Wirtschafts- und Sozialstatistik
Universität Trier

Elements of Statistics

Chapter 7:
Selected distributions

Ralf Münnich, Jan Pablo Burgard and Florian Ertz

University of Trier
Faculty IV
Economic and Social Statistics Department

Winter term 2021/22

The Bernoulli distribution

Wirtschafts- und Sozialstatistik
Universität Trier

A discrete random variable X is said to be bernoulli-distributed with parameter p $(0 < p < 1)$ if its probability or distribution function, respectively, satisfies:

$$be(x|p) = \begin{cases} 1-p & \text{for } x = 0 \\ p & \text{for } x = 1 \\ 0 & \text{else} \end{cases} \quad \text{or} \quad Be(x|p) = \begin{cases} 0 & \text{for } x < 0 \\ 1-p & \text{for } 0 \leq x < 1 \\ 1 & \text{for } x \geq 1 \end{cases}$$

Hence, it is valid:

$$\begin{aligned} \operatorname{E} X &= 0 \cdot (1-p) + 1 \cdot p = p \\ \operatorname{Var} X &= 0^2 \cdot (1-p) + 1^2 \cdot p - p^2 = p \cdot (1-p)\,. \end{aligned}$$

Access on the Bernoulli distribution in `R`:

```
dbinom(x = c(1,0), size = 1, prob = p)
```


The Binomial distribution (1)

A discrete random variable X is said to be distributed binomially with parameters n and θ ($n \in \mathbb{N}, 0 < \theta < 1$) if its probability function satisfies:

$$b(x|n,\theta) = \begin{cases} \binom{n}{x} \cdot \theta^x \cdot (1-\theta)^{n-x} & \text{for } x = 0, 1, 2, \ldots, n \\ 0 & \text{else} \end{cases}$$

or its distribution function complies with

$$B(x|n,\theta) = \begin{cases} 0 & x < 0 \\ \sum\limits_{\nu=0}^{[x]} \binom{n}{\nu} \cdot \theta^\nu \cdot (1-\theta)^{n-\nu} & \text{for } x \geq 0\,. \end{cases}$$

n balls are drawn with replacement (WR). The probability that x balls of the type of interest are sampled is given by $b(x|n,\theta)$ where θ is the probability associated with each ball.

The Binomial distribution (2)

Since the draws of x balls of the type of interest and $n-x$ balls of the remaining type are independent of each other, we say:
1. There are $\binom{n}{x}$ possible combinations to draw the x or rather $n-x$ balls, given the n draws.
2. Each and every combination has a probability of $W(x|n,\theta) = \theta^x \cdot (1-\theta)^{n-x}$.

A binomially distributed random variable is the sum of n bernoulli distributed random variables. Since the experiment is performed with replacement, the draws are stochastically independent $(X = X_1 + \ldots + X_n)$. Hence:

$$\mathrm{E}\,X = \sum_{i=1}^{n} \mathrm{E}\,X_i = \sum_{i=1}^{n} \theta = n \cdot \theta$$

and

$$\mathrm{Var}\,X = \sum_{i=1}^{n} \mathrm{Var}\,X_i = \sum_{i=1}^{n} \theta \cdot (1-\theta) = n \cdot \theta \cdot (1-\theta) \quad .$$

Example 7.1: see ex. 6.4 (1)

Let $\theta = \frac{3}{10} = 0.3$. We then have for $n = 3$

$$b(0|3;0.3) = \binom{3}{0} \cdot 0.3^0 \cdot 0.7^3 = 0.343$$

$$b(1|3;0.3) = \binom{3}{1} \cdot 0.3^1 \cdot 0.7^2 = 0.441$$

$$b(2|3;0.3) = \binom{3}{2} \cdot 0.3^2 \cdot 0.7^1 = 0.189$$

$$b(3|3;0.3) = \binom{3}{3} \cdot 0.3^3 \cdot 0.7^0 = 0.027$$

Those are the very same probabilities as reported in the previous chapter.

Example 7.1: see ex. 6.4 (2)

The Binomial distribution in R:

```
x7_1 <- 0:3
Theta7_1 <- 0.3 ; n <- 3
f_x7_1 <- dbinom(x7_1, size = n, prob = Theta7_1)
names(f_x7_1) <- x7_1
f_x7_1
```

```
    0     1     2     3
0.343 0.441 0.189 0.027
```

Furthermore, we get:

$$EX = 3 \cdot 0.3 = 0.9$$

EX in R:

```
Mean_X <- weighted.mean(x = x7_1, w = f_x7_1)
Mean_X
```

```
[1] 0.9
```

Alternatively:

```
Mean_X <- n * Theta7_1
```

Wirtschafts- und Sozialstatistik
Universität Trier

Example 7.1: see ex. 6.4 (3)

… as well as

$$\mathrm{Var}X = 3 \cdot 0.3 \cdot 0.7 = 0.63.$$

Var$X$ in R:

```
Var <- sum(f_x7_1 * (x7_1 - Mean_X)^2)
Var
```

```
[1] 0.63
```

Alternatively:

```
Var <- n * Theta7_1 * (1 - Theta7_1)
```

The multinomial distribution

A discrete random variable X is said to follow a multinomial distribution with parameters n and $\theta_1, ..., \theta_k$ ($n \in \mathbb{N}, 0 < \theta < 1$ and $\theta_1 + ... + \theta_k = 1$) if its probability function complies to:

$$m(x_1, \ldots, x_k | n, \theta_1, \ldots, \theta_k) = \begin{cases} \dfrac{n!}{x_1! \cdot \ldots \cdot x_k!} \cdot \theta_1^{x_1} \cdot \ldots \cdot \theta_k^{x_k} & \text{for } x_i \in \mathbb{N}_0 \text{ and} \\ & x_1 + \ldots + x_k = n \\ 0 & \text{else}\,. \end{cases}$$

In practice, the parameter n of the function m is mostly neglected due to the relation $x_1 + \ldots + x_k = n$.
The binomial distribution is a special case of the multinomial distribution with probabilities $(\theta, 1 - \theta)$ associated with the two possible outcomes $(x, n - x)$.

Example 7.2: see Ex. 6.16

Let now $\theta_1 = 0.3$, $\theta_2 = 0.2$ and $\theta_3 = 0.5$ (balls' colours). Then

$$f(1,0) = m(\underbrace{1,0,2}_{=3} \,|\, \underbrace{0.3;0.2;0.5}_{=1}) = \frac{3!}{1!\cdot 0!\cdot 2!}\cdot 0.3^1 \cdot 0.2^0 \cdot 0.5^2 = 0.225\,.$$

$f(1,0)$ (Probability for one red and no white ball) in `R`:

```
x7_2 <- c(1, 0, 2); n <- 3
Theta7_2 <- c(0.3, 0.2, 0.5)
f_1_0_2 <- dmultinom(x7_2, size = n, prob = Theta7_2)
f_1_0_2
```

```
[1] 0.225
```

The probability $f(2,2) = m(2,2,0|0.3,0.2,0.5)$ can be derived from the *else* node of the probability function of the multinomial distribution because of $2+2+0 > n$.

The hypergeometric distribution

Wirtschafts- und Sozialstatistik
Universität Trier

A discrete random variable X is said to follow a hypergeometric distribution with parameters n, M and N ($n, M, N \in \mathbb{N}$ with $n \leq N$ and $M < N$) if its probability function complies to:

$$h(x|n, N, M) = \begin{cases} \dfrac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}} & \text{for } x = 0, \ldots, n \text{ as well as} \\ & \max(0, M + n - N) \leq x \leq \min(n, M) \\ 0 & \text{else} \end{cases}$$

M equals the number of balls of the type of interest sampled from an urn containing N balls in total ($\theta = \dfrac{M}{N}$). Furthermore, n balls are drawn without replacement (WOR). Thus:

$$\operatorname{E} X = n \cdot \frac{M}{N} \quad \text{and} \quad \operatorname{Var} X = n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \cdot \frac{N-n}{N-1}\,.$$

Wirtschafts- und Sozialstatistik
Universität Trier

Example 7.3: Another urn example (1)

Given an urn containing $N = 100$ balls of which $M = 30$ are white. $n = 3$ balls are drawn from the urn **without replacement**. The resulting probabilities are:

$$h(0|3;0.3) = \frac{\binom{30}{0} \cdot \binom{70}{3}}{\binom{100}{3}} = 0.3385 \qquad b(0|3;0.3) = 0.343$$

$$h(1|3;0.3) = \frac{\binom{30}{1} \cdot \binom{70}{2}}{\binom{100}{3}} = 0.4481 \qquad b(1|3;0.3) = 0.441$$

$$h(2|3;0.3) = \frac{\binom{30}{2} \cdot \binom{70}{1}}{\binom{100}{3}} = 0.1883 \qquad b(2|3;0.3) = 0.189$$

$$h(3|3;0.3) = \frac{\binom{30}{3} \cdot \binom{70}{0}}{\binom{100}{3}} = 0.0251 \qquad b(3|3;0.3) = 0.027$$

Example 7.3: Another urn example (2)

The hypergeometric distribution in R

```
x7_3 <- 0:3
n <- 3
N7_3 <- 100
M7_3 <- 30
Theta7_3 <- M7_3 / N7_3

f_x7_3 <- dhyper(x7_3, m = M7_3, n = N7_3 - M7_3, k = n)
names(f_x7_3) <- x7_3
round(f_x7_3, digits = 3)
```

```
    0     1     2     3
0.339 0.448 0.188 0.025
```

The Poisson distribution

Wirtschafts- und Sozialstatistik
Universität Trier

A discrete random variable X is said to be poisson distributed with parameter λ ($\lambda > 0$) if its probability function satisfies:

$$po(x|\lambda) = \begin{cases} \dfrac{e^{-\lambda} \cdot \lambda^x}{x!} & \text{for } x = 0, 1, \dots \\ 0 & \text{else} \end{cases}$$

We have:

$$\mathrm{E}\, X = \lambda \qquad \text{and} \qquad \mathrm{Var}\, X = \lambda \quad .$$

The poisson distribution is used to model rare events, e.g. natural disasters. In addition to that, it is the limit distribution of the binomial distribution.

Example 7.4: Earthquake (1)

Wirtschafts- und Sozialstatistik
Universität Trier

Within a certain region an earthquake occurs more or less every ten years. We are interested in the probability that this specific region experiences three earthquakes within one single year.

The information 'once in every ten years' translates to $\lambda = 0.1$, hence it follows:

$$\begin{aligned}
P(X \geq 3) &= 1 - W(X \leq 2) = 1 - Po(2|0.1) \\
&= 1 - \left(\frac{0.1^0}{0!} \cdot e^{-0.1} + \frac{0.1^1}{1!} \cdot e^{-0.1} + \frac{0.1^2}{2!} \cdot e^{-0.1} \right) \\
&= 1 - \left(\frac{0.1^0}{0!} + \frac{0.1^1}{1!} + \frac{0.1^2}{2!} \right) \cdot e^{-0.1} \\
&= 1 - 1.105 \cdot e^{-0.1} \\
&\approx 0.000155\,.
\end{aligned}$$

Wirtschafts- und Sozialstatistik
Universität Trier

Example 7.4: Earthquake (2)

Calculation of $W(X \geq 3)$ in R:

```
x7_4 <- 0:5
Lambda7_4 <- 0.1
f_x7_4 <- dpois(x = x7_4, lambda = Lambda7_4)
names(f_x7_4) <- x7_4
F_x7_4 <- ppois(q = x7_4, lambda = Lambda7_4)
names(F_x7_4) <- x7_4
round(f_x7_4, digits = 6)
```

```
       0        1        2        3        4        5
0.904837 0.090484 0.004524 0.000151 0.000004 0.000000
```

```
F_x7_4
```

```
        0         1         2         3         4         5
0.9048374 0.9953212 0.9998453 0.9999962 0.9999999 1.0000000
```

```
Prob <- 1 - F_x7_4["2"]
Prob_old <- sum(f_x7_4[4:6])
```

```
round(Prob, digits=6)
```

```
[1] 0.000155
```

```
Prob_old
```

```
[1] 0.000155
```

Wirtschafts- und Sozialstatistik
Universität Trier

Interrelation between poisson and binomial distribution

We have:

$$\lim_{\substack{n\to\infty\\ \theta\to 0}} b(x|n,\theta) = po(x|\lambda) \quad ,$$

where $\lambda = n \cdot \theta$. Thus the expected values are identical and the variances are approximatively identical..

Example 7.5:

Suppose $\lambda = 2 = n \cdot \theta$:

Distribution	$X=0$	$X=1$	$X=2$	$X=3$	$X=4$	$X=5$	$X=6$	Var X
$b(x\|10;0.2)$	0.107	0.268	0.302	0.201	0.088	0.026	0.006	1.6
$b(x\|50;0.04)$	0.130	0.271	0.276	0.184	0.090	0.035	0.011	1.92
$b(x\|100;0.02)$	0.133	0.271	0.273	0.182	0.090	0.035	0.011	1.96
$b(x\|200;0.01)$	0.134	0.271	0.272	0.181	0.090	0.036	0.012	1.98
$b(x\|500;0.004)$	0.135	0.271	0.271	0.181	0.090	0.036	0.012	1.992
$b(x\|1000;0.002)$	0.135	0.271	0.271	0.181	0.090	0.036	0.012	1.996
$p(x\|2)$	0.135	0.271	0.271	0.180	0.090	0.036	0.012	2

Interrelation between binomial and hypergeometric distribution

The distributions differ due to drawing with/without replacement. For smaller samples (n/N) we thus have negligible differences.

Example 7.6:

For $n = 6$ and $M/N = 0.2$ we get:

Distribution	$X = 0$	$X = 1$	$X = 2$	$X = 3$	$X = 4$	$X = 5$	$X = 6$	$\operatorname{Var} X$
$h(x\|6, 25, 5)$	0.219	0.438	0.274	0.064	0.005	0.000	0.000	0.76
$h(x\|6, 50, 10)$	0.242	0.414	0.259	0.075	0.010	0.001	0.000	0.862
$h(x\|6, 100, 20)$	0.252	0.403	0.252	0.079	0.013	0.001	0.000	0.912
$h(x\|6, 200, 40)$	0.257	0.398	0.249	0.080	0.014	0.001	0.000	0.936
$h(x\|6, 500, 100)$	0.260	0.395	0.247	0.081	0.015	0.001	0.000	0.950
$h(x\|6, 100, 200)$	0.261	0.394	0.246	0.082	0.015	0.001	0.000	0.955
$b(x\|6, 0.2)$	0.262	0.393	0.246	0.082	0.015	0.002	0.000	0.96

An approximation of the hypergeometric distribution with the Poisson distribution can be done in two steps using the binomial distribution.

The exponential distribution

Wirtschafts- und Sozialstatistik
Universität Trier

A continuous random variable X is said to be exponentially distributed with parameter λ ($\lambda > 0$) if its density function satisfies:

$$f(x|\lambda) = \begin{cases} \lambda \cdot \exp\big(-\lambda \cdot x\big) & \text{for } x > 0 \\ 0 & \text{else} \end{cases}$$

We have:

$$\mathrm{E}\, X = \frac{1}{\lambda} \qquad \text{and} \qquad \mathrm{Var}\, X = \frac{1}{\lambda^2} \qquad .$$

The exponential distribution is used to model lifespans, e.g. the one of light bulbs. It is a distribution *without* memory (see Schira 2005, p. 365.)

Access on the exponential distribution in `R`:

```
dexp(x, rate = Lambda)
```

The normal distribution

Wirtschafts- und Sozialstatistik
Universität Trier

A continuous random variable X is called normally distributed with parameters μ and σ^2 ($\mu \in \mathbb{R}$ and $\sigma > 0$), if it exhibits the following density function:

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

We use the short hand $N(\mu, \sigma^2)$.

We have:

$$\mathrm{E}\, X = \mu \qquad \text{and} \qquad \mathrm{Var}\, X = \sigma^2 \qquad .$$

The expected value and the variance are the explicit parameters of the normal distribution.

Example 7.7: Normal distributions (1)

In the following, we will illustrate the normal distributions $N(0, 1)$, $N(5, 1)$, $N(-3, 0.35^2)$ and $N(5, 2^2)$ graphically.

Density function of the normal distribution in R:

```
x7_7 <- seq(from = -5, to = 10, by = 0.00001)

f_x7_7_1 <- dnorm(x = x7_7, mean = 0, sd = 1)
f_x7_7_2 <- dnorm(x = x7_7, mean = 5, sd = 1)
f_x7_7_3 <- dnorm(x = x7_7, mean = -3, sd = 0.35)
f_x7_7_4 <- dnorm(x = x7_7, mean = 5, sd = 2)
```

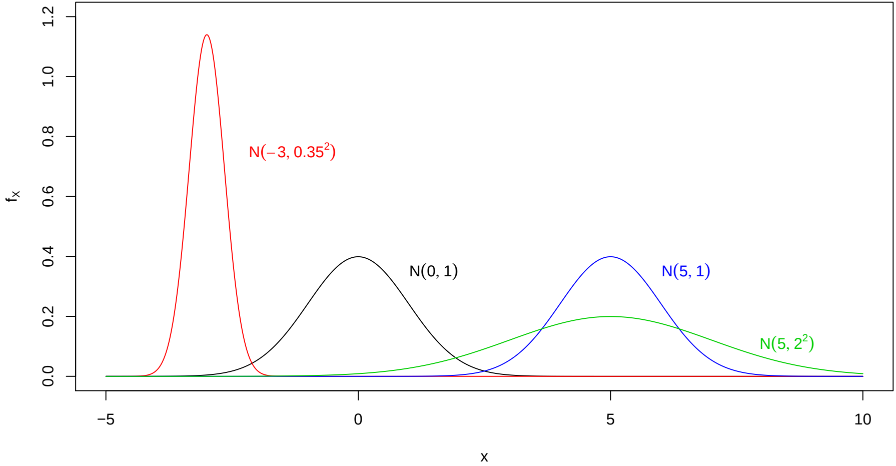
Illustration in R:

```
plot(x = x7_7, y = f_x7_7_1, type = "l", xlab = "x",
     ylab = "f(x)", ylim = c(0, 1.2))
lines(x = x7_7, y = f_x7_7_2, type = "l", col = "blue")
lines(x = x7_7, y = f_x7_7_3, type = "l", col = "red")
lines(x = x7_7, y = f_x7_7_4, type = "l", col = "green")
```

Example 7.7: Normal distributions (2)

Wirtschafts- und Sozialstatistik
Universität Trier

Normal distributions $N(0,1)$, $N(5,1)$, $N(-3,0.35^2)$ and $N(5,2^2)$



Properties of the normal distribution

Wirtschafts- und Sozialstatistik
Universität Trier

1. The normal distribution is symmetric to the axis $x = \mu$.

2. The graph of the normal distribution reaches its maximum at the point $\left(\mu, \frac{1}{\sqrt{2\pi}\sigma}\right)$.

3. The inflexion points of the normal distribution are at $x = \mu - \sigma$ and $x = \mu + \sigma$.

4. The distribution function of the normal distribution

$$\Phi(x|\mu, \sigma^2) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{(s-\mu)^2}{2\sigma^2}\right) ds$$

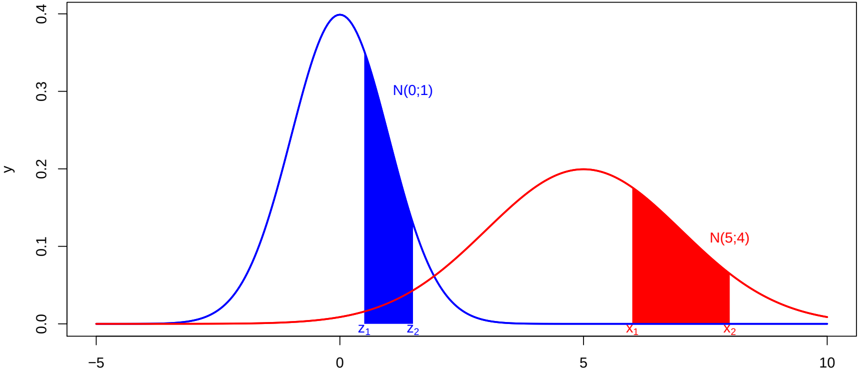
cannot be given in a closed-form expression.

Wirtschafts- und Sozialstatistik
Universität Trier

Standard normal distribution and standard transformation

If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$, where $N(0,1)$ is the standard normal distribution.

Particularly $P(x_1 < X \leq x_2) = P(z_1 < Z \leq z_2) = \Phi(z_2) - \Phi(z_1)$, with $z_1 = \frac{x_1-\mu}{\sigma}$ and $z_2 = \frac{x_2-\mu}{\sigma}$.



Example 7.8: Standard normal distribution (1)

Wirtschafts- und Sozialstatistik
Universität Trier

Using tabulated values of the standard normal distribution or R, we reach:

$$\Phi(1) = 0.8413 \qquad \Phi(1.96) = 0.9750 \qquad \Phi(3.29) = 0.9995$$
$$\Phi(-1) = 1 - \Phi(1) = 0.1587 \qquad \Phi(-1.96) = 0.0250 \qquad \Phi(-3.29) = 0.0005$$

Calculation of the values of the distribution function in R

```
F_x7_8 <- pnorm(q = c(1,1.96,3.29,-1,-1.96,-3.29),
                mean = 0, sd = 1)
round(F_x7_8, digits = 4)
```

```
[1] 0.8413 0.9750 0.9995 0.1587 0.0250 0.0005
```

Example 7.8: Standard normal distribution (2)

From this, we obtain the following probabilities:

$$
\begin{aligned}
P(1 \leq Z \leq 2) &= \Phi(2) - \Phi(1) \\
&= 0.9772 - 0.8413 = 0.1359 \\
P(-2 \leq Z \leq -1) &= \Phi(-1) - \Phi(-2) = 1 - \Phi(1) - (1 - \Phi(2)) \\
&= 0.1359 \\
P(-2 \leq Z \leq 1) &= \Phi(1) - (1 - \Phi(2)) = \Phi(1) + \Phi(2) - 1 \\
&= 0.8185 \\
P(Z \geq -1) &= 1 - \Phi(-1) = 1 - (1 - \Phi(1)) = \Phi(1) \\
&= 0.8413
\end{aligned}
$$

Calculation of $P(1 \leq Z \leq 2)$ in R:

```
Prob7_8 <- pnorm(q=2,mean=0,sd=1) - pnorm(q=1,mean=0,sd=1)
round(Prob7_8, digits=4)
```

```
[1] 0.1359
```

Example 7.9: Simplifications (1)

Wirtschafts- und Sozialstatistik
Universität Trier

Let a random variable X be $N(20; 64)$ distributed.

$$\Phi(28|20;64) = \Phi\left(\frac{28-20}{8}\right) = \Phi(1) = 0.8413$$

$$\Phi(12|20;64) = \Phi\left(\frac{12-20}{8}\right) = \Phi(-1) = 0.1587$$

$$\Phi(35.68|20;64) = \Phi\left(\frac{35.68-20}{8}\right) = \Phi(1.96) = 0.9750$$

$$\Phi(4.32|20;64) = \Phi(-1.96) = 0.0250$$

$$\Phi(46.32|20;64) = \Phi(3.29) = 0.9995$$

$$\Phi(-6.32|20;64) = \Phi(-3.29) = 0.0005$$

Wirtschafts- und Sozialstatistik
Universität Trier

Example 7.9: Simplifications (2)

Calculations in R:

```
Phi7_9 <- pnorm(q = c(28,12,35.68,4.32,46.32,-6.32),
                mean = 20, sd = sqrt(64))
round(Phi7_9, digits=4)
```

```
[1] 0.8413 0.1587 0.9750 0.0250 0.9995 0.0005
```

Alternative calculation with the standard normal distribution in R:

```
Phi7_9 <- pnorm(q = c(1,-1,1.96,-1.96,3.29,-3.29),
                mean = 0, sd = 1)
```

$$P(28 \leq X \leq 36) = 0.1359 \qquad P(4 \leq X \leq 12) = 0.1359$$
$$P(4 \leq X \leq 28) = 0.8185 \qquad P(X \geq 12) = 0.8413$$

Calculation of $P(28 \leq X \leq 36)$ in R:

```
Prob7_9 <- pnorm(36,mean=20,sd=8) - pnorm(28,mean=20,sd=8)
round(Prob7_9, digits=4)
```

```
[1] 0.1359
```

Example 7.10: Quantiles of the normal distribution (1)

Wirtschafts- und Sozialstatistik
Universität Trier

Find the p-quantiles of $N(0,1)$ and $N(6,2^2)$ for $p = 0.8; 0.9; 0.95; 0.15; 0.01$.

As $x(p) = \mu + \sigma \cdot z(p)$ we have:

$$
\begin{aligned}
z(0.8) &= 0.84 & x(0.8) &= 6 + 2 \cdot 0.84 = 7.68 \\
z(0.9) &= 1.28 & x(0.9) &= 6 + 2 \cdot 1.28 = 8.56 \\
z(0.95) &= 1.645 & x(0.95) &= 6 + 2 \cdot 1.645 = 9.29 \\
z(0.15) &= -z(0.85) = -1.04 & x(0.15) &= 6 + 2 \cdot (-1.04) = 3.92 \\
z(0.01) &= -z(0.99) = -2.33 & x(0.01) &= 6 + 2 \cdot (-2.33) = 1.34
\end{aligned}
$$

Example 7.10: Quantiles of the normal distribution (2)

Determination of the quantiles in R:

```
Quantile_z7_10 <- qnorm(p = c(0.8,0.9,0.95,0.15,0.01),
                        mean = 0, sd = 1)
Quantile_x7_10 <- qnorm(p = c(0.8,0.9,0.95,0.15,0.01),
                        mean = 6, sd= 2)
round(Quantile_z7_10, digits=2)
```

```
[1]  0.84  1.28  1.64 -1.04 -2.33
```

```
round(Quantile_x7_10, digits=2)
```

```
[1] 7.68 8.56 9.29 3.93 1.35
```

Alternative calculation for $N(6, 2^2)$:

```
Quantile_x7_10_alternative <- 6 + 2 * Quantile_z7_10
round(Quantile_x7_10_alternative, digits=2)
```

```
[1] 7.68 8.56 9.29 3.93 1.35
```

Wirtschafts- und Sozialstatistik
Universität Trier

Linear functions of normally distributed random variables

Let the random variables $X_1, \ldots, X_n$ all be $N(\mu_i, \sigma_i^2)$ $(i = 1, \ldots, n)$ distributed and overall stochastically independent. Then the linear combination

$$Y = a_0 + a_1 \cdot X_1 + \ldots + a_n \cdot X_n$$

is normally distributed with

$$N\left(a_0 + \sum_{i=1}^{n} a_i \cdot \mu_i \, ; \quad \sum_{i=1}^{n} a_i^2 \cdot \sigma_i^2\right) \qquad .$$

Particularly, the arithmetic mean of identically distributed variables $\left(X_i \sim N(\mu; \sigma^2)\right)$ is

$$N\left(\mu; \frac{\sigma^2}{n}\right)$$

distributed.

Example 7.11: Linear combination

Wirtschafts- und Sozialstatistik
Universität Trier

Let the random variables X_1, X_2, X_3 be $N(2;1^2)$, $N(5;2^2)$ and $N(8;3^2)$ distributed and stochastically independent. Furthermore, let

$$Y = 30 + 5X_1 + 2X_2 - 5X_3 \quad .$$

We get

$$\mathrm{E}\, Y = 30 + 5 \cdot 2 + 2 \cdot 5 - 5 \cdot 8 = 10$$

and

$$\mathrm{Var}\, Y = 5^2 \cdot 1^2 + 2^2 \cdot 2^2 + (-5)^2 \cdot 3^2 = 266 \quad .$$

Therefore, Y is $N(10;266)$ distributed.

Notice the difference:
An interviewer questions 10 people or he questions one person and replicates the result 10 times.

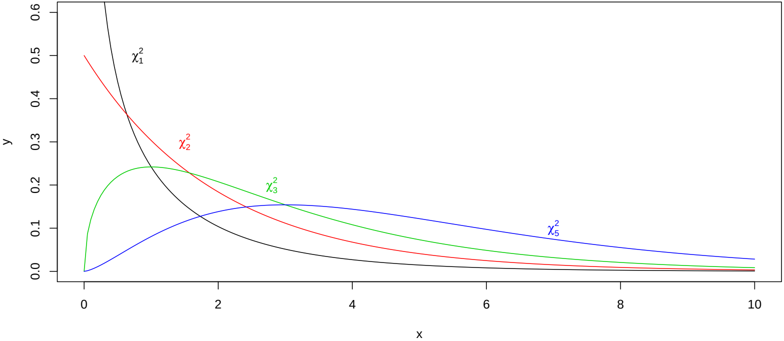
The χ^2 distribution

The sum of k squared, stochastically independent random variables Z_i $(i = 1, \ldots, k)$ following a standard normal distribution

$$Y = Z_1^2 + Z_2^2 + \ldots + Z_k^2$$

is χ^2-distributed with k degrees of freedom.

We have $\mathrm{E}\,Y = k$ and $\mathrm{Var}\,Y = 2 \cdot k$. In R: `dchisq(x, df = k)`



Example 7.12: Probabilities with the χ^2 distribution

Let the random variable Y be χ^2-distributed with $k = 10$ degrees of freedom. We have:

$$P(Y \leq 3.94) = F(3.94|10) = 0.05$$

```
round(pchisq(q = 3.94, df = 10), digits = 2)
```

```
[1] 0.05
```

$$P(Y \geq 15.987) = 1 - F(15.987|10) = 0.1$$

```
round(1 - pchisq(q = 15.987, df = 10), digits = 2)
```

```
[1] 0.1
```

$$P(4.87 \leq Y \leq 20.48) = F(20.48|10) - F(4.87|10) = 0.975 - 0.1 = 0.875$$

```
round(pchisq(q = 20.48,df = 10)- pchisq(q = 4.87, df = 10),
    digits = 3)
```

```
[1] 0.875
```

Approximation using the normal distribution

Wirtschafts- und Sozialstatistik
Universität Trier

1. If $Y \sim \chi^2_k$, then

$$\sqrt{2 \cdot Y} - \sqrt{2 \cdot k - 1} \sim N(0,1) \qquad \text{for } k > 30$$

2. $$\chi^2_k \sim N(k; 2 \cdot k) \qquad \text{for } k > 100$$

Example 7.13: Approximation (1)

Wirtschafts- und Sozialstatistik
Universität Trier

1. Let $Y \sim \chi^2_{40}$. Then we have

$$P(Y \leq 55.758) = F(55.758|40) = 0.95 \qquad \text{or}$$

$$F(55.758|40) = \Phi(\sqrt{2 \cdot 55.758} - \sqrt{2 \cdot 40 - 1}) = \Phi(1.67) = 0.9525\,.$$

2. Let $Y \sim \chi^2_{200}$. Then we have

$$P(Y \leq 220) = F(220|200) = 0.8417$$

Calculation of $P(Y \leq 220)$ in R:

```
Prob7_13 <- pchisq(q = 220, df = 200)
round(Prob7_13, digits = 4)
```

```
[1] 0.8417
```

Example 7.13: Approximation (2)

Wirtschafts- und Sozialstatistik
Universität Trier

Alternative 1:

$$F(220|200) = \Phi(\sqrt{2 \cdot 220} - \sqrt{2 \cdot 200 - 1}) = \Phi(1{,}00) = 0.8416$$

```
Prob7_13_1 <- pnorm(q = sqrt(2*220)-sqrt(2*200-1),
                      mean = 0, sd = 1)
round(Prob7_13_1, digits = 4)
```

```
[1] 0.8416
```

Alternative 2:

$$F(220|200) = \Phi(220|200;400) = \Phi\left(\frac{220-200}{\sqrt{400}}\right) = \Phi(1{,}00) = 0.8413$$

```
Prob7_13_2 <- pnorm((220-200)/sqrt(400), mean = 0, sd = 1)
round(Prob7_13_2, digits=4)
```

```
[1] 0.8413
```

The t distribution (1)

Let Z follow a standard normal distribution and let Y be χ^2_k-distributed. Furthermore, let Z and Y be stochastically independent. Then, the random variable

$$T = \frac{Z}{\sqrt{Y/k}}$$

follows a t-distribution with k degrees of freedom.

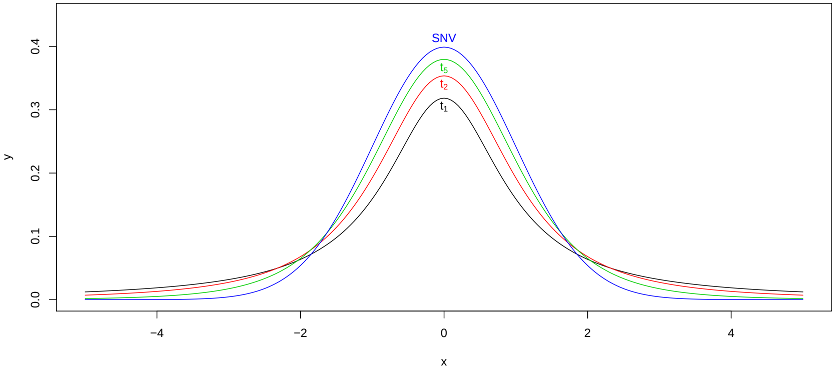
We have $\mathrm{E}\, T = 0$ $(k > 1)$ and $\mathrm{Var}\, T = \frac{k}{k-2}$ $(k > 2)$ as well as $t(1-p,k) = -t(p,k)$.

Access on the t distribution in R:

```
dt(x, df = k)
```

Wirtschafts- und Sozialstatistik
Universität Trier

The t distribution (2)



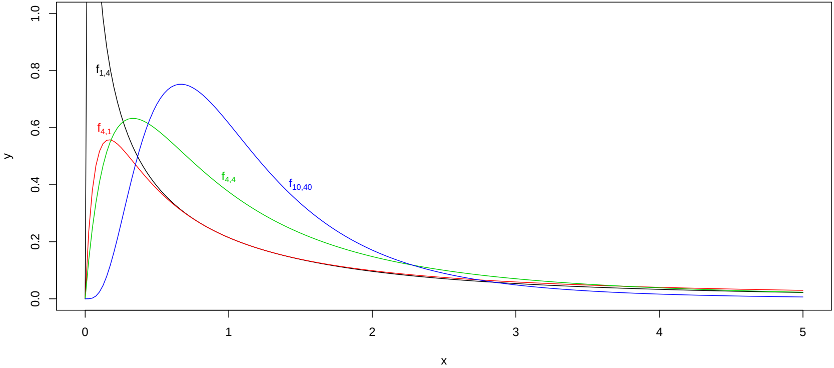
The F distribution (1)

Wirtschafts- und Sozialstatistik
Universität Trier

Let the random variables Y_1 and Y_2 be χ^2-distributed with k_1 and k_2 degrees of freedom, respectively. Then, the random variable

$$X = \frac{Y_1/k_1}{Y_2/k_2}$$

follows an F-distribution with (k_1, k_2) degrees of freedom.



Wirtschafts- und Sozialstatistik
Universität Trier

The F distribution (2)

We have:

$$\operatorname{E} X = \frac{k_2}{k_2 - 2} \qquad \text{for } k_2 > 2$$

$$\operatorname{Var} X = \frac{2(k_1 + k_2 - 2)}{k_1(k_2 - 4)} \left(\frac{k_2}{k_2 - 2} \right)^2 \qquad \text{for } k_2 > 4.$$

Furthermore, we have

$$f(p; k_1, k_2) = \frac{1}{f(1 - p; k_2, k_1)}$$

for the p- and $(1 - p)$-quantile of the F distribution, respectively.

Access on the F distribution in `R`:

```
df(x, df1 = k_1, df2 = k_2)
```

Law of large numbers

Bernoulli's law of large numbers

Assume $\{X_n\}$ to be a sequence of binomially distributed random variables with parameters n and θ $(0 < \theta < 1)$. This sequence $\{X_n\}$ satisfies the *weak law of large numbers* if

$$\lim_{n \to \infty} P(|\frac{X_n}{n} - \theta| > \varepsilon) = 0$$

is valid for each $\varepsilon > 0$.

- The law of large numbers constitutes the foundation of the statistical concept of probability.
- Practical application of the law: Suppose the very same experiment is conducted with a sufficiently larger number of repetitions. With increasing n, the sample mean of this repetitions will approach the theoretical mean.

Central limit theorem

Assume $\{X_n\}$ to be a sequence of independent random variables with existing and positive variances.
In addition,

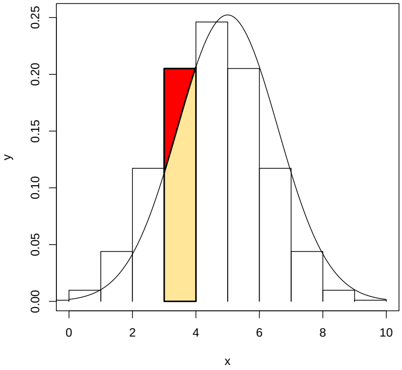
$$Y_n := \frac{\sum_{i=1}^{n} (X_i - \operatorname{E} X_i)}{\sqrt{\sum_{i=1}^{n} \operatorname{Var} X_i}}$$

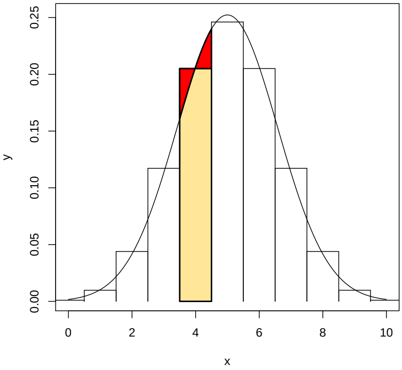
is the standardized sum variable and the random variable Y follows a standard normal distribution. The sequence of random variables $\{X_n\}$ satisfies the *central limit theorem* if

$$\lim_{n\to\infty} F_{Y_n}(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2} \cdot s^2\right) ds \,.$$

Theorem of De Moivre and Laplace

The random variables X_i are bernoulli-distributed with parameter θ $(0 < \theta < 1)$ and stochastically independent. The resulting sequence $\{Y_n\}$ satisfies the central limit theorem.





$b_X(4|10;0.5) = 0.2051$

$\Phi_X(4|5;2.5) - \Phi_X(3|5;2.5) = 0.1606$

$\Phi_X(4.5|5;2.5) - \Phi_X(3.5|5;2.5) = 0.2045$

Theorem of Lindeberg and Lévy

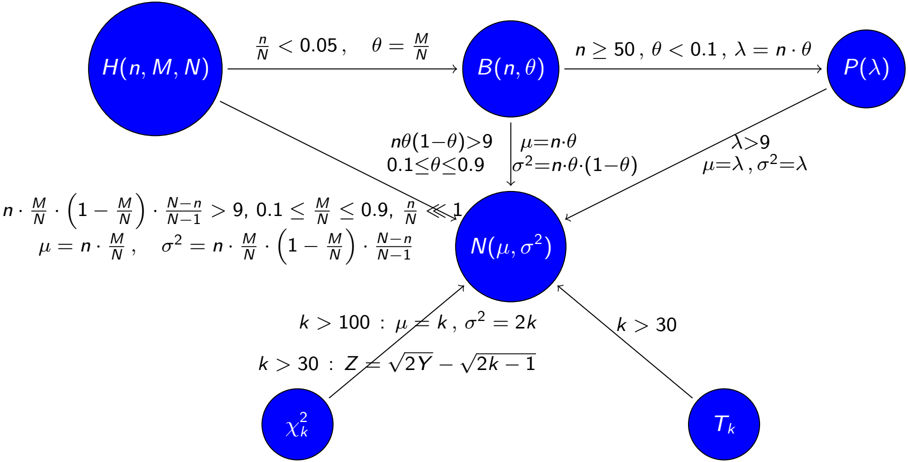
Wirtschafts- und Sozialstatistik
Universität Trier

$\{X_n\}$ being a sequence of independent identically distributed random variables with existing and positive variances. Thus the sequence $\{X_n\}$ satisfies the central limit theorem.

- In contrast to the theorem of de Moivre and Laplace, there is no specific distribution required
- In practice, approximations for n sufficiently large
- Convergence is reached despite of very different distributions
- *Technically* only applicable to sampling models **with replacement**
- There are numerous other central limit theorems

Wirtschafts- und Sozialstatistik
Universität Trier

Options of approximations between distributions



Continuity correction binomial approximation

Wirtschafts- und Sozialstatistik
Universität Trier

$$\begin{aligned}
P(X \leq x) &= B(x|n,\theta) \\
&= \Phi\big(x + 0.5|n\theta, n\theta(1-\theta)\big) \\
P(x_1 < X \leq x_2) &= B(x_2|n,\theta) - B(x_1|n,\theta) \\
&= \Phi\big(x_2 + 0.5|n\theta, n\theta(1-\theta)\big) - \Phi\big(x_1 + 0.5|n\theta, n\theta(1-\theta)\big) \\
P(X \geq x) &= 1 - B(x-1|n,\theta) \\
&= 1 - \Phi\big(x - 1 + 0.5|n\theta, n\theta(1-\theta)\big) \\
&= 1 - \Phi\big(x - 0.5|n\theta, n\theta(1-\theta)\big) \\
P(x_1 \leq X \leq x_2) &= B(x_2|n,\theta) - B(x_1 - 1|n,\theta) \\
&= \Phi\big(x_2 + 0.5|n\theta, n\theta(1-\theta)\big) - \Phi\big(x_1 - 0.5|n\theta, n\theta(1-\theta)\big)
\end{aligned}$$

Example 7.14: Continuity correction (1)

The sample site $n = 100$ and the parameter $\theta = 0.36$ re given. Initially, the conditions of approximation have to be checked.

$$100 \cdot 0.36 \cdot 0.64 = 23.04 > 9 \qquad \surd$$

Now we have $\mu = 100 \cdot 0.36 = 36$ and $\sigma^2 = 23.04 = 4.8^2$.

Calculation of μ and σ^2 in R:

```
n <- 100
Theta7_14 <- 0.36

Mean7_14 <- n * Theta7_14
Var <- n * Theta7_14 * (1 - Theta7_14)
```

```
Mean7_14
[1] 36
```

```
Var
[1] 23.04
```

Example 7.14: Continuity correction (2)

Wirtschafts- und Sozialstatistik
Universität Trier

With this, we get

$$
\begin{aligned}
P(X \leq 40) &= B(40|100; 0.36) \\
&\approx \Phi(40.5|36; 4.8^2) = \Phi(0.94) = 0.8257\,,
\end{aligned}
$$

Three possible ways to calculate $P(X \leq 40)$ in R:

```
Prob7_14_1 <- pbinom(q = 40, size = n, prob = Theta7_14)
Prob7_14_2 <- pnorm(q = (40.5-Mean7_14)/sqrt(Var),
                     mean = 0, sd = 1)
Prob7_14_3 <- pnorm(q = 40.5, mean = Mean7_14, sd = sqrt(Var
    ))
```

```
round(Prob7_14_1, 4)
```

```
[1] 0.8261
```

```
round(Prob7_14_3, 4)
```

```
[1] 0.8257
```

Example 7.14: Continuity correction (3)

... as well as

$$P(X \geq 40) = 1 - B(39|100; 0.36)$$
$$\approx 1 - \Phi(39.5|36; 4.8^2) = 1 - \Phi(0.73) = 0.2329\,.$$

Three possible ways to calculate $P(X \geq 40)$ in R:

```
Prob7_14_4 <- 1 - pbinom(q=39, size=n, prob=Theta7_14)
Prob7_14_5 <- 1 - pnorm(q=(39.5-Mean7_14)/sqrt(Var),
                              mean=0,sd=1)
Prob7_14_6 <- 1 - pnorm(q=39.5,mean=Mean7_14,sd=sqrt(Var))
```

```
round(Prob7_14_4, 4)
[1] 0.2316
```

```
round(Prob7_14_6, 4)
[1] 0.2329
```

Suppose a hypergeometric distribution (model **without replacement**, $N = 1000$, hence $0.1 \ll 1$) is used instead, the additional correction factor would lead to an approximation by $N(36; 20.76)$.

Two-dimensional normal distribution

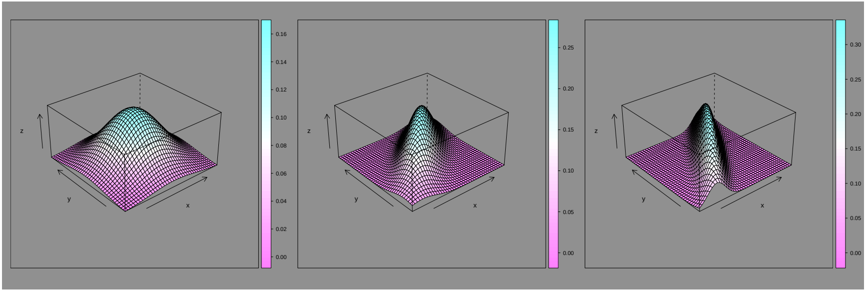
Wirtschafts- und Sozialstatistik
Universität Trier

The two-dimensional normal distribution has the density function $(|\rho| < 1)$

$$\varphi(x_1, x_2 \,|\, \mu_1; \mu_2; \sigma_1^2; \sigma_2^2; \rho) = \frac{1}{2\,\pi\,\sigma_1\,\sigma_2\,\sqrt{1-\rho^2}} \quad \cdot$$

$$\exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\,\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right).$$

The two marginal distributions are $N(\mu_1;\sigma_1^2)$ and $N(\mu_2;\sigma_2^2)$.



Representativeness of samples

A sample is called representative with respect to a population if its result can be transferred to the population in a suitable manner, e.g. it represents the same properties and proportions as that of the population of interest.

- **Small image** of the population
- Structures of the characteristics of interest have to be borne in mind
- **Sampling frame** has to be known and accessible
- Random sampling with known **inclusion probabilities**
- One should be aware of non-sampling errors
- A suitable estimation methodology is required

Probability sampling procedures

- Simple random sampling
  - Model with replacement
  - Model without replacement
- Two-stage sampling processes
  - Stratified sampling
  - Cluster sampling
  - Special two-stage processes
- Methods with unequal sampling probabilities

Please note:

- Methods with fixed sampling sizes
- Methods with alterable sampling sizes

Non-probability sampling procedures

Wirtschafts- und Sozialstatistik
Universität Trier

- Conscious sampling
- Quota sampling
  e.g.: income and consumption sample
  - Demographic features used for adjustments:
    gender, age, regional affiliation
  - Correlation between study- and quota characteristics
  - Non-random sampling within quota cells
- Concentration sample
  Elements with high concentration
  $\rightarrow$ Projection / distribution
- Snowball sampling

Simple random sample

Wirtschafts- und Sozialstatistik
Universität Trier

Random sample

A process that draws n elements from a population of size N ($n < N$) in a random and successive manner is called **random sampling procedure**. The result of such a method is a **random sample** or **probability sample**.

Simple random sample

If the random experiment corresponds to the urn model with or without replacement it is called a **simple random sample** (SRS).

Sampling with and without replacement

Wirtschafts- und Sozialstatistik
Universität Trier

Sampling with replacement

There are N^n possible different samples when drawing elements with replacement (WR). Each sample might be drawn with identical probability. Each element can be drawn 0 up to n times. The individual draws are **stochastically independent**.

Sampling without replacement

There are $N!/(N-n)!$ possible different samples when drawing elements without replacement (WOR). Each sample might be drawn with identical probability. Each element can be drawn 0 or 1 times. The individual draws are **stochastically dependent**.

Sampling functions and distributions

Wirtschafts- und Sozialstatistik
Universität Trier

A sample function is a function

$$u(x_1, x_2, \ldots, x_n),$$

which evaluates the realisations of a variable of interest observed in a sample.

Examples:

- Sample mean: $\overline{x}$
- Sample proportion: p
- Sampling variance: s^2

Which distributions do these sample functions follow?
Distributions of the random variables $\overline{X}$, P and S^2, respectively.

Sample mean under SRS

Wirtschafts- und Sozialstatistik
Universität Trier

We observe n realisations $x_1, \ldots, x_n$ of study variable X.

The sample mean is calculated as: $\overline{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$

Expected value (WR and WOR):

$$\mathrm{E}\,\overline{X} = \mathrm{E}\left(\frac{1}{n} \cdot (X_1 + \ldots + X_n)\right) = \frac{1}{n} \cdot n \cdot \mathrm{E}\,X = \mathrm{E}\,X$$

Variance (WR):

$$\mathrm{Var}\,\overline{X} = \frac{1}{n} \cdot \mathrm{Var}\,X$$

Variance (WOR):

$$\mathrm{Var}\,\overline{X} = \frac{1}{n} \cdot \mathrm{Var}\,X \cdot \frac{N-n}{N-1}$$

Sample total under SRS

The sample total is given by $\sum_{i=1}^{n} x_i$. Since the relation of population to sample size N/n has to be taken into account, a representative value of the total is given by the sample function $\frac{N}{n} \cdot \sum_{i=1}^{n} = N \cdot \overline{X}$.

Expected value (WR and WOR):

$$\mathrm{E}\, N \cdot \overline{X} = \mathrm{E}\left(N \cdot \frac{1}{n} \cdot (X_1 + \ldots + X_n)\right) = \frac{N}{n} \cdot n \cdot \mathrm{E}\, X = N \cdot \mathrm{E}\, X$$

Variance (WR):

$$\mathrm{Var}\, N \cdot \overline{X} = N^2 \cdot \frac{1}{n} \cdot \mathrm{Var}\, X$$

Variance (WOR):

$$\mathrm{Var}\, N \cdot \overline{X} = N^2 \cdot \frac{1}{n} \cdot \mathrm{Var}\, X \cdot \frac{N-n}{N-1}$$

Sample proportion under SRS

A dichotomous population with $W(X = 1) = \theta$ and $W(X = 0) = 1 - \theta$ as well as $\mathrm{E}\,X = \theta$ and $\mathrm{Var}\,X = \theta \cdot (1 - \theta)$ is considered.

The sample proportion is calculated as: $p = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$.

Expected value (WR and WOR):

$$\mathrm{E}\,P = \mathrm{E}\left(\frac{1}{n} \cdot (X_1 + \ldots + X_n)\right) = \theta$$

Variance (WR):

$$\mathrm{Var}\,P = \frac{1}{n} \cdot \theta \cdot (1 - \theta)$$

Variance (WOR):

$$\mathrm{Var}\,P = \frac{1}{n} \cdot \theta \cdot (1 - \theta) \cdot \frac{N - n}{N - 1}$$

Sampling variance under SRS

The sampling variance is calculated as: $s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \overline{x})^2$

Expected value (WR):

$$\mathrm{E}\, S^2 = \operatorname{Var} X$$

Expected value (WOR):

$$\mathrm{E}\, S^2 = \frac{N}{N-1} \cdot \operatorname{Var} X$$

Variance (WR):

$$\operatorname{Var} S^2 = \frac{2}{n-1} \cdot \left(\operatorname{Var} X\right)^2$$

General remarks on sample functions

- Aim: estimation of unknown parameters of the underlying population.
  e.g. the true, unknown mean of the population μ

  Estimation using $\widehat{\mu}$.

  - Is the sample function $\overline{x}$ applicable to this problem?
  - Which properties does $\overline{x}$ have?
- Is it possible to make statements concerning the distributions of the sample functions?
  Partially yes. See special functions of the normal distribution

In practice, only the results - and not the true values - are available!