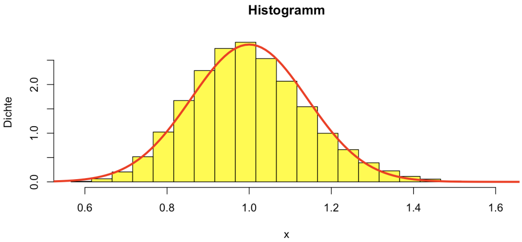
TECHNISCHE
UNIVERSITÄT
WIEN

Central limit theorem and Law of large numbers



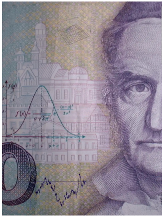
Please send you comments to: T. Levajković, tijana.levajkovic@tuwien.ac.at

Reminder

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

Reminder

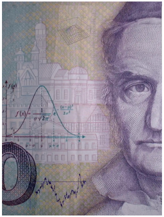
- $X \sim \mathcal{N}(\mu, \sigma^2)$
  - One of the most important distributions in statistics and probability theory.



Normal distribution on the front of the German 10 Mark banknote from 1990s

Reminder

- $X \sim \mathcal{N}(\mu, \sigma^2)$
  - One of the most important distributions in statistics and probability theory.



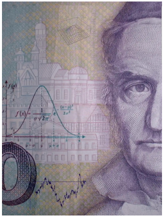
Normal distribution on the front of the German 10 Mark banknote from 1990s

  - X has a Normal distribution (or Gauss distribution) with the parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, if its pdf is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{for} \quad x \in \mathbb{R}$$

Reminder

- $X \sim \mathcal{N}(\mu, \sigma^2)$
  - One of the most important distributions in statistics and probability theory.



Normal distribution on the front of the German 10 Mark banknote from 1990s

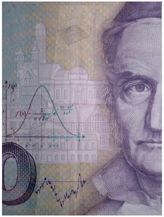
- X has a Normal distribution (or Gauss distribution) with the parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, if its pdf is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{for} \quad x \in \mathbb{R}$$

  - $E(X) = \mu, \quad Var(X) = \sigma^2$

Reminder

- $X \sim \mathcal{N}(\mu, \sigma^2)$
  - One of the most important distributions in statistics and probability theory.



Normal distribution on the front of the German 10 Mark banknote from 1990s

  - X has a Normal distribution (or Gauss distribution) with the parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, if its pdf is given by

$$f(x) = \frac{1}{\sqrt{2\pi\,\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{for} \quad x \in \mathbb{R}$$

    - $E(X) = \mu, \quad Var(X) = \sigma^2$

$Y = aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

We recall

- $Z \sim \mathcal{N}(0,1)$ is the Standard normal random variable

We recall

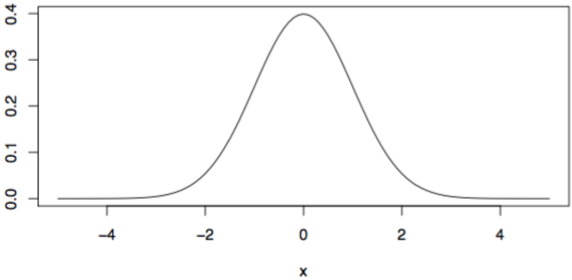
- $Z \sim \mathcal{N}(0,1)$ is the Standard normal random variable

... $E(Z) = 0, \;\; Var(Z) = 1$

We recall

- $Z \sim \mathcal{N}(0,1)$ is the Standard normal random variable

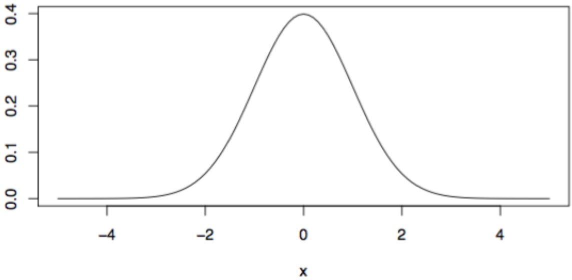
... $E(Z) = 0, \;\; Var(Z) = 1$



We recall

- $Z \sim \mathcal{N}(0,1)$ is the Standard normal random variable

... $E(Z) = 0, \;\; Var(Z) = 1$

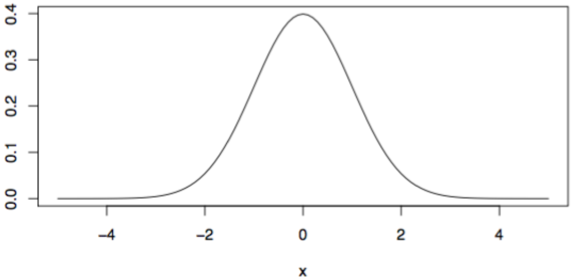


- Its cdf is given by $\Phi(z) = P(Z \leqslant z)$

We recall

- $Z \sim \mathcal{N}(0, 1)$ is the Standard normal random variable

... $E(Z) = 0, \quad Var(Z) = 1$



- Its cdf is given by $\Phi(z) = P(Z \leqslant z)$
- The following hold

$$P(-1 \leqslant Z \leqslant 1) \approx 0.68$$
$$P(-2 \leqslant Z \leqslant 2) \approx 0.95$$
$$P(-3 \leqslant Z \leqslant 3) \approx 0.997$$

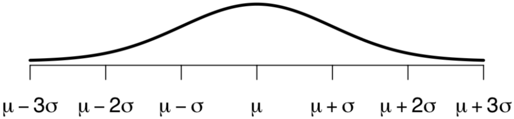
We recall

- Let $X \sim \mathcal{N}(\mu, \sigma^2)$.
  - Standardization

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

We recall

- Let $X \sim \mathcal{N}(\mu, \sigma^2)$.
  - Standardization
$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$
  - For $X \sim \mathcal{N}(\mu, \sigma^2)$ the 68-95-99.7-Rule holds

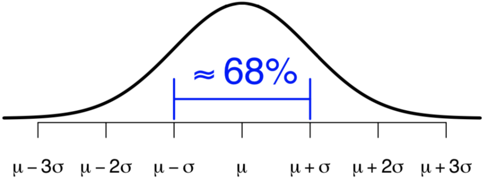


We recall

- Let $X \sim \mathcal{N}(\mu, \sigma^2)$.
  - Standardization

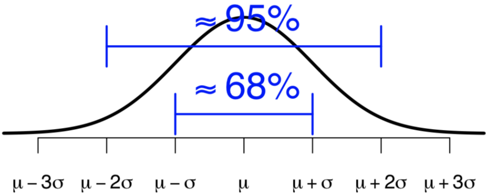
$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

  - For $X \sim \mathcal{N}(\mu, \sigma^2)$ the 68-95-99.7-Rule holds



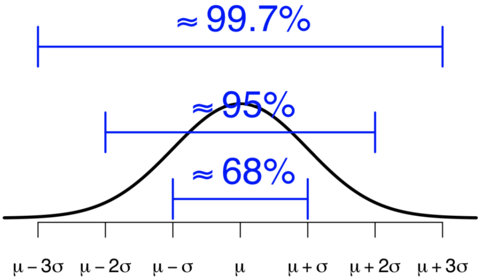
We recall

- Let $X \sim \mathcal{N}(\mu, \sigma^2)$.
  - Standardization
  $$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$
  - For $X \sim \mathcal{N}(\mu, \sigma^2)$ the 68-95-99.7-Rule holds



We recall

- Let $X \sim \mathcal{N}(\mu, \sigma^2)$.
  - Standardization
$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$
  - For $X \sim \mathcal{N}(\mu, \sigma^2)$ the 68-95-99.7-Rule holds



Sample mean

- Let $X_1, X_2, \ldots, X_n$ be independent identically distributed (i.i.d.) random variables with $E(X_1) = \mu$ and $Var(X_1) = \sigma^2$.

Sample mean

- Let $X_1, X_2, \ldots, X_n$ be independent identically distributed (i.i.d.) random variables with $E(X_1) = \mu$ and $Var(X_1) = \sigma^2$.

... $X_1, X_2, \ldots, X_n$ is allso called a random sample of size n

Sample mean

- Let $X_1, X_2, \ldots, X_n$ be independent identically distributed (i.i.d.) random variables with $E(X_1) = \mu$ and $Var(X_1) = \sigma^2$.

... $X_1, X_2, \ldots, X_n$ is allso called a random sample of size n

- For the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ it holds

$$E(\bar{X}) = \mu \qquad \text{and} \qquad Var(\bar{X}) = \frac{\sigma^2}{n}$$

Sample mean

- Let $X_1, X_2, \ldots, X_n$ be independent identically distributed (i.i.d.) random variables with $E(X_1) = \mu$ and $Var(X_1) = \sigma^2$.

... $X_1, X_2, \ldots, X_n$ is allso called a random sample of size n

- For the sample mean $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ it holds

$$E(\bar{X}) = \mu \qquad \text{and} \qquad Var(\bar{X}) = \frac{\sigma^2}{n}$$

- $$E(\bar{X}) = E\Big(\frac{1}{n} \sum_{i=1}^{n} X_i\Big) \overset{\text{linearity}}{=} \frac{1}{n} \sum_{i=1}^{n} E(X_i) \overset{\text{ident. distr.}}{=} \frac{1}{n} \sum_{i=1}^{n} E(X_1)$$
$$= \frac{1}{n} \cdot n\, E(X_1) = E(X_1) = \mu$$

Sample mean

- Let $X_1, X_2, \ldots, X_n$ be independent identically distributed (i.i.d.) random variables with $E(X_1) = \mu$ and $Var(X_1) = \sigma^2$.

... $X_1, X_2, \ldots, X_n$ is also called a random sample of size n

- For the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ it holds

$$E(\bar{X}) = \mu \qquad \text{and} \qquad Var(\bar{X}) = \frac{\sigma^2}{n}$$

- $E(\bar{X}) = E\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) \overset{\text{linearity}}{=} \frac{1}{n}\sum_{i=1}^{n} E(X_i) \overset{\text{ident. distr.}}{=} \frac{1}{n}\sum_{i=1}^{n} E(X_1)$
$= \frac{1}{n} \cdot n\, E(X_1) = E(X_1) = \mu$

- $Var(\bar{X}) = Var\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n^2} Var\Big(\sum_{i=1}^{n} X_i\Big) \overset{\text{independ.}}{=} \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i)$
$\overset{\text{ident. distr.}}{=} \frac{1}{n^2}\sum_{i=1}^{n} Var(X_1) = \frac{1}{n^2} \cdot n\, Var(X_1) = \frac{Var(X_1)}{n} = \frac{\sigma^2}{n}$

Sample mean

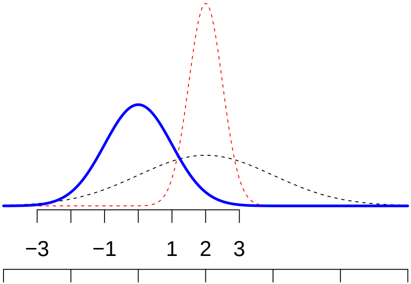
- Let $X_1, X_2, \ldots, X_n$ be independent identically distributed (i.i.d.) random variables with $E(X_1) = \mu$ and $Var(X_1) = \sigma^2$.

... $X_1, X_2, \ldots, X_n$ is allso called a random sample of size n

- For the sample mean $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ it holds

$$E(\bar{X}) = \mu \qquad \text{and} \qquad Var(\bar{X}) = \frac{\sigma^2}{n}$$

  - Especially if $X_1, \ldots X_n$ are normally distributed, the sample mean is also normally distributed $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$



Convergence

- Two important theorems of probability theory:

Convergence

- Two important theorems of probability theory:
  - Law of Large Numbers (LLN)

Convergence

- Two important theorems of probability theory:
  - Law of Large Numbers (LLN)
    - in German: Gesetz der großen Zahlen (GGZ)

Convergence

- Two important theorems of probability theory:
    - Law of Large Numbers (LLN)
        - in German: Gesetz der großen Zahlen (GGZ)
    - Central Limit Theorem (CLT)

Convergence

- Two important theorems of probability theory:
  - Law of Large Numbers (LLN)
    - in German: Gesetz der großen Zahlen (GGZ)
  - Central Limit Theorem (CLT)
    - in German: Zentraler Grenzwertsatz (ZGWS)

Convergence

- Two important theorems of probability theory:
  - Law of Large Numbers (LLN)
    - in German: Gesetz der großen Zahlen (GGZ)
  - Central Limit Theorem (CLT)
    - in German: Zentraler Grenzwertsatz (ZGWS)

... We state the theorems for

independent and identically distributed random variables

Theorems

- Let $X_1, \dots, X_n$ be independent and identically distributed (i.i.d.) random variables with $E(X_1) = \mu$ and finite $Var(X_1) = \sigma^2 < \infty$.

... n i.i.d. random variables $X_1, \dots, X_n$ are called a random sample of size n

  - Note that $X_1, \dots, X_n$ are not necessarily normally distributed.

- We consider the sample mean

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

  - The Law of large numbers and the Central limit theorem provide information about the value and the distribution of $\bar{X}_n$.

    - LLN: As n grows, the probability of $\bar{X}_n$ being in the neighborhood of μ tends to 1.

    - CLT: For large n, the distribution of $\bar{X}_n$ is approximately $\mathcal{N}(\mu, \frac{\sigma^2}{n})$.

Law of large numbers

- The Law of large numbers says that the sample mean $\bar{X}_n$ will be with high probability very close to the expectation μ of the underlying distribution

Law of large numbers

- The Law of large numbers says that the sample mean $\bar{X}_n$ will be with high probability very close to the expectation μ of the underlying distribution

- Formal statement:

Law of large numbers

- The Law of large numbers says that the sample mean $\bar{X}_n$ will be with high probability very close to the expectation μ of the underlying distribution

- Formal statement:

  Suppose $X_1, X_2, \ldots, X_n \ldots$ are i.i.d. random variables with expectation μ and finite variance σ^2. For each n, let $\bar{X}_n$ be the mean of the first n variables. Then for any $a > 0$, we have

$$\lim_{n \to \infty} P(|\bar{X}_n - \mu| < a) = 1$$

Example

- We consider a discrete random variable with the pmf

x	0	1	2	3	4
$p(x)$	0.1	0.2	0.3	0.35	0.05

Example

- We consider a discrete random variable with the pmf

x	0	1	2	3	4
$p(x)$	0.1	0.2	0.3	0.35	0.05

- Then, $E(X) = \sum_{x=0}^{4} x\,p(x) = 2.05$ and $Var(X) = \sum_{x=0}^{4} (x - 2.05)^2\,p(x) = 1.1475$

Example

- We consider a discrete random variable with the pmf

x	0	1	2	3	4
$p(x)$	0.1	0.2	0.3	0.35	0.05

  - Then, $E(X) = \sum_{x=0}^{4} x\,p(x) = 2.05$ and $Var(X) = \sum_{x=0}^{4} (x - 2.05)^2\,p(x) = 1.1475$

    - Since the variance of X exists, the Law of Large Numbers says

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \;\to\; E(X) = 2.05 \quad \text{for large } n$$

Example

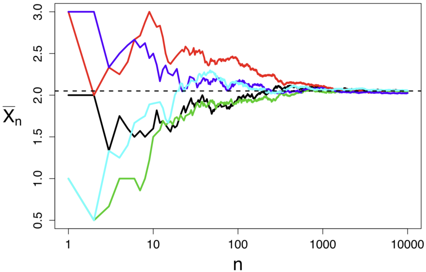
- We consider a discrete random variable with the pmf

x	0	1	2	3	4
$p(x)$	0.1	0.2	0.3	0.35	0.05

  - Then, $E(X) = \sum_{x=0}^{4} x\,p(x) = 2.05$ and $Var(X) = \sum_{x=0}^{4} (x-2.05)^2\,p(x) = 1.1475$

- Since the variance of X exists, the Law of Large Numbers says

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \;\rightarrow\; E(X) = 2.05 \quad \text{for large } n$$



Example 2

- The mean values of the Bernoulli random variables

Example 2

- The mean values of the Bernoulli random variables
  - Let $X_1 ..., X_n$ denote n independent flips of a fair coin

Example 2

- The mean values of the Bernoulli random variables
  - Let $X_1 \ldots, X_n$ denote n independent flips of a fair coin
    - $X_i \sim ber(0.5)$ and $E(X_i) = \mu = 0.5$

Example 2

- The mean values of the Bernoulli random variables
  - Let $X_1 ..., X_n$ denote n independent flips of a fair coin
    - $X_i \sim ber(0.5)$ and $E(X_i) = \mu = 0.5$
    - $S_n = X_1 + \cdots + X_n \sim B(n, 0.5)$

Example 2

- The mean values of the Bernoulli random variables
  - Let $X_1 ..., X_n$ denote n independent flips of a fair coin
    - $X_i \sim ber(0.5)$ and $E(X_i) = \mu = 0.5$
    - $S_n = X_1 + \cdots + X_n \sim B(n, 0.5)$
  - Then $\bar{X}_n = \frac{1}{n} S_n$ is the number of heads obtained in n flips.

Example 2

- The mean values of the Bernoulli random variables
  - Let $X_1 ..., X_n$ denote n independent flips of a fair coin
    - $X_i \sim ber(0.5)$ and $E(X_i) = \mu = 0.5$
    - $S_n = X_1 + \cdots + X_n \sim B(n, 0.5)$
  - Then $\bar{X}_n = \frac{1}{n} S_n$ is the number of heads obtained in n flips.
  - We expect that this number for large n is close to $\mu = 0.5$.
    - LLN: The sample mean $\bar{X}_n$ is very likely to be close to the expected value 0.5 for large n.

Example 2

- The mean values of the Bernoulli random variables
  - Let $X_1 ..., X_n$ denote n independent flips of a fair coin
    - $X_i \sim ber(0.5)$ and $E(X_i) = \mu = 0.5$
    - $S_n = X_1 + \cdots + X_n \sim B(n, 0.5)$
  - Then $\bar{X}_n = \frac{1}{n} S_n$ is the number of heads obtained in n flips.
  - We expect that this number for large n is close to $\mu = 0.5$.
    - LLN: The sample mean $\bar{X}_n$ is very likely to be close to the expected value 0.5 for large n.
    - For example, let us consider the probability of being within 0.1 from the expected value $\mu = 0.5$, i.e.

$$P(|\bar{X}_n - \mu| < 0.1) \qquad \text{or} \quad \text{equivalently} \qquad P(0.4 < \bar{X}_n < 0.6)$$

Example 2

- The mean values of the Bernoulli random variables
  - Let $X_1 ..., X_n$ denote n independent flips of a fair coin
    - $X_i \sim ber(0.5)$ and $E(X_i) = \mu = 0.5$
    - $S_n = X_1 + \cdots + X_n \sim B(n, 0.5)$
  - Then $\bar{X}_n = \frac{1}{n} S_n$ is the number of heads obtained in n flips.
  - We expect that this number for large n is close to $\mu = 0.5$.
    - LLN: The sample mean $\bar{X}_n$ is very likely to be close to the expected value 0.5 for large n.
    - For example, let us consider the probability of being within 0.1 from the expected value $\mu = 0.5$, i.e.

$$P(|\bar{X}_n - \mu| < 0.1) \qquad \text{or} \quad \text{equivalently} \qquad P(0.4 < \bar{X}_n < 0.6)$$

    - in R:

$n = 10$:	`diff(pbinom(c(3,6),10,0.5))`	= 0.65625
$n = 50$:	`diff(pbinom(c(19,30), 50, 0.5))`	= 0.8810795
$n = 100$:	`diff(pbinom(c(39,60), 100, 0.5)`	= 0.9647998
$n = 500$:	`diff(pbinom(c(199,300), 500, 0.5))`	= 0.9999941
$n = 1000$:	`diff(pbinom(c(399,600), 1000, 0.5))`	= 1

Example 2

- The mean values of the Bernoulli random variables
  - Let $X_1 \ldots, X_n$ denote n independent flips of a fair coin
    - $X_i \sim ber(0.5)$ and $E(X_i) = \mu = 0.5$
    - $S_n = X_1 + \cdots + X_n \sim B(n, 0.5)$
  - Then $\bar{X}_n = \frac{1}{n} S_n$ is the number of heads obtained in n flips.
  - We expect that this number for large n is close to $\mu = 0.5$.
    - LLN: The sample mean $\bar{X}_n$ is very likely to be close to the expected value 0.5 for large n.
    - For example, let us consider the probability of being within 0.1 from the expected value $\mu = 0.5$, i.e.

$$P(|\bar{X}_n - \mu| < 0.1) \qquad \text{or} \quad \text{equivalently} \qquad P(0.4 < \bar{X}_n < 0.6)$$

    - in R:

$n = 10$:	`diff(pbinom(c(3,6),10,0.5))`	$= 0.65625$
$n = 50$:	`diff(pbinom(c(19,30), 50, 0.5))`	$= 0.8810795$
$n = 100$:	`diff(pbinom(c(39,60), 100, 0.5)`	$= 0.9647998$
$n = 500$:	`diff(pbinom(c(199,300), 500, 0.5))`	$= 0.9999941$
$n = 1000$:	`diff(pbinom(c(399,600), 1000, 0.5))`	$= 1$

HW Repeat the calculations and determine the probability

$$P(|\bar{X}_n - \mu| < 0.01).$$

Histograms

- Histograms (frequency and density histogram)

Histograms

- Histograms (frequency and density histogram)

... A graphical representation of a frequency distribution based on a previous classification of the data.

Histograms

- Histograms (frequency and density histogram)

  ... A graphical representation of a frequency distribution based on a previous classification of the data.

  - Made by *binning* data ... data binning is a way to group a number of continuous values into a smaller number of *bins*

Histograms

- Histograms (frequency and density histogram)

  ... A graphical representation of a frequency distribution based on a previous classification of the data.

  - Made by *binning* data ... data binning is a way to group a number of continuous values into a smaller number of *bins*
  - **Frequency:** height of bar over bin = number of data points in bin.

Histograms

- Histograms (frequency and density histogram)

  ... A graphical representation of a frequency distribution based on a previous classification of the data.

  - Made by *binning* data ... data binning is a way to group a number of continuous values into a smaller number of *bins*
  - **Frequency:** height of bar over bin = number of data points in bin.
  - **Density:** area of bar is the fraction of all data points that lie in the bin.

    ... Thus, total area is 1.

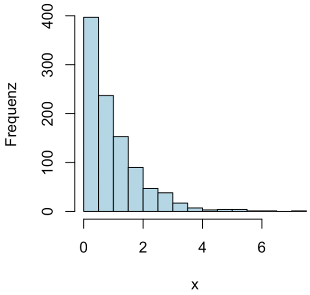
Histograms

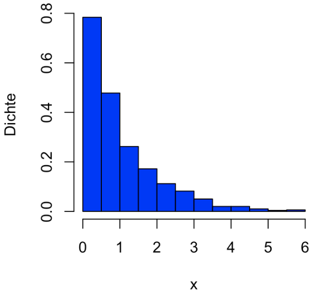
- Histograms (frequency and density histogram)

... A graphical representation of a frequency distribution based on a previous classification of the data.

  - Made by *binning* data ... data binning is a way to group a number of continuous values into a smaller number of *bins*
  - **Frequency:** height of bar over bin = number of data points in bin.
  - **Density:** area of bar is the fraction of all data points that lie in the bin.

... Thus, total area is 1.





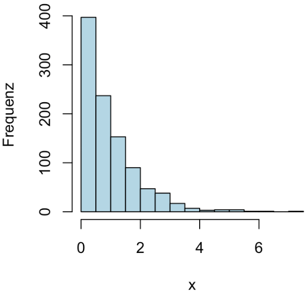
Histograms

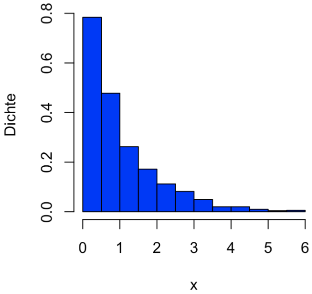
- Histograms (frequency and density histogram)

  ... A graphical representation of a frequency distribution based on a previous classification of the data.

  - Made by *binning* data ... data binning is a way to group a number of continuous values into a smaller number of *bins*
  - **Frequency:** height of bar over bin = number of data points in bin.
  - **Density:** area of bar is the fraction of all data points that lie in the bin.

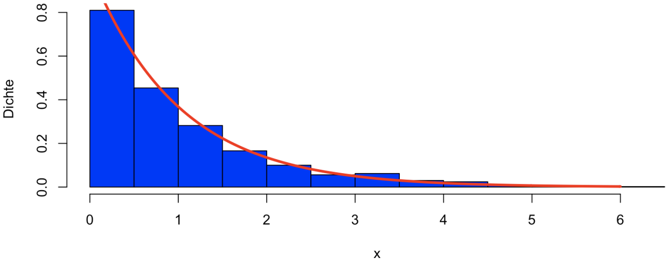
    ... Thus, total area is 1.





Law of Large Numbers and histograms

- The Law of Large Numbers implies that density histogram converges to probability density function.



The histogram with bin width 0.5 showing 1000 draws from an exponential $\exp(1)$ distribution. The pdf of $\exp(1)$ is given in red.

Central Limit Theorem

- The Central Limit Theorem states that the sum, resp. mean, of many independent copies of a random variable is approximately a normal random variable.

Central Limit Theorem

- The Central Limit Theorem states that the sum, resp. mean, of many independent copies of a random variable is approximately a normal random variable.
- Let $X_1, X_2, \dots$ be a sequence of i.i.d. with expectation μ and standard deviation σ.

Central Limit Theorem

- The Central Limit Theorem states that the sum, resp. mean, of many independent copies of a random variable is approximately a normal random variable.

- Let $X_1, X_2, \ldots$ be a sequence of i.i.d. with expectation μ and standard deviation σ.

    - For each n:
        - the sum $S_n = X_1 + X_2 + \cdots + X_n$
        - the sample mean $\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$

Central Limit Theorem

- The Central Limit Theorem states that the sum, resp. mean, of many independent copies of a random variable is approximately a normal random variable.

- Let $X_1, X_2, \ldots$ be a sequence of i.i.d. with expectation μ and standard deviation σ.

  - For each n:
    - the sum $S_n = X_1 + X_2 + \cdots + X_n$
    - the sample mean $\bar{X}_n = \frac{1}{n}\left(X_1 + X_2 + \cdots + X_n\right)$

  - The Central Limit Theorem says that for large n it holds

$$
\begin{aligned}
S_n &\approx \quad \mathcal{N}(n\mu, n\sigma^2) \\
\bar{X}_n &\approx \quad \mathcal{N}(\mu, \frac{\sigma^2}{n})
\end{aligned}
$$

Central Limit Theorem

- The Central Limit Theorem states that the sum, resp. mean, of many independent copies of a random variable is approximately a normal random variable.

- Let $X_1, X_2, \ldots$ be a sequence of i.i.d. with expectation μ and standard deviation σ.

  - For each n:
    - the sum $S_n = X_1 + X_2 + \cdots + X_n$
    - the sample mean $\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$

  - The Central Limit Theorem says that for large n it holds

$$\begin{aligned} S_n &\approx \quad \mathcal{N}(n\mu, n\sigma^2) \\ \bar{X}_n &\approx \quad \mathcal{N}(\mu, \frac{\sigma^2}{n}) \end{aligned}$$

    - Thus, the standardized $\bar{X}_n$ and S_n have approximately $\mathcal{N}(0,1)$, i.e.

$$\frac{S_n - n\mu}{\sqrt{n}\,\sigma} \approx \mathcal{N}(0,1) \quad \text{and} \quad \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \approx \mathcal{N}(0,1)$$

Example

- Simulation for CLT

Example

- Simulation for CLT
  - Let $X_1, X_2, \ldots$ be a sequence of i.i.d. $\exp(1)$ random variables.

Example

- Simulation for CLT
  - Let $X_1, X_2, \ldots$ be a sequence of i.i.d. $\exp(1)$ random variables.
    - Then, $E(X_i) = 1$ and $Var(X_i) = 1 < \infty$.
  - We consider sample means $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$.

Example

- Simulation for CLT
  - Let $X_1, X_2, \dots$ be a sequence of i.i.d. $\exp(1)$ random variables.
    - Then, $E(X_i) = 1$ and $Var(X_i) = 1 < \infty$.
  - We consider sample means $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$.
  - Central Limit Theorem:

Example

- Simulation for CLT
  - Let $X_1, X_2, \ldots$ be a sequence of i.i.d. $\exp(1)$ random variables.
    - Then, $E(X_i) = 1$ and $Var(X_i) = 1 < \infty$.
  - We consider sample means $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$.
  - Central Limit Theorem:
    - $\bar{X}_n$: $\quad E(\bar{X}_n) = 1 \quad$ and $\quad Var(\bar{X}_n) = \frac{1}{n}$

$$\bar{X}_n \approx \mathcal{N}(1, \frac{1}{n})$$

Example

- Simulation for CLT
  - Let $X_1, X_2, \ldots$ be a sequence of i.i.d. $\exp(1)$ random variables.
    - Then, $E(X_i) = 1$ and $Var(X_i) = 1 < \infty$.
  - We consider sample means $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$.
  - Central Limit Theorem:
    - $\bar{X}_n$: $E(\bar{X}_n) = 1$ and $Var(\bar{X}_n) = \frac{1}{n}$

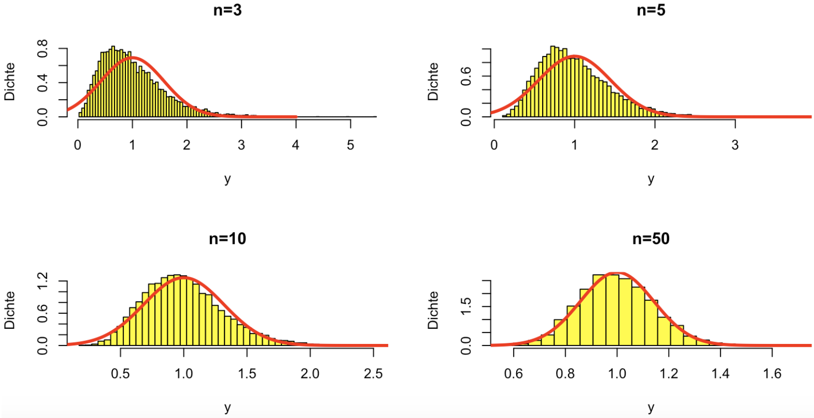
$$\bar{X}_n \approx \mathcal{N}(1, \frac{1}{n})$$

    - The standardized means Y_n

$$Y_n = \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{Var(\bar{X}_n)}} = \frac{\bar{X}_n - 1}{\frac{1}{\sqrt{n}}} \approx \mathcal{N}(0, 1)$$

Example

- The following plots show the results for respectively $n = 3, 5, 10, 50$ based on 10000 simulated values for X_n. The density of the $\mathcal{N}(1, \frac{1}{n})$ distribution is drawn in red.



Normal approximation of $B(n,p)$ distribution

- The sum $S_n = X_1 + \cdots + X_n$ of independent $ber(p)$ random variables $X_1, \ldots, X_n$ has $B(n,p)$ distribution.

Normal approximation of $B(n, p)$ distribution

- The sum $S_n = X_1 + \cdots + X_n$ of independent $ber(p)$ random variables $X_1, \ldots, X_n$ has $B(n, p)$ distribution.
  - $E(S_n) = np$ and $Var(S_n) = np(1 - p)$

Normal approximation of $B(n, p)$ distribution

- The sum $S_n = X_1 + \cdots + X_n$ of independent $ber(p)$ random variables $X_1, \ldots, X_n$ has $B(n, p)$ distribution.
  - $E(S_n) = np$ and $Var(S_n) = np(1-p)$
- Using the continuity correction it holds for $a \leqslant b$, where $a, b \in \{0, 1, \ldots, n\}$:

$$P(a \leqslant S_n \leqslant b) \approx \Phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

Normal approximation of $B(n,p)$ distribution

- The sum $S_n = X_1 + \cdots + X_n$ of independent $ber(p)$ random variables $X_1, \ldots, X_n$ has $B(n,p)$ distribution.
  - $E(S_n) = np$ and $Var(S_n) = np(1-p)$
- Using the continuity correction it holds for $a \leqslant b$, where $a, b \in \{0, 1, \ldots, n\}$:

$$P(a \leqslant S_n \leqslant b) \approx \Phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

  - Let T be a continuous random variable such that $E(T) = E(S_n) = np$ and $Var\,(T) = Var\,(S_n) = np(1-p)$. Then,

$$\begin{aligned}
P(a \leqslant S_n \leqslant b) &= P\left(a - \frac{1}{2} \leqslant S_n < b + \frac{1}{2}\right) \approx P\left(a - \frac{1}{2} \leqslant T < b + \frac{1}{2}\right) \\
&= P\left(\frac{a - \frac{1}{2} - E(T)}{\sqrt{Var\,(T)}} \leqslant \frac{T - E(T)}{\sqrt{Var\,(T)}} < \frac{b + \frac{1}{2} - E(T)}{\sqrt{Var\,(T)}}\right) \\
&\overset{\text{CLT}}{\approx} P\left(\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}} \leqslant Z < \frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) \\
&= \Phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)
\end{aligned}$$

Normal approximation of $B(n, p)$ distribution

- The sum $S_n = X_1 + \cdots + X_n$ of independent $ber(p)$ random variables $X_1, \ldots, X_n$ has $B(n, p)$ distribution.
  - $E(S_n) = np$ and $Var(S_n) = np(1-p)$
- Using the continuity correction it holds for $a \leqslant b$, where $a, b \in \{0, 1, \ldots, n\}$:

$$P(a \leqslant S_n \leqslant b) \approx \Phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

Normal approximation of $B(n,p)$ distribution

- The sum $S_n = X_1 + \cdots + X_n$ of independent $ber(p)$ random variables $X_1, \ldots, X_n$ has $B(n,p)$ distribution.
  - $E(S_n) = np$ and $Var(S_n) = np(1-p)$
- Using the continuity correction it holds for $a \leqslant b$, where $a, b \in \{0, 1, \ldots, n\}$:

$$P(a \leqslant S_n \leqslant b) \approx \Phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

- Approximation of the cdf of S_n:

$$P(S_n \leqslant x) \approx \Phi\left(\frac{x + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

Φ is the cdf of $\mathcal{N}(0,1)$

Normal approximation of $B(n, p)$ distribution

- The sum $S_n = X_1 + \cdots + X_n$ of independent $ber(p)$ random variables $X_1, \ldots, X_n$ has $B(n, p)$ distribution.
  - $E(S_n) = np$ and $Var(S_n) = np(1-p)$
- Using the continuity correction it holds for $a \leqslant b$, where $a, b \in \{0, 1, \ldots, n\}$:

$$P(a \leqslant S_n \leqslant b) \approx \Phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

- Approximation of the cdf of S_n:

$$P(S_n \leqslant x) \approx \Phi\left(\frac{x + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

Φ is the cdf of $\mathcal{N}(0, 1)$

- Rule of thumb: The approximation is considered reasonable, when $\min\{np, np(1-p)\} \geqslant 10$

Normal approximation of $\mathcal{P}(\lambda)$ distribution

- Let $X \sim \mathcal{P}(\lambda)$. Then, by providing the continuity correction for $a \leqslant b$, with $a, b \in \mathbb{N}_0$

$$P(a \leqslant X \leqslant b) \approx \Phi\left(\frac{b + \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{a - \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right),$$

Normal approximation of $\mathcal{P}(\lambda)$ distribution

- Let $X \sim \mathcal{P}(\lambda)$. Then, by providing the continuity correction for $a \leqslant b$, with $a, b \in \mathbb{N}_0$

$$P(a \leqslant X \leqslant b) \approx \Phi\left(\frac{b + \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{a - \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right),$$

- Approximation of the cdf of X:

$$P(X \leqslant x) \approx \Phi\left(\frac{x + \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right)$$

Normal approximation of $\mathcal{P}(\lambda)$ distribution

- Let $X \sim \mathcal{P}(\lambda)$. Then, by providing the continuity correction for $a \leqslant b$, with $a, b \in \mathbb{N}_0$

$$P(a \leqslant X \leqslant b) \approx \Phi\left(\frac{b + \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{a - \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right),$$

- Approximation of the cdf of X:

$$P(X \leqslant x) \approx \Phi\left(\frac{x + \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right)$$

- Rule of thumb: The approximation is considered to be reasonable when $\lambda > 15$.

Example

- An accountant rounds to the nearest euro. We assume the error in rounding follows uniform distribution on $(-0.5, 0.5)$. Estimate the probability that the total error in 300 entries is more than 5 euro.

Example

- An accountant rounds to the nearest euro. We assume the error in rounding follows uniform distribution on $(-0.5, 0.5)$. Estimate the probability that the total error in 300 entries is more than 5 euro.

  Answer

Example

- An accountant rounds to the nearest euro. We assume the error in rounding follows uniform distribution on $(-0.5, 0.5)$. Estimate the probability that the total error in 300 entries is more than 5 euro.

  Answer

  - Let X_j be the rounding error of the jth entry. Then, $X_j \sim U(-0.5, 0.5)$

Example

- An accountant rounds to the nearest euro. We assume the error in rounding follows uniform distribution on $(-0.5, 0.5)$. Estimate the probability that the total error in 300 entries is more than 5 euro.

  Answer

  - Let X_j be the rounding error of the jth entry. Then, $X_j \sim U(-0.5, 0.5)$

    $E(X_j) = 0$ and $Var\left(X_j\right) = \frac{1}{12}$

Example

- An accountant rounds to the nearest euro. We assume the error in rounding follows uniform distribution on $(-0.5, 0.5)$. Estimate the probability that the total error in 300 entries is more than 5 euro.

  Answer

  - Let X_j be the rounding error of the jth entry. Then, $X_j \sim U(-0.5, 0.5)$
    $E(X_j) = 0$ and $Var\,(X_j) = \frac{1}{12}$
  - The overall error is $S = X_1 + \cdots + X_{300}$.

Example

- An accountant rounds to the nearest euro. We assume the error in rounding follows uniform distribution on $(-0.5, 0.5)$. Estimate the probability that the total error in 300 entries is more than 5 euro.

  Answer

  - Let X_j be the rounding error of the jth entry. Then, $X_j \sim U(-0.5, 0.5)$
    $E(X_j) = 0$ and $Var\,(X_j) = \frac{1}{12}$
  - The overall error is $S = X_1 + \cdots + X_{300}$.
    $E(S) = 0$, $Var\,(S) = \frac{300}{12} = 25$ and $\sigma_S = 5$.

Example

- An accountant rounds to the nearest euro. We assume the error in rounding follows uniform distribution on $(-0.5, 0.5)$. Estimate the probability that the total error in 300 entries is more than 5 euro.

  Answer

  - Let X_j be the rounding error of the jth entry. Then, $X_j \sim U(-0.5, 0.5)$
    $E(X_j) = 0$ and $Var\,(X_j) = \frac{1}{12}$
  - The overall error is $S = X_1 + \cdots + X_{300}$.
    $E(S) = 0$, $Var\,(S) = \frac{300}{12} = 25$ and $\sigma_S = 5$.
  - CLT: for the standardized S it holds

$$Z = \frac{S - E(S)}{\sqrt{Var(S)}} = \frac{S}{5} \sim \mathcal{N}(0, 1)$$

Example

- An accountant rounds to the nearest euro. We assume the error in rounding follows uniform distribution on $(-0.5, 0.5)$. Estimate the probability that the total error in 300 entries is more than 5 euro.

  Answer

  - Let X_j be the rounding error of the jth entry. Then, $X_j \sim U(-0.5, 0.5)$
    $E(X_j) = 0$ and $Var\,(X_j) = \frac{1}{12}$
  - The overall error is $S = X_1 + \cdots + X_{300}$.
    $E(S) = 0$, $Var\,(S) = \frac{300}{12} = 25$ and $\sigma_S = 5$.
  - CLT: for the standardized S it holds

$$Z = \frac{S - E(S)}{\sqrt{Var(S)}} = \frac{S}{5} \sim \mathcal{N}(0, 1)$$

  - Then,

$$P(|S| > 5) = 1 - P(|S| \leqslant 5) \approx 1 - P(|Z| \leqslant 1) \approx 0.32$$

Examples

HW Let $X_1, X_2, \ldots, X_{25}$ be independent and identically distributed (i.i.d.) random variables and $X_1 \sim \mathcal{N}(1, 4)$. Find the probability $P(X_1 + X_2 + \ldots\ldots X_{25} \geqslant 26)$.

HW Transportation officials tell us that 60% of the population wear their seatbelts while driving. A random sample of 1000 drivers has been taken. What is the probability that between 580 and 630 of the drivers were wearing their seatbelts?

HW Let $X_1, X_2, \ldots X_{121}$ be i.i.d. with the expectation $\mu = 35$ and variance $\sigma^2 = 25$. Denote by

$$\bar{X}_{121} = \frac{1}{121}(X_1 + \cdots + X_{121})$$

the sample mean. Approximate the probability $P(\bar{X}_{121} > 35.2)$ using the Central limit theorem.

HW We toss a fair coin 100 times.
What is the probability of obtaining 60 or more heads?

Questions

A few multiple-choice questions

(1) Let $X_1, X_2, \ldots X_{81}$ be i.i.d. sample from a population with population mean $\mu = 5$ and population variance $\sigma^2 = 4$ and let $S = X_1 + X_2 + \ldots X_{81}$. Approximate the probability $P(S \notin [387, 423])$ using the Central limit theorem.

a. 68%
b. 78%
c. 45%
d. 32%

(2) Assume that X is a binomial random variable with $n = 100$ and $p = 0.1$. Use the normal probability distribution to compute $P(X \leqslant 15)$.

a. 0.5336
b. 0.9664
c. 0.0336
d. 0.4664

A few multiple-choice questions

(3) Transportation officials tell us that 60% of the population wear their seatbelts while driving. A random sample of 1000 drivers has been taken. What is the probability that between 220 and 550 of the drivers were wearing their seatbelts?

a. 1.0
b. 0.4015
c. 0.9066
d. 0

Thank you for your attention!