Wirtschafts- und Sozialstatistik
Universität Trier

Elements of Statistics

Chapter 8:
Estimation

Ralf Münnich, Jan Pablo Burgard and Florian Ertz

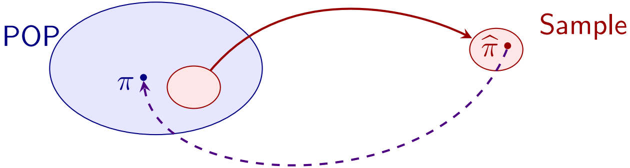
University of Trier
Faculty IV
Economic and Social Statistics Department

Winter term 2021/22

General idea

We are interested in population parameters which are generally unknown (here: π).

Before, we analysed populations using methods of descriptive statistics. Now, we draw a sample of the population and analyse this sample. The aim is to transfer results to the population ($\rightarrow$ point estimation).

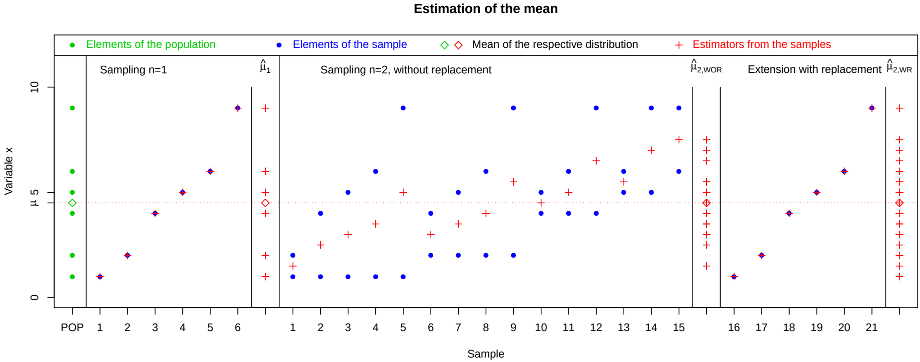


Additionally, we want to specify an interval of *plausible* values ($\rightarrow$ interval estimation).

Distribution of the sample mean

Wirtschafts- und Sozialstatistik
Universität Trier

We draw all possible samples of size $n = 1$ and $n = 2$, respectively, out of a population of $N = 6$ elements. We have:



Sample function and estimating function

Wirtschafts- und Sozialstatistik
Universität Trier

Specification of an estimating function

An estimating function for an unknown population parameter π is a sample function which qualifies to be used to estimate the parameter π by virtue of its properties. It is labelled $u_\pi(x_1, \ldots, x_n)$. The realisation of the estimating function is the estimate $\widehat{\pi} = u_\pi(x_1, \ldots, x_n)$.

Attention: We distinguish between the parameter to be estimated π, the estimate $\widehat{\pi}$ and the distribution of the latter or the corresponding random variable. The latter results when we substitute the sample variables $X_1, \ldots, X_n$ for the corresponding realisations $x_1, \ldots, x_n$. To be concrete, e. g. when estimating the mean of the population, we have μ and $\widehat{\mu} = \overline{x}$. We label the distribution of the estimator $U_\pi(X_1, \ldots, X_n)$ or U, and in this case $\overline{X}$. We may write $U(X_1, \ldots, X_n|\pi)$ as well.

Example 8.1: Four estimating functions (1)

Wirtschafts- und Sozialstatistik
Universität Trier

We want to estimate the mean μ of the population. With a sample size of n we have four estimating functions at our disposal:

$$\widehat{\mu}_1 = U_1(X_1, \ldots, X_n|\pi) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\widehat{\mu}_2 = U_2(X_1, \ldots, X_n|\pi) = \frac{1}{n+1} \cdot \left(2 \cdot X_1 + \sum_{i=2}^{n} X_i\right)$$

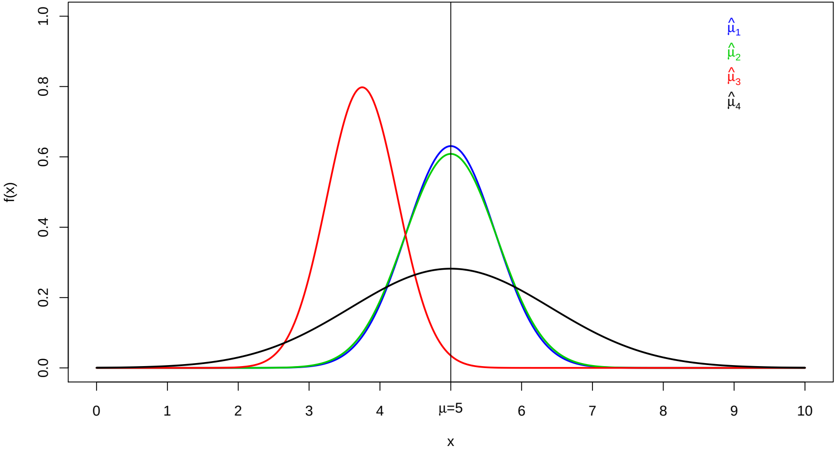
$$\widehat{\mu}_3 = U_3(X_1, \ldots, X_n|\pi) = \frac{1}{n+6} \cdot \left(2 \cdot X_1 + 2 \cdot X_n + \sum_{i=2}^{n-1} X_i\right)$$

$$\widehat{\mu}_4 = U_4(X_1, \ldots, X_n|\pi) = \frac{1}{2} \cdot (X_1 + X_n)$$

Example 8.1: Four estimating functions (2)

Wirtschafts- und Sozialstatistik
Universität Trier

Let the population be normally distributed with parameters $\mu = 5$ and $\sigma^2 = 4$. We draw a sample of size $n = 10$ with replacement.



Example 8.1: Four estimating functions (3)

Calculation of $\widehat{\mu}_1$, $\widehat{\mu}_2$, $\widehat{\mu}_3$ and $\widehat{\mu}_4$ in R:

```
mu <- 5
sigma <- sqrt(4)
n <- 10

Mean_U1 <- 1/n * (n * mu)
Mean_U2 <- 1/(n + 1) * (2 * mu + 9 * mu)
Mean_U3 <- 1/(n + 6) * (2 * mu + 2 * mu + 8 * mu)
Mean_U4 <- 1/2 * (2 * mu)

Means<-cbind(Mean_U1,Mean_U2,Mean_U3,Mean_U4)

Means
```

```
     Mean_U1   Mean_U2   Mean_U3   Mean_U4
[1,]       5         5      3.75         5
```

Example 8.1: Four estimating functions (4)

Calculation of $\widehat{\sigma}_1^2$, $\widehat{\sigma}_2^2$, $\widehat{\sigma}_3^2$ and $\widehat{\sigma}_4^2$ in R:

```
Var_U1 <- (1/n)^2 * (n * sigma^2)
Var_U2 <- (2/(n + 1))^2 * sigma^2 + (1/(n + 1))^2 *
          (9 * sigma^2)
Var_U3 <- (2/(n + 6))^2 * (2 * sigma^2) +
          (1/(n + 6))^2 * (8 * sigma^2)
Var_U4 <- (1/2)^2 * (2 * sigma^2)

Variances <- cbind(Var_U1,Var_U2,Var_U3,Var_U4)

Variances
```

```
     Var_U1    Var_U2    Var_U3    Var_U4
[1,]    0.4 0.4297521      0.25         2
```

Example 8.1: Four estimating functions (5)

Creation of the graphics in R:

```
x8_1 <- seq(from = 0, to = 10, length.out = 1000)

f_x8_1_1 <- dnorm(x = x8_1, mean = Mean_U1,
                  sd = sqrt(Var_U1))
f_x8_1_2 <- dnorm(x = x8_1, mean = Mean_U2,
                  sd = sqrt(Var_U2))
f_x8_1_3 <- dnorm(x = x8_1, mean = Mean_U3,
                  sd = sqrt(Var_U3))
f_x8_1_4 <- dnorm(x = x8_1, mean = Mean_U4,
                  sd = sqrt(Var_U4))

plot(x = x8_1, y = f_x8_1_1, type = "l", xlab = "x",
     ylab = "f(x)", ylim = c(0,1), lwd = 2, col = "blue")
lines(x=x8_1,y=f_x8_1_2,type="l",lwd=2,col="green")
lines(x=x8_1,y=f_x8_1_3,type="l",lwd=2,col="red")
lines(x=x8_1,y=f_x8_1_4,type="l",lwd=2,col="black")
abline(v = 5)
```

Properties of estimating functions (1)

Wirtschafts- und Sozialstatistik
Universität Trier

Definition 8.1 (Estimation error):

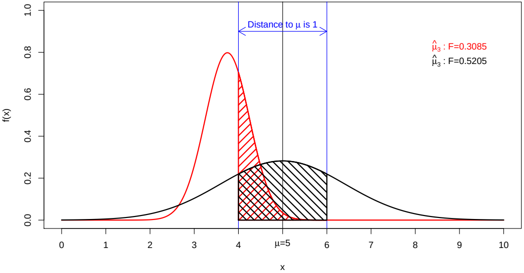
An estimation error e is the actual error resulting from an estimation:

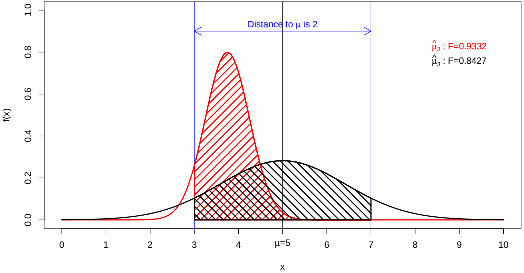
$$e = \widehat{\mu} - \mu \quad .$$

- Sampling is *random*. Therefore, the results of the different estimating functions will most likely lead to different evaluations for different samples.
- How can we compare estimating functions with regard to differing sample realisations?
- How should estimating functions behave for large samples ($n \to \infty$)?
- To what extent are such considerations useful in practice?

Wirtschafts- und Sozialstatistik
Universität Trier

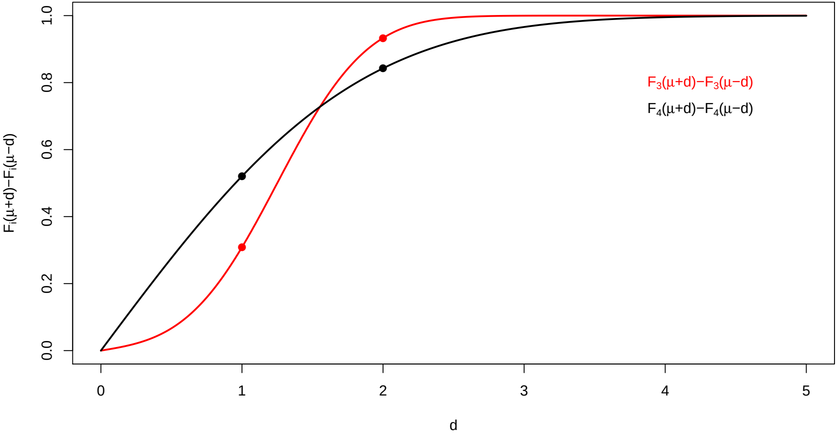
Example 8.2: see Ex. 8.1 (1)





Example 8.2: see Ex. 8.1 (2)

Probability for the interval $[\mu - d; \mu + d]$ of the distributions of $\widehat{\mu}_3$ and $\widehat{\mu}_4$:



Properties of estimating functions (2)

Wirtschafts- und Sozialstatistik
Universität Trier

Definition 8.2 (Unbiasedness):

An estimating function $U_\pi(X_1, \ldots, X_n)$ (short hand: U) is called unbiased for parameter π if we have

$$\mathrm{E}(U) = \pi\,.$$

The *average* estimate is equal to the parameter to be estimated π. Otherwise it is called biased. The extent of the bias may be quantified as follows:

$$\mathrm{Bias}(U) = \mathrm{E}(U) - \pi \quad .$$

We speak of asymptotical unbiasedness, if the following holds: $\lim\limits_{n\to\infty} E(U_n) = \pi$.

Example 8.3: see Ex. 8.1 (1)

Wirtschafts- und Sozialstatistik
Universität Trier

The estimating function $U = \sum_i \gamma_i X_i$ with $\sum_i \gamma_i = 1$ is unbiased because

$$\mathrm{E}\,U = \sum_{i=1}^{n} \gamma_i \cdot \underbrace{\mathrm{E}\,X_i}_{=\mu} = \mu \sum_{i=1}^{n} \gamma_i = \mu \quad .$$

Therefore, $\widehat{\mu}_1$, $\widehat{\mu}_2$ and $\widehat{\mu}_4$ are unbiased as their weights are $\gamma_i = 1/n$ for $\widehat{\mu}_1$, $\gamma_1 = 2/(n+1)$ and $\gamma_i = 1/(n+1)$ $(i > 1)$ for $\widehat{\mu}_2$ as well as $\gamma_1 = \gamma_n = 1/2$ and $\gamma_i = 0$ $(i \neq 1, n)$ for $\widehat{\mu}_4$.
For $\widehat{\mu}_3$ follows:

$$\mathrm{E}\,U_3 = \mathrm{E}\left(\frac{1}{n+6} \cdot \left(2 \cdot X_1 + 2 \cdot X_n + \sum_{i=2}^{n-1} X_i\right)\right)$$

$$= \frac{1}{n+6} \cdot \left(2 \cdot \mathrm{E}\,X_1 + 2 \cdot \mathrm{E}\,X_n + \sum_{i=2}^{n-1} \mathrm{E}\,X_i\right) = \frac{n+2}{n+6} \cdot \mu \quad .$$

Wirtschafts- und Sozialstatistik
Universität Trier

Example 8.3: see Ex. 8.1 (2)

Calculations for $\widehat{\mu}_3$ in R:

```
Mean_U3 <- (n + 2)/(n + 6) * mu
Mean_U3
```

```
[1] 3.75
```

U_3 is biased but asymptotically unbiased as $\lim\limits_{n\to\infty} \dfrac{n+2}{n+6} \cdot \mu = \mu$.

Calculation of the bias of $\widehat{\mu}_3$ in R:

```
Bias_U3 <- Mean_U3 - mu
Bias_U3
```

```
[1] -1.25
```

Calculation of the bias with $n = 10,000$ in R:

```
n_new <- 10000
Bias_U3_new <- (n_new + 2)/(n_new + 6) * mu - mu
round(Bias_U3_new, digits = 4)
```

```
[1] -0.002
```

Wirtschafts- und Sozialstatistik
Universität Trier

Example 8.4:

The estimating function $p = \widehat{\theta}$ is unbiased for the proportion θ of a certain type of interest in the population. This follows immediately from an application of the arithmetic mean in Example 8.3 on dichotomous variables.

Example 8.5:

The sample variance

$$S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \overline{X})^2$$

is unbiased for the population variance σ^2. Therefore,

$$S^{*\,2} = \frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \overline{X})^2$$

must be biased. Nevertheless, $S^{*\,2}$ is asymptotically unbiased.

Example 8.6: Two estimating functions

For the estimation of population parameter π we have two different unbiased estimating functions U_1 and U_2 at our disposal. We only know that $\text{Var}\, U_1 = 0.9 \cdot \text{Var}\, U_2$. Using Tchebysheff's inequality (theorem 7.2) we have:

$$P(|U_1 - \pi| \geq \varepsilon) \leq \frac{\text{Var}\, U_1}{\varepsilon^2} = 0.9 \cdot \frac{\text{Var}\, U_2}{\varepsilon^2}$$
$$P(|U_2 - \pi| \geq \varepsilon) \leq \frac{\text{Var}\, U_2}{\varepsilon^2} \quad .$$

The probability of *committing* an estimation error of at least ε is smaller for U_1 and depends on the variance of the estimating functions.
In case of biased estimating functions we may use the extended version of Tchebysheff's inequality (see Schaich and Münnich, 2001, p. 21):

$$P(|U_2 - \pi| \geq \varepsilon) \leq \frac{\text{Var}\, U_2 + (\text{E}\, U_2 - \pi)^2}{\varepsilon^2} = \frac{\overbrace{\text{Var}\, U_2 + \text{Bias}^2(U_2)}^{:=\text{MSE}(U_2)}}{\varepsilon^2} \quad .$$

Properties of estimating functions (3)

Wirtschafts- und Sozialstatistik
Universität Trier

Definition 8.3 (Efficiency):

An unbiased estimating function U is called efficient (best) estimating function for parameter π if there is no other unbiased estimating function U' for π with $\text{Var}\,(U') \leq \text{Var}\,(U)$.

Out of a number of unbiased estimating functions we choose the one with the smallest variance.

In practice, it's far from easy to find the best estimating function. With the aid of sufficient estimating functions (estimating functions that use all information of a sample about the parameter that one wants to estimate) and the Rao-Blackwell theorem, one can construct *better* estimating functions (see lecture *Elements of Statistics and Econometrics* in the masters program *M.Sc. Applied Statistics*).

Example 8.7: Arithmetic mean

Wirtschafts- und Sozialstatistik
Universität Trier

Out of the linear unbiased estimating functions, the arithmetic mean is the best estimating function for μ. Using the Lagrange multiplier we get:

$$\frac{\partial \left[\operatorname{Var}\left(\sum_{i=1}^{n} \gamma_i X_i \right) + \lambda \left(1 - \sum_{i=1}^{n} \gamma_i \right) \right]}{\partial \gamma_i} =$$

$$\frac{\partial}{\partial \gamma_i} \left[\sum_{i=1}^{n} \gamma_i^2 \operatorname{Var} X_i + \lambda \left(1 - \sum_{i=1}^{n} \gamma_i \right) \right] =$$

$$\frac{\partial}{\partial \gamma_i} \left[\sigma^2 \cdot \sum_{i=1}^{n} \gamma_i^2 + \lambda \left(1 - \sum_{i=1}^{n} \gamma_i \right) \right] =$$

$$\sigma^2 \cdot 2 \cdot \gamma_i - \lambda \stackrel{!}{=} 0 \quad .$$

Finally, after equating we get $\gamma_i = \gamma_j$ for all $i, j = 1, \ldots, n$ and therefore the proposition.

We say that the arithmetic mean is the *best linear unbiased estimator* (BLUE) for μ.

Properties of estimating functions (4)

Wirtschafts- und Sozialstatistik
Universität Trier

Definition 8.4 (Consistency):

An estimating function $U(X_1, \ldots, X_n|\pi)$ is called consistent for the estimation of the population parameter π if

$$\lim_{n\to\infty} P(|U_n - \pi| > \varepsilon) = 0$$

for any arbitrarily small $\varepsilon > 0$.

We say that U_n converges stochastically to the parameter to be estimated π.

Example 8.8: Consistency of $\overline{X}$

Wirtschafts- und Sozialstatistik
Universität Trier

$\overline{x}_n$ is the arithmetic mean for sample size n (with replacement). Using Tchebysheff's inequality and $\operatorname{Var}\overline{X}_n = \operatorname{Var}X/n$ ($\overline{X}_n$ is unbiased) we get

$$\begin{aligned} P(|\overline{X}_n - \operatorname{E}\overline{X}_n| > \varepsilon) &\leq P(|\overline{X}_n - \mu| \geq \varepsilon) \\ &\leq \frac{\operatorname{Var}\overline{X}_n}{\varepsilon^2} = \frac{\operatorname{Var}X}{n \cdot \varepsilon^2} \end{aligned}$$

for every $\varepsilon > 0$. Finally, we then have

$$0 \leq \lim_{n\to\infty} P(|\overline{X}_n - \operatorname{E}\overline{X}_n| > \varepsilon) \leq \lim_{n\to\infty} \frac{\operatorname{Var}X}{n \cdot \varepsilon^2} = 0 \quad .$$

$\overline{X}_n$ is consistent.

Methods to gain estimating functions

Wirtschafts- und Sozialstatistik
Universität Trier

- **Ordinary least squares (OLS)**:
  The sum of the squared errors is minimised. Examples are the OLS regression (see Chapter 4) or $\widehat{\mu}_{\mathrm{KQ}}$:
  $\sum_i (x_i - \widehat{\mu}_{\mathrm{KQ}})^2 \rightarrow \min$ leads to $\widehat{\mu}_{\mathrm{KQ}} = \overline{x}$.
- **Method of moments**:
  The empirical moments $\frac{1}{n}\sum_i x_i^k$ are made equal to the theoretical moments $\mathrm{E}\left(X^k\right)$. From this, one obtains the estimates. With $\widehat{\mu} = \overline{x}$ and $\widehat{\mu}^2 + \widehat{\sigma}^2 = \frac{1}{n}\sum_i x_i^2$ for $k = 1, 2$, one would finally $\widehat{\sigma}^2 = s^{*\,2}$ with unknown μ.
- **Maximum Likelihood method (ML)**
- **Bayesian estimation**

Maximum Likelihood method

Wirtschafts- und Sozialstatistik
Universität Trier

Given the n stochastically independent realisations of a random sample, the explicit parameters of a known distribution have to be estimated. From the set of all possible estimates, those estimates are selected which have the highest probability or probability density given the available sample result. Hence:

$$
\begin{aligned}
L(x_1, \ldots, x_n | \widehat{\pi}_1, \ldots, \widehat{\pi}_r) &= \max_{\pi_1, \ldots, \pi_r} L(x_1, \ldots, x_n | \pi_1, \ldots, \pi_r) \\
&= \max_{\pi_1, \ldots, \pi_r} \prod_{i=1}^{n} f(x_i | \pi_1, \ldots, \pi_r) \quad .
\end{aligned}
$$

In most cases, the log likelihood function $\ln L$ is maximized instead of the likelihood function L, whereby a sum instead of a product is maximized.

Properties of the Maximum Likelihood method

Wirtschafts- und Sozialstatistik
Universität Trier

- Given there is an efficient estimate for a parameter π, the ML method yields it
- ML estimation functions are consistent, but generally not unbiased
- ML estimators are asymptotically normal distributed for $n \to \infty$
- If U is an ML estimation function for π, then $\tau(U)$ is also an ML estimation function for a wide class of functions τ

Example 8.9: One urn, two colours (1)

Wirtschafts- und Sozialstatistik
Universität Trier

An urn contains $N = 50$ balls. Those balls are either black or yellow but the respective proportions θ are unknown. A sample of size $n = 10$ (WR) yields four black balls. We are looking for the θ which maximizes $b(4|10;\theta)$. Because of $N = 50$, θ can only be a multiple of 0.02. Resulting in:

θ	0.34	0.36	0.38	0.40	0.42	0.44	0.46
$b(4\|10;\theta)$	0.2320	0.2424	0.2487	0.2508	0.2488	0.2427	0.2331

Creation of the table in `R`:

```
x8_9 <- 4
n <- 10
theta <- seq(from = 0.34, to = 0.46, by = 0.02)
theta
```

```
[1] 0.34 0.36 0.38 0.40 0.42 0.44 0.46
```

```
f_x8_9 <- dbinom(x = x8_9, size = n, prob = theta)
round(f_x8_9, digits = 4)
```

```
[1] 0.2320 0.2424 0.2487 0.2508 0.2488 0.2427 0.2331
```

Example 8.9: One urn, two colours (2)

Wirtschafts- und Sozialstatistik
Universität Trier

Thus, $\widehat{\theta} = 0.4$ is used as the ML estimate in this case.

Determination of $\widehat{\theta} = 0{,}4$ in R:

```
theta_hat <- theta[which.max(f_x8_9)]
theta_hat
```

```
[1] 0.4
```

Example 8.10: ML estimation of θ - Binomial distribution (1)

In a sample of size n, the outcome 1 results $n \cdot p$ times whereas the outcome 0 results $n \cdot (1-p)$ times. Thus, the likelihood function is given by:

$$L(x_1, \ldots, x_n | \theta) = \prod_{i=1}^{n} \theta^{x_i} \cdot (1-\theta)^{1-x_i} = \theta^{np} \cdot (1-\theta)^{n(1-p)} \quad .$$

Taking the logarithm results in:

$$\ln L(x_1, \ldots, x_n | \theta) = np \ln \theta + n(1-p) \ln(1-\theta)$$

finally, differentiation yields

$$\frac{\partial \ln L(x_1, \ldots, x_n | \theta)}{\partial \theta} = \frac{np}{\theta} - \frac{n(1-p)}{1-\theta} \stackrel{!}{=} 0 \quad .$$

Thus the necessary criterion for a maximum finally results in $\widehat{\theta} = p$. Sufficient criterion still has to be checked!

Wirtschafts- und Sozialstatistik
Universität Trier

Example 8.10: ML estimation of θ - Binomial distribution (2)

Log likelihood and partial derivative in R:

```
Log_Likelihood <- expression(n * p * log(Theta) +
                             n * (1 - p) * log(1 - Theta))

D_Log_Likelihood <- D(expr = Log_Likelihood, name = "Theta")

Log_Likelihood
```

```
expression(n * p * log(Theta) + n * (1 - p) * log(1 - Theta))
```

```
D_Log_Likelihood
```

```
n * p * (1/Theta) - n * (1 - p) * (1/(1 - Theta))
```

Example 8.11: ML estimation of μ and σ^2 of a normal distribution I

The following is valid:

$$
\begin{aligned}
L(x_1, \ldots, x_n | \mu; \sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2\right) \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \cdot \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right)
\end{aligned}
$$

or:

$$
\ln L(x_1, \ldots, x_n | \mu; \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 \quad .
$$

Wirtschafts- und Sozialstatistik
Universität Trier

Example 8.11: ML estimation of μ and σ^2 of a normal distribution II

Finally, partial derivation with respect to the parameters μ and σ^2

$$\frac{\ln L(x_1, \ldots, x_n|\mu;\sigma^2)}{\partial\mu} = \frac{1}{2\sigma^2} \cdot 2 \cdot \sum_{i=1}^{n}(x_i - \mu) \stackrel{!}{=} 0 \qquad \text{and}$$

$$\frac{\ln L(x_1, \ldots, x_n|\mu;\sigma^2)}{\partial\sigma^2} = -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n}(x_i - \mu)^2 \stackrel{!}{=} 0$$

yields the estimators $\widehat{\mu} = \overline{x}$ and $\widehat{\sigma}^2 = s^{*\,2}$.

Bayesian estimation (see Fahrmeir et al., 2016)

- $x_1, \ldots, x_n$ are n independent realisations of a random variable X which follows a distribution F with parameter θ
- θ is a realisation of a random variable Θ
- $f(x, \theta)$ is the joint density; $f(x|\theta)$ is the conditional and $f(x)$ the boundary distribution of X
- $f(\theta)$ is the a-priori distribution of the parameter Θ
- $f(\theta|x)$ is the a-posteriori distribution of Θ

Bayesian inference

Let $f(x|\theta)$ be the density of X given θ and $L(\theta) = f(x_1, \ldots, x_n|\theta)$ constitutes the corresponding likelihood function. Then, the a-priori density $f(\theta)$ can be used to derive the a-posteriori density of θ

$$f(\theta|x_1, \ldots, x_n) = \frac{f(x_1|\theta) \ldots f(x_n|\theta) \cdot f(\theta)}{\int f(x_1|\theta) \ldots f(x_n|\theta) \cdot f(\theta)\, d\theta} = \frac{L(\theta) f(\theta)}{\int L(\theta) f(\theta)\, d\theta}$$

(discrete distributions and multidimensional Θ are also possible).

Bayesian estimator und Bayesian learning

Wirtschafts- und Sozialstatistik
Universität Trier

A-posteriori expected value

$$\widehat{\theta}_{\mathrm{E}} = \mathrm{E}\,(\theta|x_1,\ldots,x_n) = \int \theta f(\theta|x_1,\ldots,x_n)\,d\theta$$

Maximum a-posteriori estimator (MAP)

$$\widehat{\theta}_{\mathrm{MAP}} = \arg\max_{\theta} L(\theta)f(\theta) \quad \text{or} \quad \widehat{\theta}_{\mathrm{MAP}} = \arg\max_{\theta} \big(\ln L(\theta) + \ln f(\theta)\big)$$

The calculation of the a-posteriori density of θ is often no longer analytically feasible $\rightarrow$ numerical or Monte Carlo integration or MCMC.

If the a-priori distribution of Θ is *very flat* (non-informative prior), then one obtains the Maximum Likelihood estimation. Otherwise, the subjective conceptions of the a-priori distribution is used in the estimation.

Example 8.12: see Fahrmeir et al., 2016

Let $x_1, \ldots, x_n$ be independent realisations from $X \sim N(\mu, \sigma^2)$ with known σ^2. We want to estimate the parameter μ. We use $N(\mu_0, \sigma_0^2)$ as a-priori density for the parameter we want to estimate. σ_0^2 controls the precision of the a-priori information.

With some effort, we can show that the a-posteriori distribution of μ is

$$\mu|x_1, \ldots, x_n \sim N\left(\frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2}\overline{x} + \frac{\sigma^2}{n\sigma_0^2 + \sigma^2}\mu_0; \frac{\sigma^2}{n + \sigma^2/\sigma_0^2}\right)$$

The *trust parameter* σ_0^2 controls the evaluation of the sample information. If σ_0^2 is very large ($\to \infty$), then we obtain the classical MLE. If on the other hand σ_0^2 is very small, then the a-priori information changes little with the sample information.

General idea of interval estimation

Wirtschafts- und Sozialstatistik
Universität Trier

Besides the point estimate derived from the sample, we need some kind of *quality criterion* for this point estimate. Some options:

- Variance of the estimator
  (requires an approximate normal distribution)
- Standard error (standard deviation of estimator)
- Coefficient of variation of estimate

The problem of each of those options is that missing information regarding the population forces us to *estimate* their respective values using the sample.

Another option is to state a certain *range of variation* around the point estimate. We would like to state an interval based on quantiles of an estimator's distribution, like $[x_{0.025}; x_{0.975}]$.

Example 8.13: Random interval (1)

Wirtschafts- und Sozialstatistik
Universität Trier

Let the random variable X be normally distributed with known variance $\sigma^2 = 900$. To estimate the population mean μ we draw a sample of size $n = 36$ with replacement.

We use the estimator $\overline{X}$. We use

- $\overline{X}_l = \overline{X} + z(0.025) \cdot \frac{\sigma}{\sqrt{n}} = \overline{X} - 1.96 \cdot 5 = \overline{X} - 9.8$
- $\overline{X}_u = \overline{X} + z(0.975) \cdot \frac{\sigma}{\sqrt{n}} = \overline{X} + 1.96 \cdot 5 = \overline{X} + 9.8$

as the limits of the interval motivated above. We get the following random interval $[\overline{X}_l, \overline{X}_u] = [\overline{X} - 9.8; \overline{X} + 9.8]$.

What is the probability that the parameter to be estimated μ lies within the limits of this random interval?

Example 8.13: Random interval (2)

Wirtschafts- und Sozialstatistik
Universität Trier

$$
\begin{aligned}
P(\overline{X}_l \leq \mu \leq \overline{X}_u) &= P(\overline{X} - 9.8 \leq \mu \leq \overline{X} + 9.8) \\
&= P(-9.8 \leq \overline{X} - \mu \leq 9.8) \\
&= P\Bigg(-1.96 \leq \underbrace{\frac{\overline{X} - \mu}{5}}_{\sim \text{SND}} \leq 1.96 \Bigg) \\
&= 0.975 - 0.025 = 0.95
\end{aligned}
$$

A random interval constructed in this fashion *covers* the true parameter μ with a probability of 95%.

Data input of relevant parameters in `R`:

```
alpha <- 0.05
sigma <- 30
n <- 36
```

Example 8.13: Random interval (3)

Attention:
Such a statement may only be given in terms of probabilities and therefore only **before** an experiment is carried out. As soon as a concrete interval $[\overline{x}_l, \overline{x}_u]$ is determined, we can only state if the true parameter is covered by the interval or not. But in reality this information will not be available in most cases.

Confidence intervals (1)

- As we assume that the probability of the interval $[\overline{X}_l, \overline{X}_u]$ covering the true parameter μ is 0.95 before the experiment is carried out,
  - we have a respective level of *confidence* that
  - the true parameter actually lies within the limits of the confidence interval after the experiment has been carried out.
- Therefore, the interval $[\overline{X}_l, \overline{X}_u]$ is called 95% confidence interval for μ.
- Generally, depending on the question at hand, we use values of 0.95, 0.99 or 0.90.

Confidence intervals (2)

Wirtschafts- und Sozialstatistik
Universität Trier

Definition 8.5 (Confidence interval):

Let the confidence level $(1-\alpha)$ be given. The interval $[\pi_l, \pi_u]$ with $\pi_l = f(X_1, \ldots, X_n)$ and $\pi_u = f(X_1, \ldots, X_n)$ $(\pi_l \leq \pi_u)$ is called $(1-\alpha)$ confidence interval for π, if we have $P(\pi_l \leq \pi \leq \pi_u) = 1-\alpha$.

Questions about the properties of such a confidence interval, like its symmetry or its minimal length, immediately arise.

CI for μ, POP is normally distributed, σ^2 is known

Wirtschafts- und Sozialstatistik
Universität Trier

The random variable

$$Z = \frac{\overline{X} - \mu}{\sigma} \cdot \sqrt{n}$$

follows the standard normal distribution. The resulting $(1-\alpha)$ confidence interval is

$$\left[\overline{X} - z(1-\alpha/2) \cdot \frac{\sigma}{\sqrt{n}}; \overline{X} + z(1-\alpha/2) \cdot \frac{\sigma}{\sqrt{n}}\right].$$

- This $(1-\alpha)$ confidence interval is as short as possible and is symmetric to $\overline{X}$.
- The larger σ, the longer the CI
- The larger n, the shorter the CI
- The larger $(1-\alpha)$, the longer the CI

Wirtschafts- und Sozialstatistik
Universität Trier

Example 8.14: see Ex. 8.13

The evaluation of the sample yielded $\overline{x} = 72$. Therefore, the 95% confidence interval is

$$\left[72 - 1.96 \cdot \frac{30}{\sqrt{36}}; 72 + 1.96 \cdot \frac{30}{\sqrt{36}}\right]$$

and finally

$$\left[62.2; 81.8\right] \quad .$$

95% confidence interval in R:

```
SpMean <- 72
CI <- vector()
CI[1] <- SpMean - qnorm(p = 1 - (alpha/2))*(sigma/sqrt(n))
CI[2] <- SpMean + qnorm(p = 1 - (alpha/2))*(sigma/sqrt(n))

CI_lower_alternative <- SpMean + qnorm(p = alpha/2) *
                              (sigma/sqrt(n))
round(CI, digits = 1)
```

```
[1] 62.2 81.8
```

Wirtschafts- und Sozialstatistik
Universität Trier

CI for μ, POP is normally distributed, σ^2 is unknown

The random variable

$$T = \frac{\overline{X} - \mu}{S} \cdot \sqrt{n}$$

follows the t distribution with $n - 1$ degrees of freedom. The resulting $(1 - \alpha)$ confidence interval is

$$\left[\overline{X} - t(1 - \frac{\alpha}{2}, n - 1) \cdot \sqrt{\frac{S^2}{n}}; \overline{X} + t(1 - \frac{\alpha}{2}, n - 1) \cdot \sqrt{\frac{S^2}{n}}\right].$$

- We have $\dfrac{n-1}{\sigma^2} \cdot S^2 = \dfrac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \overline{X})^2 \sim \chi^2_{n-1}$.
- Cochran's theorem holds and therefore $\dfrac{1}{\sigma^2} S^2 (n-1)$ and $\dfrac{1}{\sigma} \cdot (\overline{X} - \mu) \cdot \sqrt{n}$ are stochastically independent.
- $\dfrac{1}{\sigma} \cdot (\overline{X} - \mu) \cdot \sqrt{n} \Big/ \sqrt{\left(\dfrac{1}{\sigma^2} S^2 (n-1)\right) / (n-1)} = \dfrac{\overline{X} - \mu}{S} \cdot \sqrt{n}$

Example 8.15: Unknown variance (1)

Wirtschafts- und Sozialstatistik
Universität Trier

Now, let σ^2 be unknown. As an estimate of σ^2 we use $s^2 = 33^2$ which is derived from the sample. We get the 95% confidence interval

$$\left[72 - 2.0315 \cdot \frac{33}{\sqrt{36}}; 72 + 2.0315 \cdot \frac{33}{\sqrt{36}}\right]$$

and finally

$$\left[72 - 11.173; 72 + 11.173\right] = \left[60.8268; 83.1733\right] \quad .$$

Attention:
$t(0.975; 35)$ is not tabulated. We used the arithmetic mean of the tabulated values $t(0.975; 30)$ and $t(0.975; 40)$ as the normal approximation would still yield inexact values (small n).
Thanks to R, this is not a problem anymore (see next slide).

Example 8.15: Unknown variance (2)

Wirtschafts- und Sozialstatistik
Universität Trier

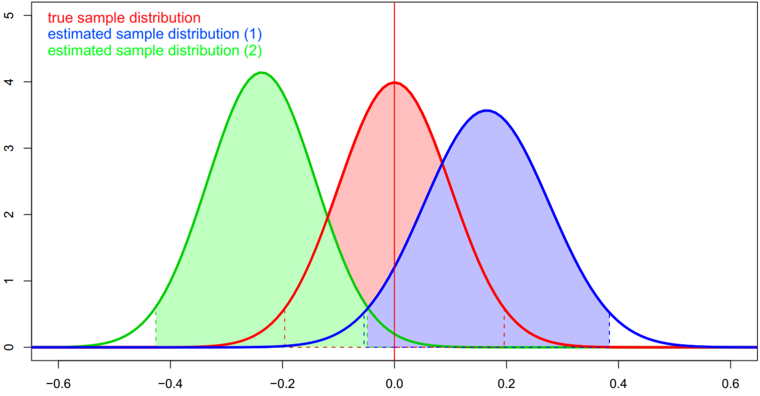
95% Confidence interval in R:

```
alpha <- 0.05
SpMean <- 72
SpVar <- 33^2
n <- 36

CI <- vector()
CI[1] <- SpMean - qt(p = 1 - (alpha/2), df = n - 1) *
                   sqrt(SpVar/n)
CI[2] <- SpMean + qt(p = 1 - (alpha/2), df = n - 1) *
                   sqrt(SpVar/n)

round(CI, digits = 1)
```

```
[1] 60.8 83.2
```

Example 8.16: Sample distributions (1)

Let the population be normally distributed with unknown variance σ^2. A sample of size $n = 10$ is drawn. We can compare the true but unknown sample distribution as well as two estimated distributions resulting from two different samples.



Example 8.16: Sample distributions (2)

Wirtschafts- und Sozialstatistik
Universität Trier

Confidence interval simulations (see next slide)

Point vs. variance estimates (upper left)

$\rightarrow$ Cochran's theorem (given normal distribution)

Estimated distributions (lower left)

True distribution and estimated distributions of $\overline{X}$

Confidence intervals (upper right)

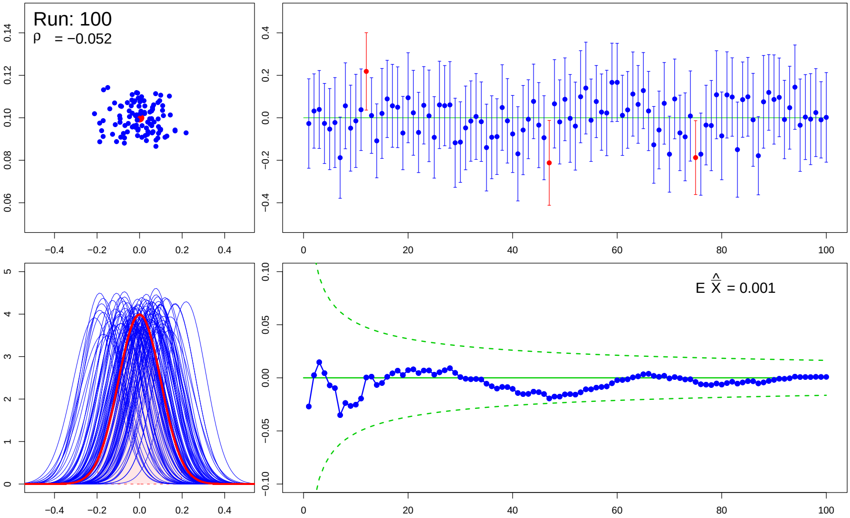
For $R = 100$ simulation runs;
red intervals do not cover true value

Convergence of $\mathrm{E}\left(\overline{X}\right)$ (lower right)

For r-th simulation run;
check of law of large numbers

Simulation of standard normal distribution

Wirtschafts- und Sozialstatistik
Universität Trier



CI for σ^2, POP is normally distributed

The random variable

$$\frac{n-1}{\sigma^2} \cdot S^2$$

follows a χ^2 distribution with $n-1$ degrees of freedom. We get the $(1-\alpha)$ confidence interval

$$\left[\frac{(n-1)S^2}{\chi^2(1-\frac{\alpha}{2}; n-1)}; \frac{(n-1)S^2}{\chi^2(\frac{\alpha}{2}; n-1)}\right]$$

- The CI follows from a rearrangement of
  $\chi^2(\frac{\alpha}{2}; n-1) \leq \dfrac{n-1}{\sigma^2} \cdot S^2 \leq \chi^2(1-\frac{\alpha}{2}; n-1)$.
- The CI does not have a minimal length. For very large n the normal approximation ensures the property of symmetry and minimal length.

Example 8.17: CI for variance (1)

Wirtschafts- und Sozialstatistik
Universität Trier

Let a population be normally distributed. A sample of size $n = 25$ yields $s^2 = 7.244$. We search the 90% confidence interval for σ^2.

We have $\chi^2(0.05; 24) = 13.848$ and $\chi^2(0.95; 24) = 36.415$. Therefore, we get the 90% confidence interval

$$\left[\frac{24 \cdot 7.244}{36.415}; \frac{24 \cdot 7.244}{13.848}\right]$$

and finally

$$\left[4.774; 12.555\right].$$

Example 8.17: CI for variance (2)

90% confidence interval in R:

```
alpha <- 0.1
n <- 25
SpVar <- 7.244

CI <- vector()

CI[1] <- ((n - 1) * SpVar) /
         qchisq(p = 1 - alpha/2, df = n-1)

CI[2] <- ((n - 1) * SpVar) /
         qchisq(p = alpha/2, df = n-1)

round(CI, digits = 3)
```

```
[1]  4.774 12.554
```

CI for $\operatorname{E} X$, arbitrary distribution, $\operatorname{Var} X$ known

The random variable

$$Z = \frac{\overline{X} - \operatorname{E} X}{\sqrt{\operatorname{Var} X}} \cdot \sqrt{n}$$

does approximatively follow a standard normal distribution. For $n > 30$, following the central limit theorem of Lindeberg and Lévy, the $(1-\alpha)$ confidence interval is

$$\left[\overline{X} - z(1-\alpha/2) \cdot \sqrt{\frac{\operatorname{Var} X}{n}} ; \overline{X} + z(1-\alpha/2) \cdot \sqrt{\frac{\operatorname{Var} X}{n}}\right].$$

Example 8.18: see Ex. 8.13

Wirtschafts- und Sozialstatistik
Universität Trier

Now, let a normal distribution of the population be questionable. As the sample size is $n = 36$, we again have the 95% confidence interval $\left[62.2; 81.8\right]$, but now it is not exact but approximative.

95% confidence interval in R:

```
n <- 36
alpha <- 0.05
VarX <- 30^2

CI_new <- vector()
CI_new[1] <- SpMean - qnorm(p=1-alpha/2) * sqrt(VarX/n)
CI_new[2] <- SpMean + qnorm(p=1-alpha/2) * sqrt(VarX/n)

round(CI_new, digits = 1)
```

```
[1] 62.2 81.8
```

CI for $\mathrm{E}\,X$, arbitrary distribution, $\mathrm{Var}\,X$ unknown

Wirtschafts- und Sozialstatistik
Universität Trier

The random variable

$$Z = \frac{\overline{X} - \mathrm{E}\,X}{\sqrt{S^2}} \cdot \sqrt{n}$$

does approximately follow a standard normal distribution. For $n > 30$, following the central limit theorem of Lindeberg and Lévy, the $(1 - \alpha)$ confidence interval is

$$\left[\overline{X} - z(1-\alpha/2) \cdot \sqrt{\frac{S^2}{n}}; \overline{X} + z(1-\alpha/2) \cdot \sqrt{\frac{S^2}{n}}\right].$$

Example 8.19: see Ex. 8.15 (1)

Analoguously to Example 8.18 as an approximative 95% confidence interval we have

$$\left[72 - 1.96 \cdot \frac{33}{\sqrt{36}}; 72 + 1.96 \cdot \frac{33}{\sqrt{36}}\right]$$

and therefore

$$\left[72 - 10.78; 72 + 10.78\right] = \left[61.22; 82.78\right] \quad .$$

Check of approximation conditions in R:

```
SpVar <- 33^2
n > 30
[1] TRUE
```

95% confidence interval in R:

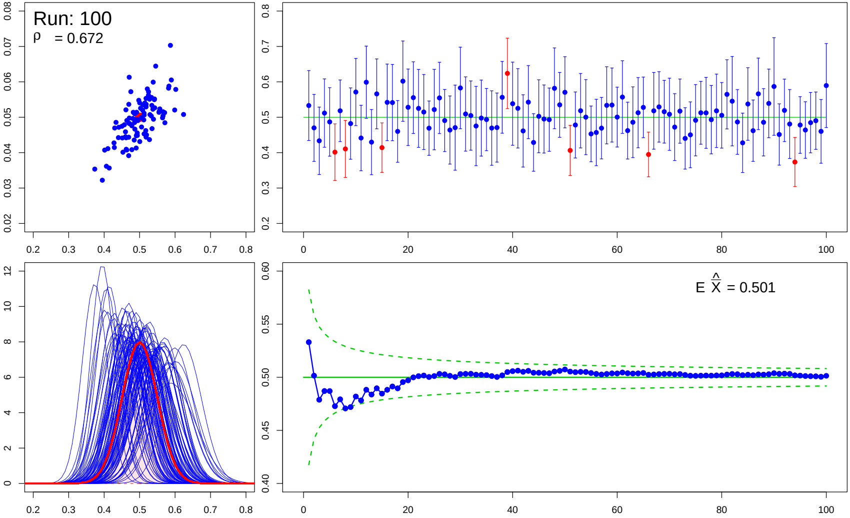
```
CI_new <- vector()
CI_new[1] <- SpMean - qnorm(p = 1 - alpha/2) *
                      sqrt(SpVar/n)
CI_new[2] <- SpMean + qnorm(p = 1 - alpha/2) *
                      sqrt(SpVar/n)
round(CI_new, digits = 2)
[1] 61.22 82.78
```

Example 8.19: see Ex. 8.15 (2)

This approximative CI is shorter than the respective CI using the t distribution: $\left[60.8268; 83.1733\right]$. Notice the problems which may arise when using approximations.

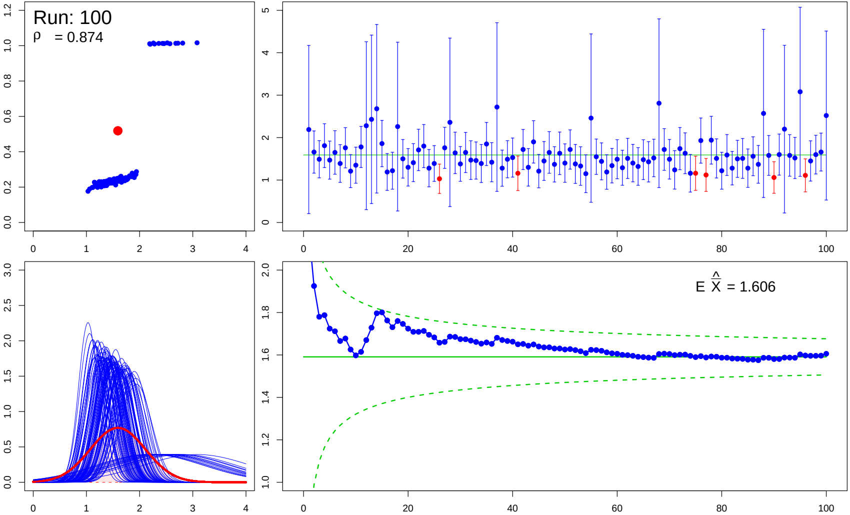
The following examples illustrate this effect, e.g. that approximations may not always be used without concern.

Simulation using the exponential distribution ($\lambda = 2$)



Simulation using discrete distribution with outlier

Wirtschafts- und Sozialstatistik
Universität Trier



Wirtschafts- und Sozialstatistik
Universität Trier

CI for $\operatorname{E} X$, arbitrary distribution, $\operatorname{Var} X$ unknown, without replacement

The random variable

$$Z = \frac{\overline{X} - \operatorname{E} X}{\sqrt{\frac{S^2}{n} \cdot \frac{N-n}{N-1}}}$$

does approximately follow a standard normal distribution. For $n > 30$, following the central limit theorem of Lindeberg and Lévy, the $(1-\alpha)$ confidence interval is

$$\left[\overline{X} - z(1-\alpha/2) \cdot \sqrt{\frac{S^2}{n} \cdot \frac{N-n}{N-1}}; \overline{X} + z(1-\alpha/2) \cdot \sqrt{\frac{S^2}{n} \cdot \frac{N-n}{N-1}}\right]$$

- In case $\operatorname{Var} X$ is known, we substitute $\operatorname{Var} X$ for S^2.
- Mind the approximation conditions: n large and n not close to N

CI for proportions, variance unknown

Instead of $\overline{X}$ we use the sample proportion P. We estimate the population variance using $P \cdot (1-P)$. As the estimator distribution, using de Moivre and Laplace's theorem, the standard normal distribution is used. Mind the approximation conditions. We forego the continuity correction. The $(1-\alpha)$ confidence interval is

$$\left[P - z\left(1-\frac{\alpha}{2}\right) \cdot \sqrt{\frac{P(1-P)}{n}}; P + z\left(1-\frac{\alpha}{2}\right) \cdot \sqrt{\frac{P(1-P)}{n}}\right].$$

Example 8.20: CI for proportions (1)

A survey of $n = 100$ students yielded a number of 15 students having a job. We get the 99% confidence interval

$$\left[0.15 - 2.575 \cdot \sqrt{\frac{0.15 \cdot 0.85}{100}}; 0.15 + 2.575 \cdot \sqrt{\frac{0.15 \cdot 0.85}{100}}\right] = \left[0.058; 0.242\right].$$

Data input in R:

```
alpha <- 0.01
n <- 100
p <- 15/100
```

Check of approximation conditions in R:

```
n * p * (1 - p) > 9
```

```
[1] TRUE
```

```
0.1 <= p & p <= 0.9
```

```
[1] TRUE
```

Example 8.20: CI for proportions (2)

Wirtschafts- und Sozialstatistik
Universität Trier

99% confidence interval in R:

```
CI <- vector()
CI[1] <- p - qnorm(p = 1 - alpha/2)*sqrt((p * (1 - p))/n)
CI[2] <- p + qnorm(p = 1 - alpha/2)*sqrt((p * (1 - p))/n)

round(CI, digits = 3)
```

```
[1] 0.058 0.242
```

Example 8.21: see Ex. 8.14 (1)

Wirtschafts- und Sozialstatistik
Universität Trier

Determination of needed sample size

We search the sample size for which the 95% CI is at most 5 units long.

We have

$$\left[72 - 1.96 \cdot \frac{30}{\sqrt{n}}; 72 + 1.96 \cdot \frac{30}{\sqrt{n}}\right] .$$

This yields a length of $d = 2 \cdot 1.96 \cdot 30/\sqrt{n}$. Using

$$2 \cdot 1.96 \cdot \frac{30}{\sqrt{n}} \leq 5$$

we finally get

$$n \geq \left(2 \cdot 1.96 \cdot \frac{30}{5}\right)^2 = 553.1904 \quad .$$

We need a sample size of $n \geq 554$.

Example 8.21: see Ex. 8.14 (2)

Wirtschafts- und Sozialstatistik
Universität Trier

Calculation of n in R:

```
alpha <- 0.05
Quantile <- qnorm(p = 1 - alpha/2)
sigma <- 30
d <- 5

n_min <- ceiling((2 * Quantile * sigma/d)^2)

n_min
```

```
[1] 554
```