $\frac{1}{n}\sum_{i=1}^{n} X_i$ of independent dice rolls.

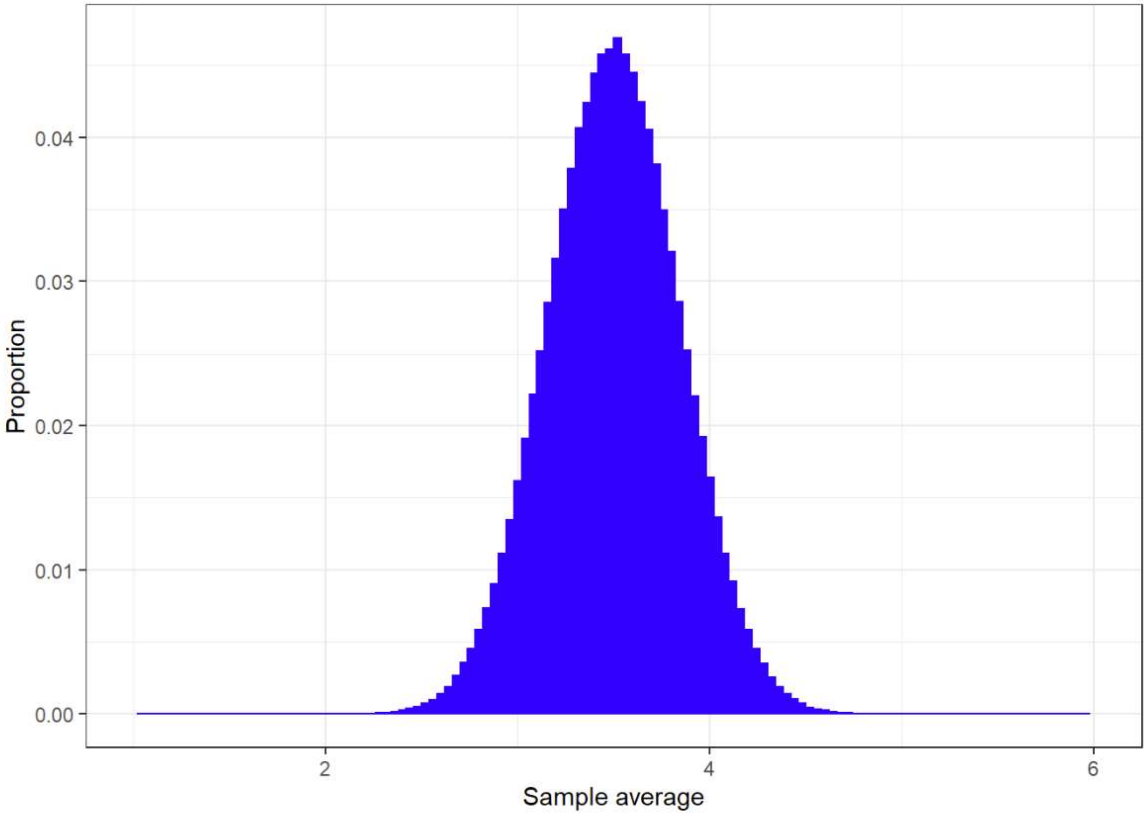
Sample size $n = 10$



Sample mean of dice rolls

The sample average of the sequence $\frac{1}{n}\sum_{i=1}^{n} X_i$ of independent dice rolls.

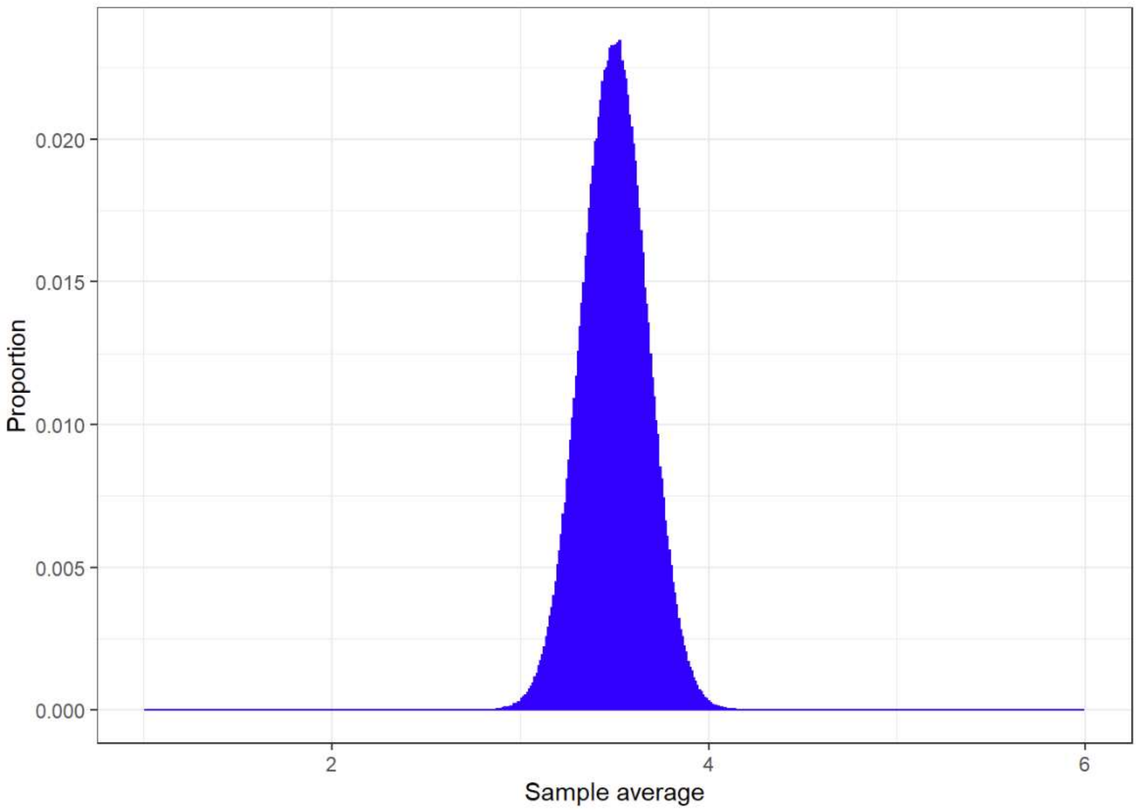
Sample size $n = 25$



Sample mean of dice rolls

The sample average of the sequence $\frac{1}{n}\sum_{i=1}^{n} X_i$ of independent dice rolls.

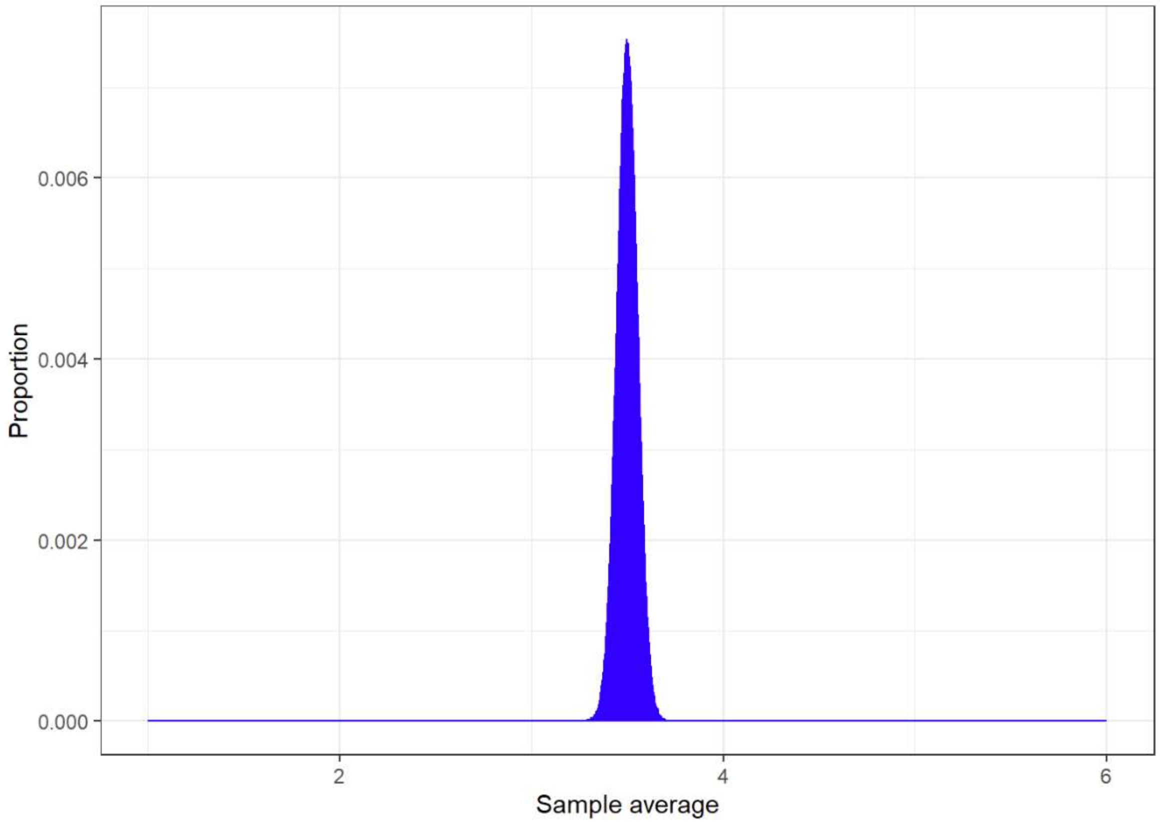
Sample size $n = 100$



Sample mean of dice rolls

The sample average of the sequence $\frac{1}{n}\sum_{i=1}^{n} X_i$ of independent dice rolls.

Sample size $n = 1000$



The law of large numbers

The law of large numbers tells us that the sample average converges towards the expectation, for sequences of independent and identically distributed random variables.

Theorem (Bernoulli, circa. 1700, the weak laws of large numbers)

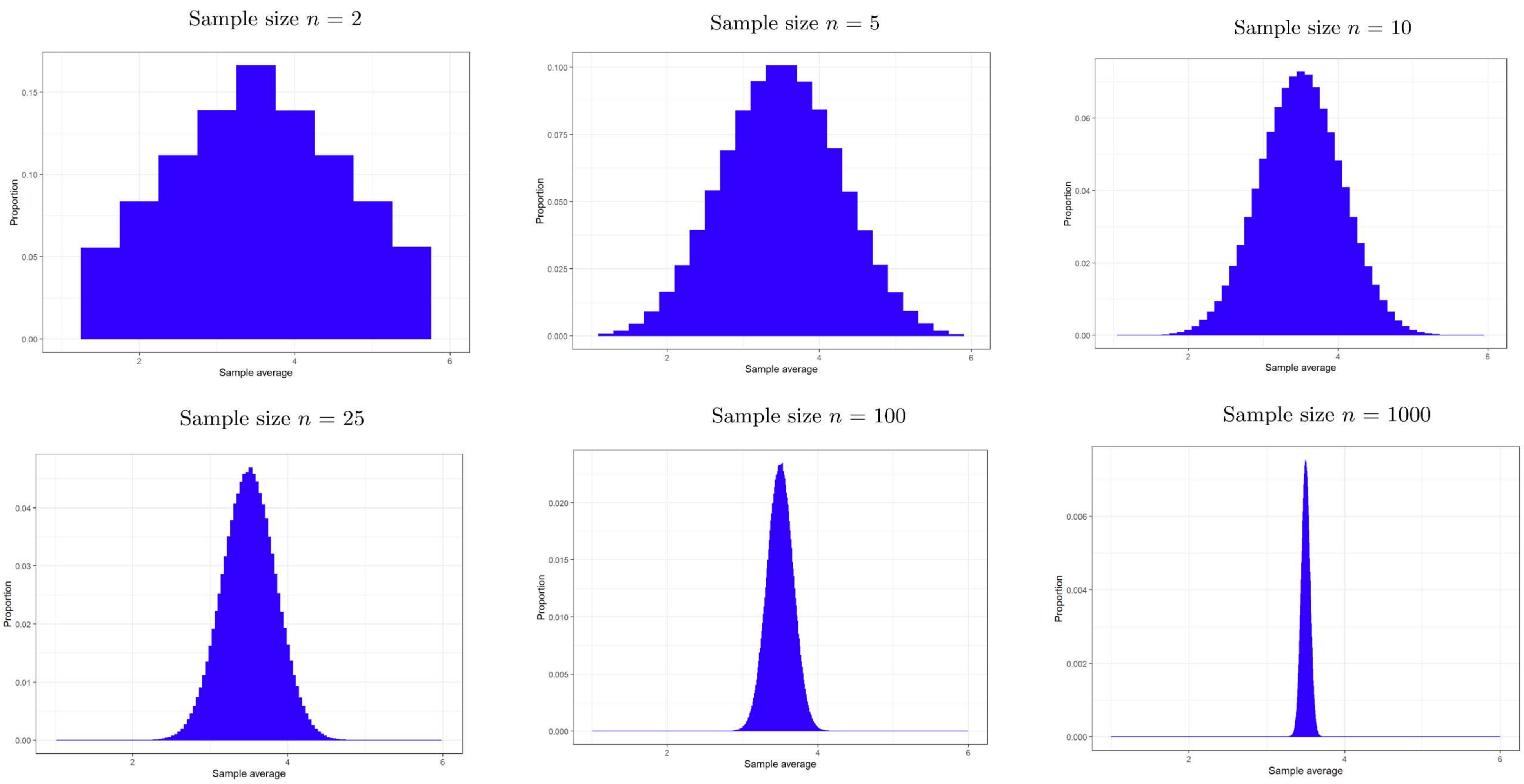
Let $X : \Omega \to \mathbb{R}$ be a random variable with a well-defined expectation $\mu = \mathbb{E}(X)$. Let $X_1, \cdots, X_n : \Omega \to \mathbb{R}$ be a sequence of independent copies of X. Then for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum X_i - \mu\right| \geq \epsilon\right) = 0.$$

The probability of the event that the distance between the sample mean $\frac{1}{n}\sum_{i=1}^{n} X_i$ and the population mean bigger than a small number ϵ can be arbitrarily small (if n is large enough). And we say that the sample mean converges to the population mean in probability.

Sample mean of dice rolls

The sample average of the sequence $\frac{1}{n}\sum_{i=1}^{n} X_i$ of independent dice rolls.



The law of large numbers

Recall the weak law of large number: Let $X : \Omega \to \mathbb{R}$ be a random variable with a well-defined expectation $\mu = \mathbb{E}(X)$. Let $X_1, \cdots, X_n : \Omega \to \mathbb{R}$ be a sequence of independent copies of X. Then for all $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}\big(\big|\frac{1}{n}\sum X_i - \mu\big| \geq \epsilon\big) = 0$.

Intuition:

Suppose that X has expectation $\mathbb{E}(X) = \mu$ and variance $\mathrm{Var}(X) = \sigma^2$.

Since $X_1, \cdots, X_n$ are independent copies of X, we have $\mathbb{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$. So

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$$

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \left(\frac{1}{n}\right)^2\sum_{i=1}^{n}\mathrm{Var}(X_i) = \left(\frac{1}{n}\right)^2(n\cdot\sigma^2) = \frac{\sigma^2}{n}.$$

This suggests that $\frac{1}{n}\sum_{i=1}^{n} X_i - \mu$ is a random variable which shrinks to zero as n goes to infinity.

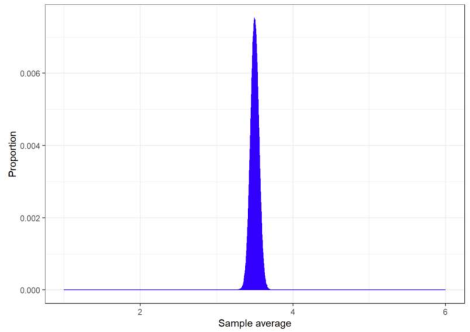
The law of large numbers

Recall the weak law of large number: Let $X : \Omega \to \mathbb{R}$ be a random variable with a well-defined expectation $\mu = \mathbb{E}(X)$. Let $X_1, \cdots, X_n : \Omega \to \mathbb{R}$ be a sequence of independent copies of X. Then for all $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum X_i - \mu\right| \geq \epsilon\right) = 0$.

Suppose that X has expectation $\mathbb{E}(X) = \mu$ and variance $\mathrm{Var}(X) = \sigma^2$.

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} \mathrm{Var}(X_i) = \left(\frac{1}{n}\right)^2 (n \cdot \sigma^2) = \frac{\sigma^2}{n}.$$

Sample size $n = 1000$



Looking into the distribution of $\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)$:

To gain further insight we can zoom in and consider the renormalised deviations

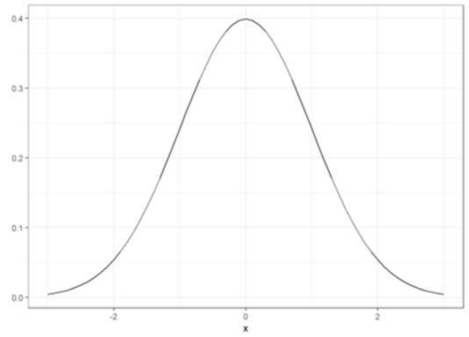
$$G_n := \sqrt{\frac{n}{\sigma^2}}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right).$$

Hence $\mathrm{Var}(G_n) = 1$ for all $n \in \mathbb{N}$.

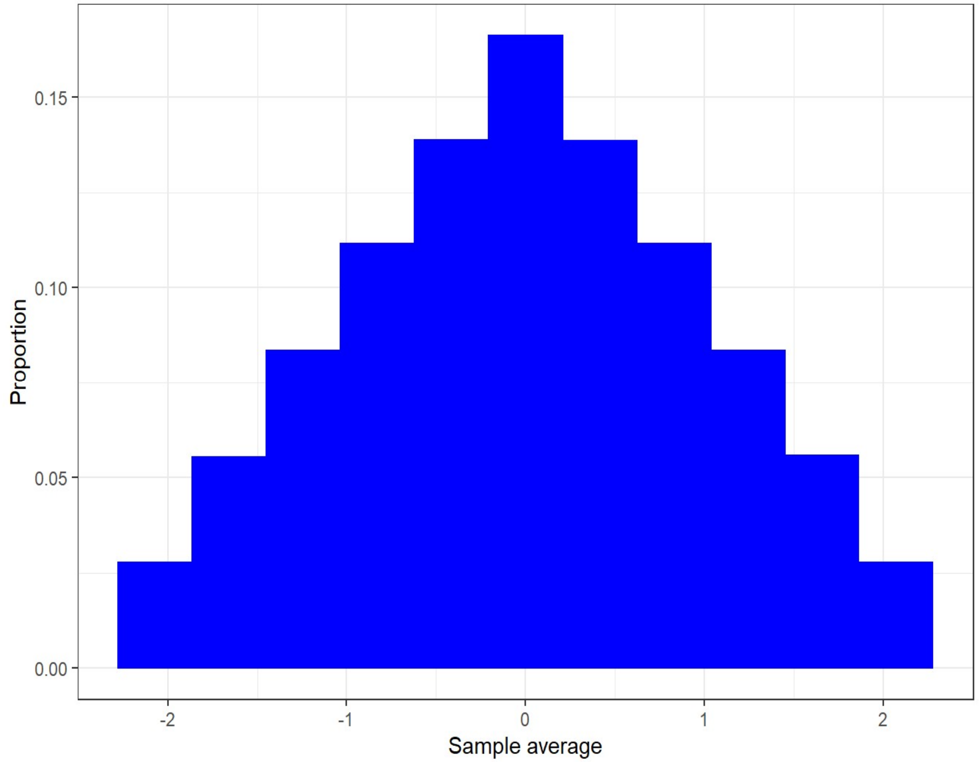
Sample mean of dice rolls (renormalized)

Renormalised sample average of independent dice rolls: $G_n := \sqrt{\frac{n}{\sigma^2}}\left(\frac{1}{n}\sum_{i=1}^n X_i - \mu\right)$.

Gaussian density function



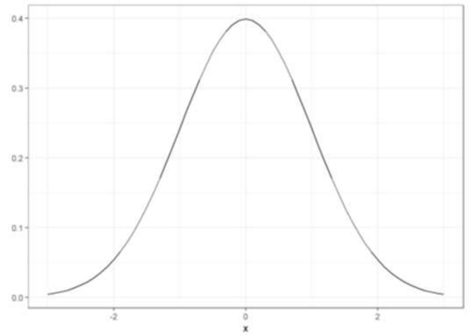
Sample size $n = 2$



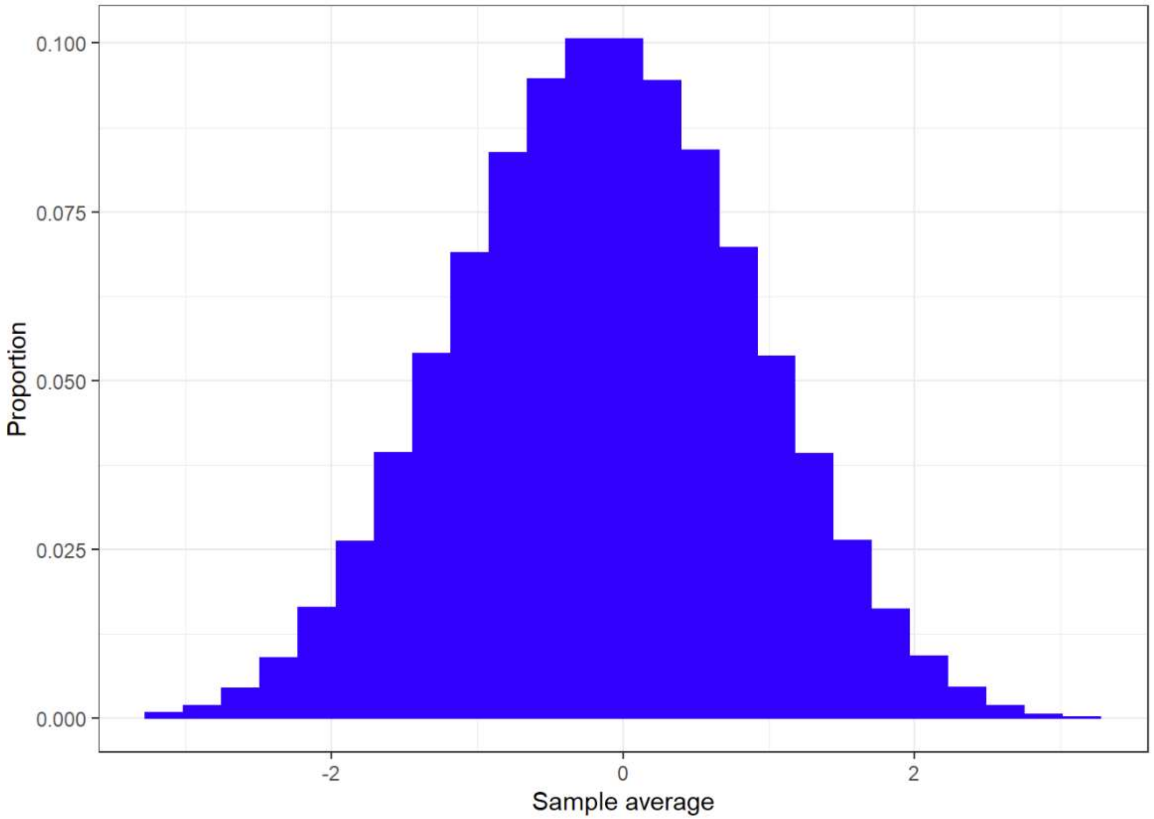
Sample mean of dice rolls (renormalized)

Renormalised sample average of independent dice rolls: $G_n := \sqrt{\frac{n}{\sigma^2}}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)$.

Gaussian density function



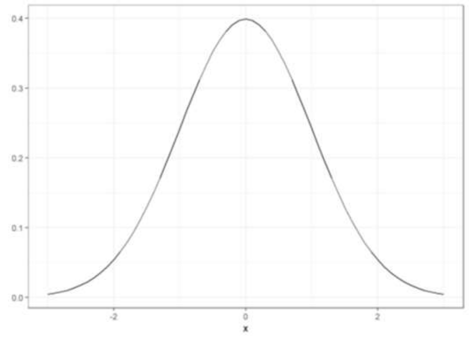
Sample size $n = 5$



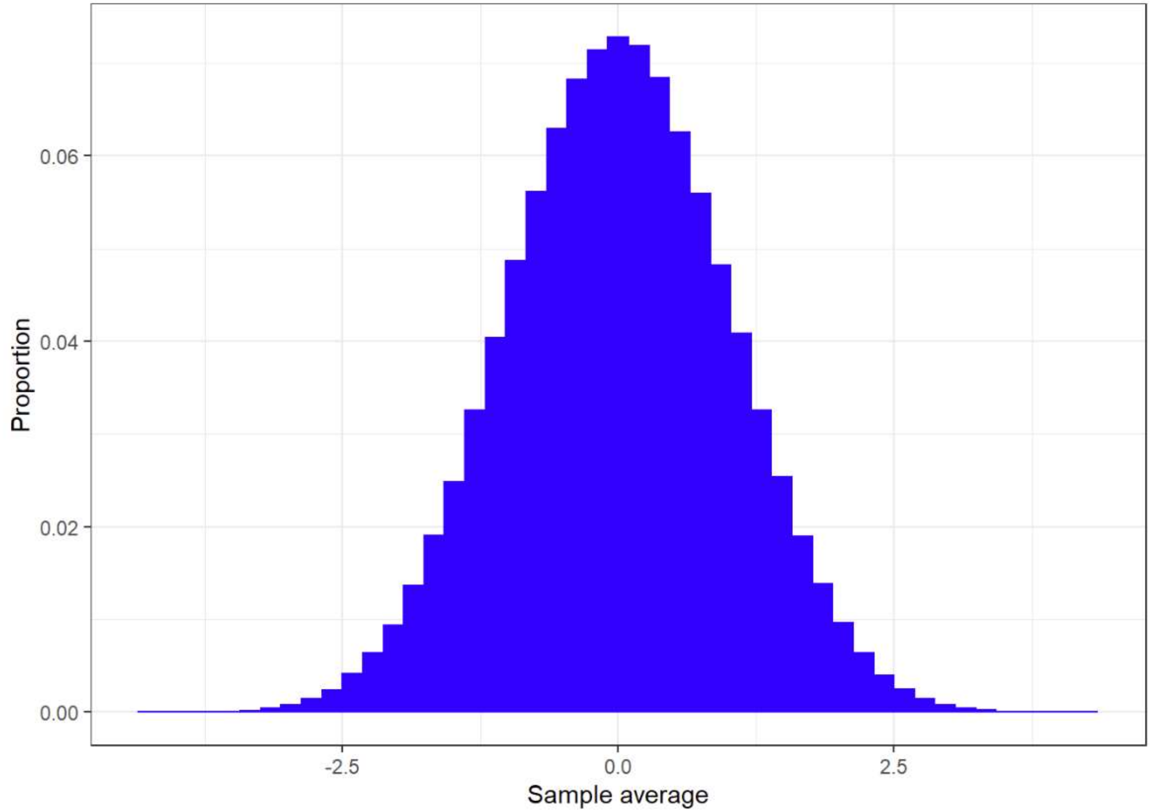
Sample mean of dice rolls (renormalized)

Renormalised sample average of independent dice rolls: $G_n := \sqrt{\frac{n}{\sigma^2}}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right).$

Gaussian density function



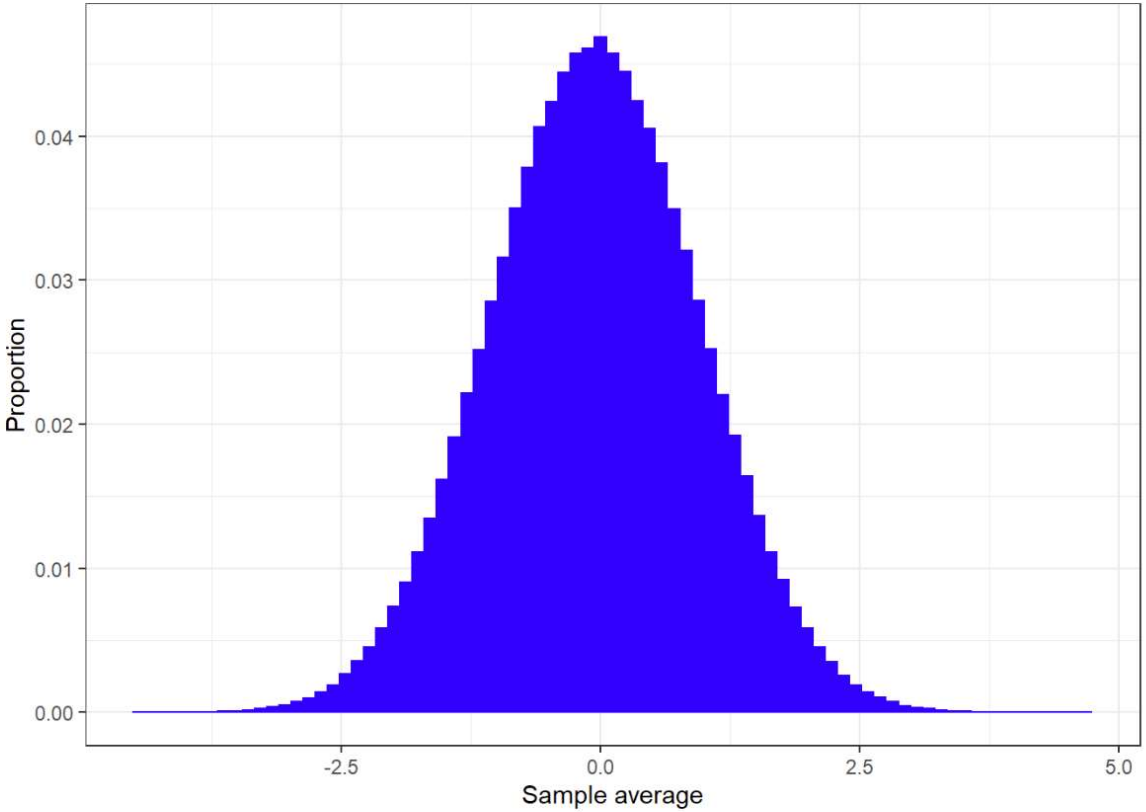
Sample size $n = 10$



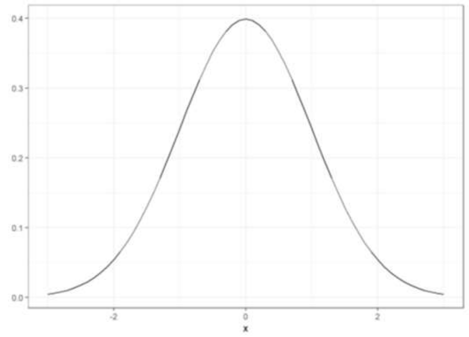
Sample mean of dice rolls (renormalized)

Renormalised sample average of independent dice rolls: $G_n := \sqrt{\frac{n}{\sigma^2}}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)$.

Sample size $n = 25$



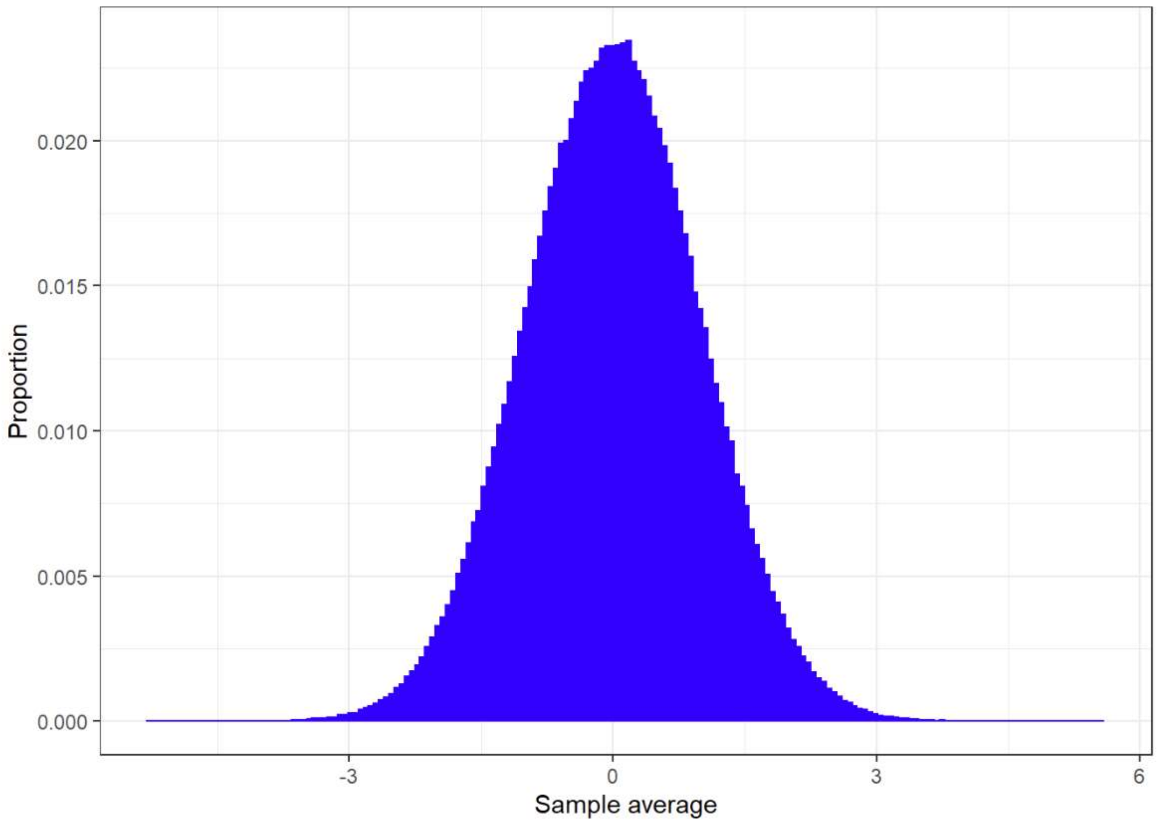
Gaussian density function



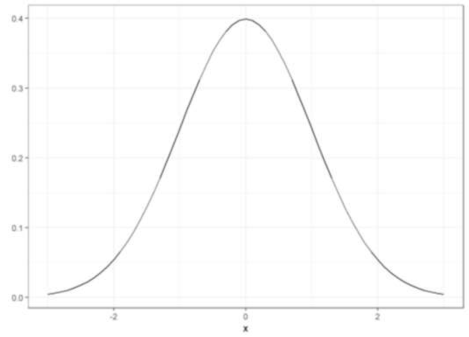
Sample mean of dice rolls (renormalized)

Renormalised sample average of independent dice rolls: $G_n := \sqrt{\frac{n}{\sigma^2}}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)$.

Sample size $n = 100$



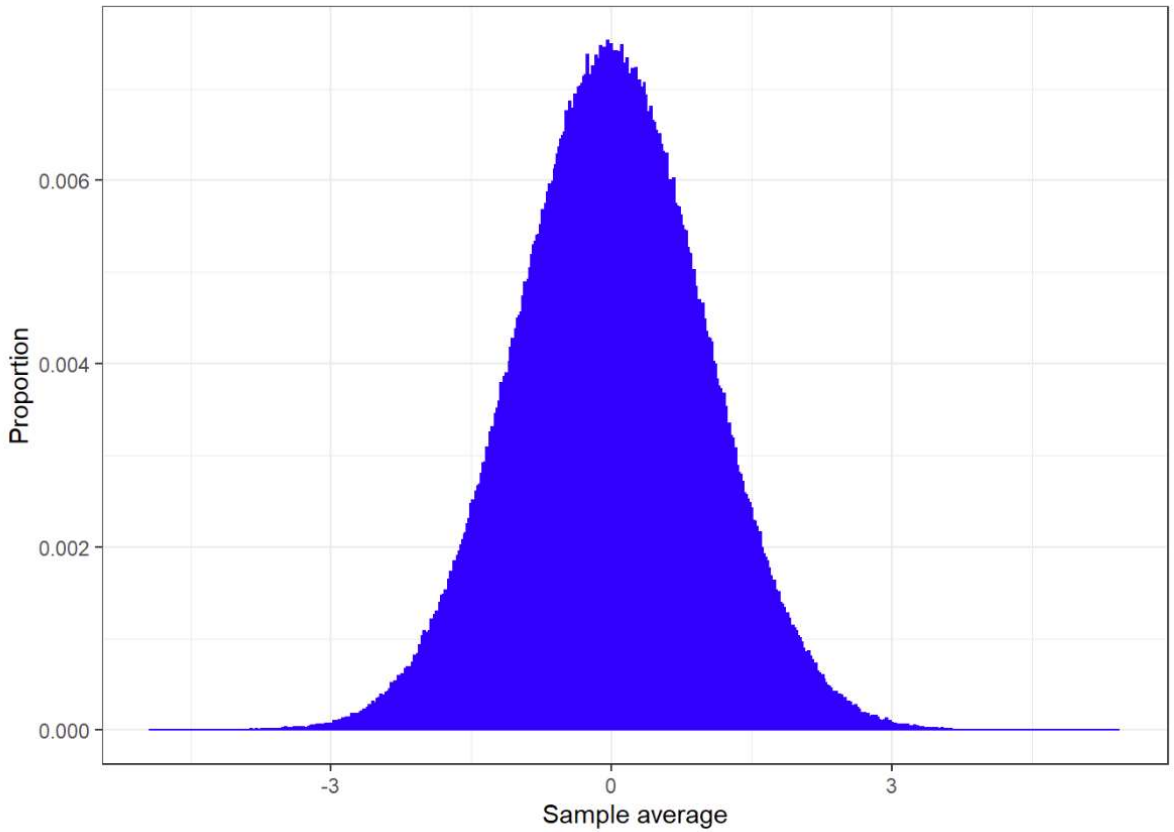
Gaussian density function



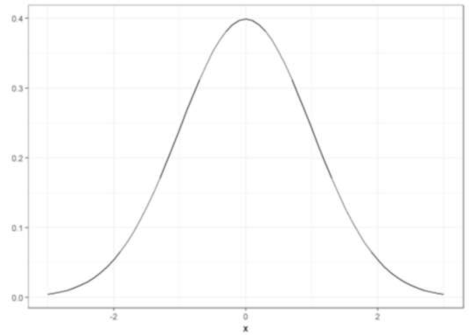
Sample mean of dice rolls (renormalized)

Renormalised sample average of independent dice rolls: $G_n := \sqrt{\frac{n}{\sigma^2}}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)$.

Sample size $n = 1000$



Gaussian density function



The central limit theorem

Theorem (Lindeberg—Lévy) The central limit theorem

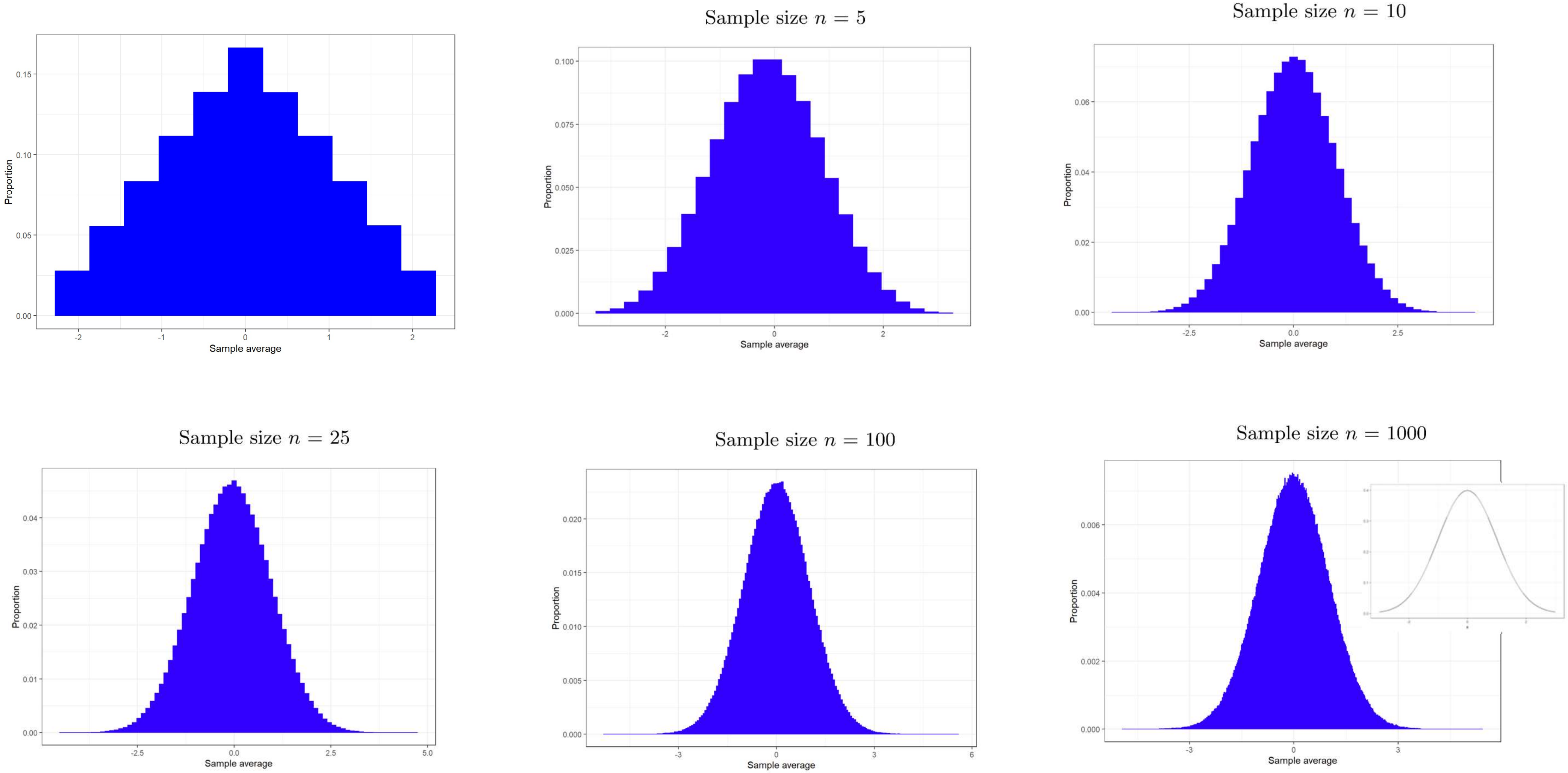
Let $X : \Omega \to \mathbb{R}$ be a random variable with expectation $\mu = \mathbb{E}(X)$ and variance $\sigma^2 = \mathrm{Var}(X)$. Let $X_1, \cdots, X_n : \Omega \to \mathbb{R}$ be a sequence of independent copies of X. Let $Z \sim \mathcal{N}(0,1)$ be a standard Gaussian random variable. Then for all $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\left\{\sqrt{\frac{n}{\sigma^2}} \cdot \left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \leq x\right\} = \mathbb{P}(Z \leq x).$$

The distribution of $G_n := \sqrt{\frac{n}{\sigma}}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)$ converges to the standard Gaussian distribution $\mathcal{N}(0,1)$.

Sample mean of dice rolls (renormalized)

Renormalised sample average of independent dice rolls: $G_n := \sqrt{\frac{n}{\sigma^2}}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)$.



What have we covered?

We introduced the concept of a continuous random variable.

We saw how continuous random variables can be understood via the probability density function.

We looked at expectation, variance, standard deviation, covariance and correlation in this context.

We discussed Gaussian random variables and briefly looked at chi-square & Student's t distributions.

We saw how the law of large numbers describes the limiting behaviour of the average.

We saw how the central limit theorem gives us greater insight as we zoom in around the mean.

University of BRISTOL

Thanks for listening!

Dr. Rihuan Ke
rihuan.ke@bristol.ac.uk

Statistical Computing and Empirical Methods
Unit EMATM0061, MSc Data Science