Probability and Statistics (English) FS 2020 Session Exam

08.02.2021 **Time Limit: 180 Minutes**

Surname		**First name**		**Legi Number**								**Examnumber**		
				X	X									

the first two letters of each — *last 6 digits* — *leave blank*

Enter **the first two letters of your surname and first name now**, as well as **the last six digits of your Legi number**. If you are attaching separate sheets, write the same information and **only this information clearly at the top of each sheet**.

This exam contains 10 problems and in addition statistical tables.

Grading Table (for grading use only, please leave empty)

Question:	1	2	3	4	5	6	7	8	9	10	Total
Maximal:	10	10	10	10	10	10	10	10	10	10	100
Points:											
Checkup:											

Instructions

The exam is a closed-book exam (no pocket calculators or mobile phones either).

Before the exam:

- ◇ Put your student card (Legi) on the table.
- ◇ Fill out the cover sheet as appropriate. Do not open the exam.

During the exam:

- ◇ Read the questions carefully, and try to answer as many as you can. Start each problem on a new sheet, and write the information above on each sheet. Do not use red or green ink, nor pencil.
- ◇ You may use results from the lecture or from the formulae sheet without proof, unless you are explicitly asked to give a proof. But to get full points, it must be clear from your solutions how you found your answer. A correct result alone will not always give all points.
- ◇ Try to simplify your answers as much as possible. If a numerical calculation requires a calculator, you may give your answer as a numerical expression — e.g. $\frac{\sqrt{3}}{7^2}$ can be a valid solution (but for instance $\int_0^1 \frac{\sqrt{3}}{7^2}dx$ or $\frac{\sqrt{300}}{490}$ would not give full points).

After the exam:

- ◇ Order your answer sheets in a reasonable way, and place everything inside the envelope you are given.
- ◇ After your exam sheets have been collected by a supervisor, remain seated and follow instructions.

1. (10 points)

   (a) (5 points) A small voting district has K female and L male voters. A random sample of N voters, where $N \leq K + L$, is drawn uniformly at random from the population.

   (i) Give a probabilistic model $(\Omega, \mathcal{F}, \mathbb{P})$ to describe this situation.

   (ii) Let $n \leq \min\{K, N\}$. What is the probability that exactly n of the drawn N voters will be female? Do you recognize a known distribution? Write down its name.

   (b) (5 points) At a wedding, a group of $n \geq 3$ people (including the married couple, Alice and Bob) want to have a picture taken. They all stand in a line, with the order of the people in the line taken uniformly at random among the permutations of n elements. What is the probability that exactly k guests stand between Alice and Bob, for $k \in \{0, 1, ..., n-2\}$?

Solution:

(a) (i) First we define the set of all $K + L$ voters as $M := \{1, \ldots, K + L\}$. Then we define $\Omega := \{W \subseteq M : |W| = N\}$ as the set of all N-element subsets of M. Then we take $\mathcal{F} = 2^\Omega$ and $\mathbb{P} : \mathcal{F} \to [0,1], A \mapsto \mathbb{P}[A] = \frac{|A|}{|\Omega|} = \frac{|A|}{\binom{L+K}{N}}$, since there are $\binom{L+K}{N}$ ways to pick the N voters among the $L + K$ district voters.

(ii) There are $\binom{K}{n}\binom{L}{N-n}$ ways to choose the n female and $N-n$ male voters. Let X be the random variable that models the number of female voters when choosing N voters at random. The probability to pick n female voters when choosing N voters at random is then

$$\mathbb{P}[X = n] = \frac{\binom{K}{n}\binom{L}{N-n}}{\binom{L+K}{N}}.$$

This is a hypergeometric distribution.

(b) First, we order the guests and assign them a number between 3 and n (1 is given to Alice and 2 to Bob). We are looking for the number of permutations π such that $|\pi(1) - \pi(2)| = k + 1$. Let $\mathcal{P}_k$ be defined as the set of such permutations.

Such a permutation is constructed as follows (without loss of generality, assume that $\pi(1) < \pi(2)$; any such permutation in $\mathcal{P}_k$ corresponds to another permutation π' in $\mathcal{P}_k$ with $\pi'(1) = \pi(2)$ and $\pi'(2) = \pi(1)$):

- Choose a value for $\pi(1)$ in $\{1, 2, ..., n-k-1\}$.
- Choose k of the guests (integer in $\{3, ..., n\}$). There are $k!$ ways to assign them values between $\pi(1) + 1$ and $\pi(1) + k = \pi(2) - 1$.
- Then there are $(n-2-k)!$ ways to assign the remaining guests to the remaining places in the line.

Thus there are $2(n-k-1)\binom{n-2}{k}k!(n-2-k)!$ permutations in $\mathcal{P}_k$. Since the guest permutation is chosen at random and the total number of permutations of n elements is $n!$, the probability that exactly k guests are between Alice and Bob is given by

$$\frac{2(n-k-1)\binom{n-2}{k}k!(n-2-k)!}{n!} = \frac{2(n-k-1)}{n(n-1)}.$$

2. (10 points) A tetraeder is a geometric body with 4 equal sides. It can be used as a die with 4 possible outcomes. We have two fair tetraeder dice. One has on its sides the odd numbers $1, 3, 5, 7$, the other the even numbers $2, 4, 6, 8$. We flip a coin to decide which die is rolled. If the coin shows heads, we roll the odd-numbered die, and otherwise the even-numbered one.

   (a) (3 points)
      (i) Describe this experiment by a suitable probability space $(\Omega, \mathcal{F}, \mathbb{P})$, using a Laplace model.
      (ii) What is the cardinality $|\mathcal{F}|$ of $\mathcal{F}$? Give examples of events E_1, E_2, E_3, E_4 in $\mathcal{F}$ such that $\mathbb{P}[E_i] \neq \mathbb{P}[E_j]$ for all $i \neq j$. Compute their probabilities $\mathbb{P}[E_i]$. *Hint:* To obtain full points, write E_1, E_2, E_3, E_4 in precise mathematical notation using the previously defined probability space. It is recommended to choose very simple events.

   (b) (2 points) What is the probability that rolling the die produces a prime number?

   (c) (5 points)
      (i) Define random variables $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ such that X and Y represent the outcomes of flipping the coin and of rolling the die, respectively.
      (ii) Are the random variables X and Y independent? Prove your answer.

Solution:

(a) (i) We mentally label the sides of each die by $1, 2, 3, 4$ in increasing order of the number on each side. Then we can take $\Omega := \{0, 1\} \times \{1, 2, 3, 4\}$, $\mathcal{F} := 2^\Omega$, $\mathbb{P} : \mathcal{F} \to [0, 1], A \mapsto \mathbb{P}[A] = \frac{|A|}{|\Omega|} = \frac{|A|}{8}$.

Alternatively one could define

$$\Omega' := \{ (0, 1), (0, 3), (0, 5), (0, 7), (1, 2), (1, 4), (1, 6), (1, 8) \}.$$

Another possibility would be to define

$$\Omega'' := \{1, 2, 3, 4, 5, 6, 7, 8\}.$$

(ii) $|\mathcal{F}| = 2^8 = 256$, so there are many different possibilities to choose the examples—e.g. $E_1 = \varnothing$, $E_2 = \{(0, 1)\}$, $E_3 = \{(0, 1), (0, 3), (1, 2)\}$, $E_4 = \Omega$ with $\mathbb{P}[E_1] = 0$, $\mathbb{P}[E_2] = \frac{1}{8}$, $\mathbb{P}[E_3] = \frac{3}{8}$, $\mathbb{P}[E_4] = 1$.

(b) The prime numbers are $2, 3, 5, 7$. This corresponds to $A = \{(1, 1), (0, 2), (0, 3), (0, 4)\}$ or $A' = \{(0, 3), (0, 5), (0, 7), (1, 2)\}$ or $A'' = \{2, 3, 5, 7\}$. In every case, the event has cardinality 4 so that the desired probability is $\frac{4}{8} = \frac{1}{2}$.

(c) (i) $X : \Omega \to \mathbb{R}, \omega \mapsto X(\omega) := \omega_1$ and

$$Y : \Omega \to \mathbb{R}, \omega \mapsto Y(\omega) := \begin{cases} 2\omega_2 - 1, & \omega_1 = 0, \\ 2\omega_2, & \omega_1 = 1. \end{cases}$$

An alternative way to write this is $Y(\omega) := (2\omega_2-1)I_{\{\omega_1=0\}}+(2\omega_2)I_{\{\omega_1=1\}}$ or $Y(\omega) := 2\omega_2 - 1 + \omega_1$.

On Ω', Y can be defined more simply as $Y(\omega') := \omega_2'$.

On Ω'', Y can be defined more simply as $Y(\omega'') := \omega''$, but X would need to be defined via a case analysis.

(ii) X and Y are dependent. To show this, it is sufficient to find one example of $x, y \in \mathbb{R}$ such that $\mathbb{P}[X = x, Y = y] \neq \mathbb{P}[X = x]\,\mathbb{P}[Y = y]$. A simple choice is $x = 0$ and $y = 2$. First we compute the left-hand side as

$$\mathbb{P}[X = 0, Y = 2] = \frac{|\{\omega \in \Omega : X(\omega) = 0 \text{ and } Y(\omega) = 2\}|}{8} = 0.$$

Now, we consider the right-hand side and first calculate the probability

$$\mathbb{P}[X = 0] = \mathbb{P}[\{\omega \in \Omega : X(\omega) = 0\}] = \frac{|\{\omega \in \Omega : X(\omega) = 0\}|}{8} = \frac{4}{8} = \frac{1}{2}.$$

Analogously, we get

$$\mathbb{P}[Y = 2] = \mathbb{P}[\{\omega \in \Omega : Y(\omega) = 2\}] = \frac{|\{\omega \in \Omega : Y(\omega) = 2\}|}{8} = \frac{1}{8}.$$

So we can conclude that

$$\mathbb{P}[X = 0, Y = 2] = 0 \neq \frac{1}{16} = \mathbb{P}[X = 0]\,\mathbb{P}[Y = 2].$$

3. (10 points) Alice rolls a die and pays the resulting number to Bob in CHF if the number is 1,5 or 6. Otherwise she does not pay him anything.

   (a) (1 point) Define a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that describes rolling the die.

   (b) (1 point) Let X be the random variable that describes how many CHF Bob receives. Write down the formal definition of X.

   (c) (2 points) Compute the expected value $\mathbb{E}[X]$.

   (d) (1 point) Compute the cumulative distribution function F of X.

   (e) (1 point) Compute the median $F^{\leftarrow}(1/2)$ of X.

   (f) (2 points) Compute the variance $\mathrm{Var}[X]$.

   (g) (2 points) Compute the expectation of $3X$ and the standard deviation of $-4X$.

Solution:

(a) $\Omega := \{1,2,3,4,5,6\}$, $\mathcal{F} = 2^{\Omega}$, $\mathbb{P}[A] := \sum_{\omega \in A} p(\omega)$ with $p(\omega) := \frac{1}{6}\ \forall \omega \in \Omega$.
(or $\mathbb{P}[A] = \frac{|A|}{|\Omega|} = \frac{|A|}{6}$)

(b) $X : \Omega \to \underbrace{\{0,1,5,6\}}_{\text{alternativly a superset, e.g. } \mathbb{N}_0, \mathbb{R} \text{ could be used here}},\ \omega \mapsto X(\omega) := \begin{cases} \omega, & \text{if } \omega \in \{1,5,6\}, \\ 0, & \text{else.} \end{cases}$

(c) Using (1.9) from the lecture notes, we have

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) p(\omega) = \sum_{\omega \in \{1,2,3,4,5,6\}} X(\omega) \frac{1}{6} = \frac{1+5+6}{6} = \frac{12}{6} = 2.$$

Alternative solution: Using the right-hand side from (1.10) in the lecture notes, we have

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x p_X(x) = \sum_{x \in \{0,1,5,6\}} x p_X(x) = \frac{1+5+6}{6} = \frac{12}{6} = 2.$$

(d) $F(x) = \begin{cases} 0, & \text{if } x < 0, \\ \frac{3}{6}, & \text{if } 0 \le x < 1, \\ \frac{4}{6}, & \text{if } 1 \le x < 5, \\ \frac{5}{6}, & \text{if } 5 \le x < 6, \\ 1, & \text{if } 6 \le x. \end{cases}$

(e) $F^{\leftarrow}(1/2) = 1$

(f)

$$\mathrm{Var}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 \tag{1a}$$
$$= \sum_{\omega \in \Omega} X^2(\omega) p(\omega) - \mathbb{E}[X]^2 \tag{1b}$$
$$= \sum_{\omega \in \{1,2,3,4,5,6\}} X^2(\omega) \frac{1}{6} - 2^2 \tag{1c}$$
$$= \frac{1+25+36}{6} - 4 = \frac{62}{6} - 4 = \frac{31}{3} - 4 = \frac{19}{3} \approx 6.3333333. \tag{1d}$$

Alternative solution:

$$\begin{aligned}
\mathrm{Var}[X] &= \mathbb{E}\left[(X-\mathbb{E}[X])^2\right] && \text{(2a)}\\
&= \sum_{\omega\in\Omega}(X-\mathbb{E}[X])^2(\omega)p(\omega) && \text{(2b)}\\
&= \sum_{\omega\in\{1,2,3,4,5,6\}}(X-2)^2(\omega)\frac{1}{6} && \text{(2c)}\\
&= \frac{(1+3\cdot 4+9+16)}{6} = \frac{38}{6} = \frac{19}{3} \approx 6.3333333. && \text{(2d)}
\end{aligned}$$

(g)

$$\mathbb{E}[3X] = 3\mathbb{E}[X] = 6$$

$$\sqrt{\mathrm{Var}[-4X]} = \sqrt{(-4)^2\,\mathrm{Var}[X]} = 4\sqrt{\mathrm{Var}[X]} = 4\sqrt{\frac{19}{3}} \approx 10.066$$

4. (10 points) We consider two urns A and B. Urn A contains 2 green and 3 blue balls, urn B 2 green and 1 blue.

   (a) (4 points) We randomly choose one urn and then randomly draw from that urn 2 balls *with* replacement. What is the probability that urn A was chosen if both drawn balls are green?

   (b) (3 points) We change the experiment as follows. We randomly choose one urn and then randomly draw from that urn 2 balls *without* replacement. What is the probability that urn A was chosen and both drawn balls are green?

   (c) (3 points) We again change the experiment. We randomly draw directly from urn A *without* replacement. What is the expected number of drawn blue balls?

Solution: Seien U_A das Ereignis, dass Urne A gewählt wird, und U_B das Ereignis, dass Urne B gewählt wird. Es gilt $\mathbb{P}[U_A] = \mathbb{P}[U_B] = 1/2$.

Sei $\{(b,g)\}$ das Ereignis, dass die erste gezogene Kugel blau und die zweite gezogene Kugel grün ist. Die Ereignisse $\{(g,b)\}$, $\{(b,b)\}$ und $\{(g,g)\}$ werden analog definiert. Der Kürze halber schreiben wir nur (b,g) usw.

(a) Gesucht ist $\mathbb{P}[U_A|\,(g,g)]$. Laut Aufgabenstellung gilt

$$\mathbb{P}[(g,g)|\,U_A] = \frac{2}{5}\cdot\frac{2}{5} = \frac{4}{25} \qquad \text{und} \qquad \mathbb{P}[(g,g)|\,U_B] = \frac{2}{3}\cdot\frac{2}{3} = \frac{4}{9}.$$

Nach der Formel von Bayes gilt

$$\mathbb{P}[U_A|\,(g,g)] = \frac{\mathbb{P}[(g,g)|\,U_A]\,\mathbb{P}[U_A]}{\mathbb{P}[(g,g)|\,U_A]\,\mathbb{P}[U_A] + \mathbb{P}[(g,g)|\,U_B]\,\mathbb{P}[U_B]} = \frac{\frac{4}{25}}{\frac{4}{25}+\frac{4}{9}} = \frac{9}{34}.$$

(b) Gesucht ist $\mathbb{P}[U_A \cap \{(g,g)\}]$. Laut Aufgabenstellung gilt

$$\mathbb{P}[(g,g)|\,U_A] = \frac{2}{5}\cdot\frac{1}{4} = \frac{2}{20} = \frac{1}{10}$$

Somit gilt

$$\mathbb{P}[U_A \cap \{(g,g)\}] = \mathbb{P}[(g,g)|\,U_A]\,\mathbb{P}[U_A] = \frac{2}{20}\frac{1}{2} = \frac{1}{20}.$$

(c) Sei N_b die Anzahl der aus Urne A gezogenen blauen Kugeln. Es gilt

$$\begin{aligned}\mathbb{E}[N_b] &= \mathbb{P}[N_b = 1] + 2\mathbb{P}[N_b = 2] = \mathbb{P}[(b,g)\cup(g,b)] + 2\mathbb{P}[(b,b)] \\ &= \mathbb{P}[(b,g)] + \mathbb{P}[(g,b)] + 2\mathbb{P}[(b,b)]\,.\end{aligned}$$

Mit dem angegebenen Verfahren gilt

$$\mathbb{P}[(b,g)] = \frac{3}{5}\cdot\frac{2}{4} = \frac{3}{10}, \quad \mathbb{P}[(g,b)] = \frac{2}{5}\cdot\frac{3}{4} = \frac{3}{10} \quad \text{und} \quad \mathbb{P}[(b,b)] = \frac{3}{5}\cdot\frac{2}{4} = \frac{3}{10}.$$

Wir erhalten somit

$$\mathbb{E}[N_b] = \frac{3}{10} + \frac{3}{10} + 2\frac{3}{10} = \frac{12}{10} = \frac{6}{5}.$$

5. (10 points) Let X_1 and X_2 be independent random variables, both uniformly distributed on $[0,1]$, and set $X = \max\{X_1, X_2\}$.

   (a) (2 points) Compute the density function $f_X(x), x \in \mathbb{R}$, of the random variable X.

   (b) (2 points) Compute the probability $\mathbb{P}[X_1 \le x | X \ge y]$ for a fixed $y \in (0,1)$.

   The subsequent questions have no connection to (a) and (b).

   We randomly choose a point $B = (U, V)$ in the domain D; see Figure 1. This implies that the joint density of (U, V) has the form:

$$f_{U,V}(u,v) = \begin{cases} c, & \text{if } (u,v) \in D, \\ 0, & \text{otherwise,} \end{cases}$$

   for a constant $c \in \mathbb{R}$.

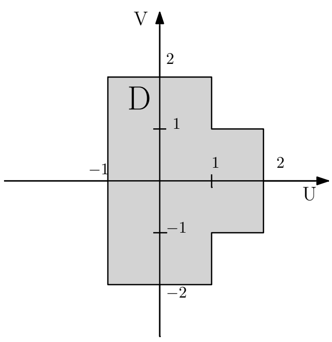


Figure 1: The point B is chosen randomly in the domain D.

   (c) (2 points) Determine the constant c.

   (d) (4 points) Are U and V independent? Are they uncorrelated? Argue your answer.

Solution:

(a) Für $x < 0$ ist $F_X(x) = 0$. Für $0 \le x \le 1$ gilt wegen der Unabhängigkeit

$$\begin{aligned} F_X(x) &= \mathbb{P}[X \le x] = \mathbb{P}[\max\{X_1, X_2\} \le x] \\ &= \mathbb{P}[X_1 \le x, X_2 \le x] = \mathbb{P}[X_1 \le x]\,\mathbb{P}[X_2 \le x] = x^2. \end{aligned}$$

Schliesslich ist $F_X(x) = 1$ für $x > 1$.
Für $x \in [0,1]$ ist $f_X(x) = \frac{d}{dx} F_X(x) = 2x$. Dann folgt

$$f_X(x) = \begin{cases} 2x, & \text{falls } 0 \le x \le 1, \\ 0, & \text{sonst.} \end{cases}$$

(b) Für $x < 0$ ist $\mathbb{P}\left[X_1 \leq x | X \geq y\right] = 0$. Für $0 \leq x \leq y$ gilt

$$\begin{aligned}
&\mathbb{P}\left[X_1 \leq x | X \geq y\right] \\
&= \frac{\mathbb{P}\left[X_1 \leq x, \max\{X_1, X_2\} \geq y\right]}{\mathbb{P}\left[\max\{X_1, X_2\} \geq y\right]} \\
&= \frac{\mathbb{P}\left[X_1 \leq x, X_2 \geq y\right]}{1 - y^2} = \frac{x(1-y)}{1-y^2} = \frac{x}{1+y}.
\end{aligned}$$

Für $y < x \leq 1$ ist

$$\begin{aligned}
&\mathbb{P}\left[X_1 \leq x | X \geq y\right] \\
&= \frac{\mathbb{P}\left[X_1 \leq x, \max\{X_1, X_2\} \geq y\right]}{\mathbb{P}\left[\max\{X_1, X_2\} \geq y\right]} \\
&= \frac{\mathbb{P}\left[X_1 \leq y, X_2 \geq y\right] + \mathbb{P}\left[y \leq X_1 \leq x\right]}{1 - y^2} \\
&= \frac{y(1-y) + (x-y)}{1-y^2} = \frac{y - y^2 + x - y}{1-y^2} = \frac{x - y^2}{1-y^2}.
\end{aligned}$$

Für $x > 1$ ist $\mathbb{P}\left[X_1 \leq x | X \geq y\right] = 1$.

(c) Für die Dichtefunktion $f_{U,V}$ muss gelten

$$1 = \iint_{\mathbb{R}^2} f_{U,V}(u,v)du\,dv = \iint_{\mathbb{R}^2} cI_D(u,v)du\,dv = c(\text{Fläche von } D) = 10c.$$

Also ist $c = \frac{1}{10}$.

(d)

$$\begin{aligned}
\mathbb{E}\left[U\right] &= c\iint_{\mathbb{R}^2} uI_D(u,v)du\,dv \\
&= c\int_{-1}^{1} u4du + c\int_{1}^{2} u2du \\
&= c(2u^2\big|_{-1}^{1} + u^2\big|_{1}^{2}) \\
&= 3c = \frac{3}{10} \\
\mathbb{E}\left[V\right] &= c\iint_{\mathbb{R}^2} vI_D(u,v)du\,dv \\
&= c\int_{-2}^{2}\int_{-1}^{1} du\, v\, dv + c\int_{-1}^{1}\int_{1}^{2} du\, v\, dv \\
&= 2c\frac{1}{2}v^2\Big|_{-2}^{2} + c\frac{1}{2}v^2\Big|_{-1}^{1} \\
&= 0. \\
\mathbb{E}\left[UV\right] &= c\iint_{\mathbb{R}^2} uvI_D(u,v)du\,dv \\
&= c\int_{-1}^{1} udu \int_{-2}^{2} vdv + c\int_{1}^{2} udu \int_{-1}^{1} vdv \\
&= 0.
\end{aligned}$$

Somit ist $\text{Cov}(U,V) = 0$, also unkorreliert.

U und V sind nicht unabhängig. Allgemein ist $\mathbb{P}[(U,V) \in A] = \iint_A f_{U,V}(u,v)du\,dv$. Also ist $\mathbb{P}\left[U \geq \frac{3}{2}, V \geq \frac{3}{2}\right] = 0$ (weil $f_{U,V}$ dort 0 ist), $\mathbb{P}\left[U \geq \frac{3}{2}\right] = c \cdot 1 = c = \frac{1}{10}$, $\mathbb{P}\left[V \geq \frac{3}{2}\right] = c \cdot 1 = c = \frac{1}{10}$ (weil $f_{U,V}$ auf D konstant c ist) und damit

$$\mathbb{P}\left[\left\{U \geq \frac{3}{2}\right\} \cap \left\{V \geq \frac{3}{2}\right\}\right] = 0 \neq \frac{1}{100} = \mathbb{P}\left[U \geq \frac{3}{2}\right]\mathbb{P}\left[V \geq \frac{3}{2}\right].$$

6. (10 points) A random sample of n items is taken from a distribution with mean μ and standard deviation σ. We want to understand how many items are needed in order to have the relation

$$\mathbb{P}\left[|\overline{X}_n - \mu| \le \frac{\sigma}{4}\right] \ge 0.95, \qquad \text{where } \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i. \qquad (\star)$$

(a) (4 points) If the X_i are not independent, find a value for n by using an appropriate inequality. Specify your assumptions.

(b) (6 points) Under different assumptions on the X_i, find a (smaller) value for n such that $(\star)$ is valid approximately. Again specify your assumptions.

Solution:

(a) Nach Chebyshev gilt

$$\mathbb{P}\left[|\overline{X}_n - \mu| > c\right] \le \frac{\mathrm{Var}\left[\overline{X}_n\right]}{c^2}$$

Weiter ist $\mathbb{E}\left[\overline{X}_n\right] = \mu$, und wenn wir Unkorreliertheit der X_i annehmen, gilt $\mathrm{Var}\left[\overline{X}_n\right] = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}\left[X_i\right] = \frac{\sigma^2}{n}$. Mit $c = \frac{\sigma}{4}$ erhalten wir also

$$\mathbb{P}\left[|\overline{X}_n - \mu| > \frac{\sigma}{4}\right] \le \frac{\sigma^2/n}{\sigma^2/16} = \frac{16}{n},$$

und wenn das $\le 1 - 0.95 = 0.05$ sein soll, brauchen wir $n \ge 16\frac{100}{5} = 320$.

(b) We assume that $X_1, \ldots, X_n$ are i.i.d. Define $Z_n = \frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu)$. By the central limit theorem, the random variable Z_n can be approximated by the standard normal distribution. Let Φ be the c.d.f. of the standard normal distribution. We have

$$\mathbb{P}\left[|\overline{X}_n - \mu| \le \frac{\sigma}{4}\right] = \mathbb{P}\left[|Z_n| \le \frac{\sqrt{n}}{4}\right] \approx \Phi\Big(\frac{\sqrt{n}}{4}\Big) - \Phi\Big(-\frac{\sqrt{n}}{4}\Big) = 2\Phi\Big(\frac{\sqrt{n}}{4}\Big) - 1\,.$$

Thus we need $\Phi\big(\frac{\sqrt{n}}{4}\big) \ge 0.975$, which implies $n \ge 16\left(\Phi^{\leftarrow}(0.975)\right)^2 = 16 \cdot 1.96^2 \approx 61.5$.

7. (10 points) We consider high-water levels in the lake of Zurich, where high-water means that the water level exceeds the critical level of 140 cm above the normal level. Let the random variable X describe the water level in cm above the critical level. We model X by the so-called generalised Pareto distribution with the density function

$$f_X(x;\theta) = \begin{cases} \frac{1}{\theta}(1+x)^{-(1+\frac{1}{\theta})}, & \text{if } x > 0, \\ 0, & \text{if } x \le 0. \end{cases}$$

Here θ is an unknown parameter which is to be estimated on the basis of data $x_1, \dots, x_n$. These are as usual viewed as realizations of random variables $X_1, \dots, X_n$ which are for each θ i.i.d. under $\mathbb{P}_\theta$ with density $f_X(x;\theta)$.

(a) (3 points) Let $\mathbf{X} = (X_1, \dots, X_n)$. Show that

$$T_n = t_n(\mathbf{X}) := \sum_{i=1}^{n} \frac{\log(1+X_i)}{n}$$

is the maximum likelihood estimator for θ.

(b) (3 points) Determine the distribution of $\log(1+X_i)$ under $\mathbb{P}_\theta$.

(c) (3 points) Compute the mean and variance of T_n in each model $\mathbb{P}_\theta$.
Hint: If you have not solved (b), assume that $Y_i := \log(1+X_i)$ is under $\mathbb{P}_\theta$ $\mathrm{G}(1, \frac{1}{\theta})$-distributed.

(d) (1 point) Is the sequence $T_n, n \in \mathbb{N}$, of estimators consistent?

Solution:

(a) Die log-Likelihood-Funktion ist gegeben durch

$$\begin{aligned}\log L(x_1, \dots, x_n; \theta) &= \log\left(\frac{1}{\theta^n}\prod_{i=1}^{n}(1+x_i)^{-(1+\frac{1}{\theta})}\right) \\ &= -n\log\theta - (1+\frac{1}{\theta})\sum_{i=1}^{n}\log(1+x_i).\end{aligned}$$

Die Ableitung nach θ ist

$$\frac{\partial}{\partial\theta}\log L(x_1, \dots, x_n; \theta) = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n}\log(1+x_i),$$

und das ist 0 für

$$\theta = \frac{1}{n}\sum_{i=1}^{n}\log(1+x_i).$$

Der ML-Schätzer für θ ist also gegeben durch

$$T_n = \frac{1}{n}\sum_{i=1}^{n}\log(1+X_i).$$

(b) Mit $Y_i := \log(1+X_i)$ ist $\mathbb{P}_\theta[Y_i \le t] = \mathbb{P}_\theta[X_i \le e^t - 1]$, also $f_Y(t;\theta) = \frac{\partial}{\partial t}\mathbb{P}_\theta[Y_i \le t] = f_X(e^t-1;\theta)e^t = \frac{1}{\theta}(e^t)^{-(1+\frac{1}{\theta})}e^t = \frac{1}{\theta}e^{-t\frac{1}{\theta}}$ für $e^t - 1 > 0$, also $t > 0$. Also ist $Y_i \sim \text{Exp}(\frac{1}{\theta})$ unter $\mathbb{P}_\theta$.

(c) Die Linearität des Erwartungswertes liefert

$$\mathbb{E}_\theta[T_n] = \mathbb{E}_\theta\left[\frac{1}{n}\sum_{i=1}^n \log(1+X_i)\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_\theta[\log(1+X_i)] = \mathbb{E}_\theta[\log(1+X_1)].$$

Nach (b) (oder dem Hinweis) ist $Y_1 := \log(1+X_1)$ $\text{Exp}(\frac{1}{\theta})$-verteilt (bzw. $G(1,\frac{1}{\theta})$-verteilt). Also ist $\mathbb{E}[Y_1] = \theta$ und $\text{Var}[Y_1] = \theta^2$. Daher gilt

$$\mathbb{E}_\theta[T_n] = \mathbb{E}_\theta[Y_1] = \frac{1}{\frac{1}{\theta}} = \theta.$$

Ähnlich rechnen wir für die Varianz mit Unabhängigkeit

$$\begin{aligned}\text{Var}_\theta[T_n] &= \text{Var}_\theta\Big[\frac{1}{n}\sum_{i=1}^n \log(1+X_i)\Big] \\ &= \frac{1}{n^2}\sum_{i=1}^n \text{Var}_\theta[\log(1+X_i)] \\ &= \frac{1}{n}\text{Var}_\theta[\log(1+X_1)] \\ &= \frac{1}{n}\text{Var}_\theta[Y_1] = \frac{1}{n}\frac{1}{\frac{1}{\theta^2}} = \frac{1}{n}\theta^2.\end{aligned}$$

(d) Die Chebyshev-Ungleichung liefert für jedes $n \in \mathbb{N}$ und für beliebiges $\epsilon > 0$

$$\mathbb{P}_\theta[|T_n - \theta| > \epsilon] \le \frac{1}{\epsilon^2}\text{Var}_\theta[T_n] = \frac{1}{\epsilon^2}\frac{1}{n}\theta^2 \overset{n\to\infty}{\longrightarrow} 0.$$

Dies beweist die Konsistenz.

8. (10 points) We want to find out whether a device for measuring air pollution gives correct values. To that end, we pump 20 ppm of CO into a closed room and then measure the CO level with our device nine times. This yields data $x_1, \ldots, x_9$ with the following values (in ppm of CO): $\overline{x}_9 = 20.44$, $s_x^2 = 0.774$ (in (ppm of CO)^2), $s_x = 0.88$. The measurements can be assumed to be normally distributed (in good approximation).

   (a) (4 points) Which test is appropriate for this situation? Formulate a suitable null hypothesis and alternative, and test at the 5% level.

   (b) (3 points) Explain how one can test the same null hypothesis if the precision $\sigma = 0.800$ (in ppm of CO) of the device (i.e. of each measurement) is known. Give the same details as in (a).

   (c) (3 points) Compute for both cases the corresponding realised 95%-confidence interval for the mean of the measurements.

Solution:

(a) t-**Test**

1. Modell: $(X_i)_{1\le i\le 9}$ i.i.d. $\sim \mathcal{N}(\mu,\sigma^2)$, $\mu = \mathbb{E}[X_i]$ unbekannt, $\sigma^2 = \mathrm{Var}[X_i]$ ebenfalls unbekannt
2. H_0: $\mu = \mu_0 = 20$
3. H_1: $\mu \neq \mu_0$ (zweiseitig)
4. Teststatistik: $T := \frac{\overline{X}_n - \mu_0}{S_x/\sqrt{n}}$
5. Unter H_0 gilt $T \sim t_8$. (Wir erwarten in diesem Fall also einen Wert $T(\omega)$ in der Nähe von 0.)
6. H_0 wird auf dem 5%-Niveau verworfen, falls $|T| > t_{8,0.975}$. Die kritischen Werte $t_{8,0.975}$ sind tabelliert. Für $n = 9$ haben wir $t_{8,0.975} = 2.306$.
7. Mit unseren Daten beträgt der Wert der Teststatistik T

$$T(\omega) = \frac{\overline{x}_n - \mu_0}{\frac{s_x}{\sqrt{n}}} = \frac{20.44 - 20}{\frac{0.88}{\sqrt{9}}} = \frac{0.44}{\frac{0.88}{3}} = \frac{3}{2} = 1.5 < t_{8,0.975} = 2.306.$$

   Das Gerät kann richtig geeicht sein. Die Nullhypothese wird nicht verworfen.

(b) z-**Test**

1. Modell: $(X_i)_{1\le i\le 9}$ i.i.d. $\sim \mathcal{N}(\mu,\sigma^2)$, μ unbekannt, σ^2 bekannt (im Gegensatz zu Aufgabe (a)).
2. H_0: $\mu = \mu_0 = 20$
3. H_1: $\mu \neq \mu_0$ (zweiseitig)
4. Teststatistik: $Z := \frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}}$
5. Unter H_0 gilt $Z \sim \mathcal{N}(0,1)$. (Wir erwarten in diesem Fall also einen Wert $Z(\omega)$ in der Nähe von 0.)
6. H_0 wird auf dem 5%-Niveau verworfen, falls $|Z| > 1.96$.

7. Mit unseren Daten beträgt der Wert der Teststatistik Z

$$Z(\omega) = \frac{\overline{x}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{20.44 - 20}{\frac{0.8}{\sqrt{9}}} = \frac{0.44}{\frac{0.8}{3}} = \frac{33}{20} = 1.65 < 1.96.$$

Das Gerät kann richtig geeicht sein. Die Nullhypothese wird nicht verworfen.

(c) (i) Vertrauensintervall beim t-Test: $[\overline{X}_n - t_{n-1}(\frac{1-\alpha}{2})\frac{S_x}{\sqrt{n}}, \overline{X}_n + t_{n-1}(\frac{1-\alpha}{2})\frac{S_x}{\sqrt{n}}]$
Mit $\overline{x}_n = 20.44$, $t_8(0.025) = 2.306$, $s_x = 0.88$ und $n = 9$ ergibt dies das realisierte Vertrauensintervall $[20.44 - 2.306\,\frac{0.88}{3}, 20.44 + 2.306\,\frac{0.88}{3}] \approx [20.44 - 0.77 \cdot 0.88, 20.44 + 0.77 \cdot 0.88] \approx [19.76, 21.12]$.

(ii) Vertrauensintervall beim z-Test: $[\overline{X}_n - z(\frac{1-\alpha}{2})\frac{\sigma}{\sqrt{n}}, \overline{X}_n + z(\frac{1-\alpha}{2})\frac{\sigma}{\sqrt{n}}]$
Mit $\overline{x}_n = 20.44$, $z(0.025) = 1.96$, $\sigma = 0.800$ und $n = 9$ ergibt dies das realisierte Vertrauensintervall $[20.44 - 1.96\,\frac{0.8}{3}, 20.44 + 1.96\,\frac{0.8}{3}] \approx [20.44 - 0.65 \cdot 0.8, 20.44 + 0.65 \cdot 0.8] \approx [19.91, 20.96]$

9. (10 points) Each of a group of 20 tennis players gets two tennis racquets for testing. In each pair of racquets, one has nylon strings and the other synthetic gut strings. After ten weeks of testing, each player is asked whether they want to use nylon or gut strings. We call X the number of players choosing gut and p the population fraction of players who play better with gut strings. As gut strings are more expensive, it seems plausible that players only will choose gut if they really notice a substantial difference in their playing level.

   In order to test statistically whether gut strings are better, we use the null hypothesis $H_0 : p = 0.5$, and we only reject H_0 if the outcome of our experiment is clearly in favour of gut strings.

   (a) (3 points) Which of the two rejection regions $K = \{15, 16, 17, 18, 19, 20\}$ and $K' = \{0, 1, 2, 3, 4, 5\}$ is better suited for this test? Formulate your assumptions and explain your choice.

   (b) (3 points) What is the probability of an error of the first kind for the region you have chosen in (a)? Does that region yield a test at the significance level $\alpha = 0.05$? Is this the best test at the level α (and if so, in which sense)?

   (c) (2 points) Compute the probability of an error of the second kind for the region you have chosen in (a), for the two cases $p = p_1 = 0.6$ and $p = p_2 = 0.8$.

   (d) (2 points) Compute, for the rejection region you have chosen in (a), the realised value of the P-value for the realised value $x = 13$ of X. Do we reject the null hypothesis at a significance level of $\alpha' = 0.10$?

   Hint: A table with values of the distribution function of a $\text{Bin}(n, p)$-distributed random variable for $n = 20$ and different p can be found with other statistical tables at the end of the exam.

Solution:

(a) Wir nehmen an, dass die Präferenzen der einzelnen Tennisspieler unabhängig voneinander und identisch verteilt sind. Also ist X binomialverteilt mit $n = 20$ und p. Die Nullhypothese und Alternative sind

$$H_0 : p = 0.5 \quad H_A : p > 0.5.$$

Wir suchen also einen Verwerfungsbereich, der die Nullhypothese verwirft, falls viel mehr als die Hälfte der Tennisspieler Darm-Saiten bevorzugen, also falls X gross ist. Der einzige Verwerfungsbereich, der dies erfüllt, ist der erste, also $K = \{15, 16, 17, 18, 19, 20\}$.

(b) Die Wahrscheinlichkeit eines Fehlers erster Art, also die Nullhypothese abzulehnen, obwohl sie stimmt, ist

$$\mathbb{P}_{H_0}[X \geq 15] = \sum_{k=15}^{20} \binom{20}{k} 0.5^k 0.5^{(20-k)} = 1 - \mathbb{P}_{p=\frac{1}{2}}[X \leq 14] \approx 1 - 0.9793 = 0.0207.$$

Alternativ kann man auch mit der Symmetrie der Binomialverteilung für $p = \frac{1}{2}$ argumentieren,

$$\mathbb{P}_{H_0}[X \geq 15] = \sum_{k=15}^{20} \binom{20}{k} 0.5^k 0.5^{(20-k)} = \mathbb{P}_{p=\frac{1}{2}}[X \leq 5] \approx 0.0207.$$

Also liefert der in (a) gewählte Verwerfungsbereich einen Test zum Niveau $\alpha = 0.05 \geq 0.0207$. Wir erhalten

$$\mathbb{P}_{H_0}[X \geq 14] = 1 - \mathbb{P}_{p=\frac{1}{2}}[X \leq 13] \approx 1 - 0.9423 = 0.0577 > 0.05,$$

(alternativ mit Sym. $\mathbb{P}_{H_0}[X \geq 14] = \mathbb{P}_{p=\frac{1}{2}}[X \leq 6] \approx 0.0577 > 0.05$,) also ist der gewählte Verwerfungsbereich der grösstmögliche zum Niveau α.

(c) Ein Fehler zweiter Art passiert, wenn wir die Nullhypothese nicht verwerfen, obwohl sie falsch ist, also

$$\mathbb{P}_{p=p_1=0.6}[X \leq 14] \approx 0.8744 \quad \text{und} \quad \mathbb{P}_{p=p_2=0.8}[X \leq 14] \approx 0.1958.$$

(d) Der P-Wert ist

$$\mathbb{P}_{H_0}[X \geq 13] = 1 - \mathbb{P}_{p=\frac{1}{2}}[X \leq 12] \approx 1 - 0.8684 = 0.1316.$$

Alternativ mit Symmetrie,

$$\mathbb{P}_{H_0}[X \geq 13] = \mathbb{P}_{p=\frac{1}{2}}[X \leq 7] \approx 0.1316.$$

Die Nullhypothese wird also auf dem Signifikanzniveau 0.10 nicht verworfen.

10. (10 points) Consider a series of coin tosses modelled by a sequence $X_1, X_2, \ldots$ of i.i.d. random variables with $\mathbb{P}[X_n = 1] = \mathbb{P}[X_n = -1] = \frac{1}{2}$. Here $X_n = 1$ signifies heads and $X_n = -1$ tails. Before each toss n, you have the opportunity to bet a (possibly random) amount V_n. If the coin comes up heads ($X_n = 1$), you win V_n, otherwise you lose V_n. Therefore the total gain after n rounds is

$$G_n = \sum_{k=1}^{n} V_k X_k.$$

Note that G_n may be negative, which signifies a loss.

Suppose you use the following strategy: Set $V_1 = 1$. For $k \geq 1$, if $X_k = 1$ you stop, meaning that you set $V_n = 0$ for $n \geq k+1$. If $X_k = -1$ and you have not yet stopped, you double your bet, that is, you set $V_{k+1} = 2\,V_k$.

Furthermore, let T denote the time you stop, $T = \inf\{k \geq 1 : X_k = 1\}$, with $T = \infty$ if $X_k = -1$ for all k.

(a) (2 points) Compute G_n for $1 \leq n < T$.

(b) (1 point) Compute G_T for $T < \infty$.

(c) (2 points) Compute $\mathbb{P}[T < \infty]$ with a rigorous argument.

(d) (2 points) Compute the expected maximal intermediate shortfall. That is, compute $\mathbb{E}\left[\min_{1 \leq n \leq T} G_n\right]$.

In reality, you would face a limit on credit. Specifically, assume you are only allowed to use strategies $(V_n)_{n \geq 1}$ such that the gain satisfies $G_n \geq -b$ almost surely for all n, where $b \in (1, \infty)$ is some fixed constant. Modify the above strategy so that you stop as soon as there is a risk of violating the bound. That is, specify V_k as before, except that if you have not yet stopped at k and $G_k - 2\,V_k < -b$, so that the bound would be violated if you lose with the new bet $V_{k+1} = 2V_k$, then set $V_n = 0$ for all $n \geq k + 1$.

Let T' denote the time you stop when using this modified strategy.

(e) (3 points) Compute your expected gain $\mathbb{E}[G_{T'}]$ under these new rules. How does this depend on the bound b?

Solution:

(a) Note that $(X_1, \ldots, X_{T-1}, X_T) = (-1, \ldots, -1, 1)$ for $T > 1$, and that $V_k = 2^{k-1}$ for all $k < T$. Thus, for $n < T$,

$$G_n = \sum_{k=1}^{n} 2^{k-1} \times (-1) = -\sum_{k=0}^{n-1} 2^k = 1 - 2^n.$$

(b) Since $X_T = 1$ and $V_T = 2^{T-1}$, we have from part (a), if $T > 1$,

$$G_T = G_{T-1} + V_T X_T = 1 - 2^{T-1} + 2^{T-1} = 1.$$

If $T = 1$, then clearly $G_T = 1 \times 1 = 1$. So in all cases, $G_T = 1$.

(c) Since $\{T = \infty\} = \{X_n = -1 \text{ for all } n \geq 1\} = \bigcap_{n\geq 1}\{X_n = -1\}$, we have

$$\mathbb{P}[T = \infty] = \lim_{n\to\infty} \mathbb{P}\left[\bigcap_{k=1}^{n}\{X_k = -1\}\right] = \lim_{n\to\infty} \prod_{k=1}^{n} \mathbb{P}[X_k = -1] = \lim_{n\to\infty} 2^{-n} = 0,$$

using the continuity property of probability measures as well as the independence of $(X_n)_{n\geq 1}$. Thus $\mathbb{P}[T < \infty] = 1$.

(d) From parts (a) and (b) we have $G_n = 1 - 2^n$ for $n < T$, and $G_T = 1$. Thus

$$\min_{1\leq n\leq T} G_n = \begin{cases} 1 - 2^{T-1} & \text{if } T > 1 \\ 1 & \text{if } T = 1. \end{cases}$$

Also, for $n \geq 1$,

$$\mathbb{P}[T = n] = \mathbb{P}[X_1 = \cdots = X_{n-1} = -1 \text{ and } X_n = 1] = \prod_{k=1}^{n} \frac{1}{2} = 2^{-n}.$$

Thus

$$\mathbb{E}\left[\min_{1\leq n\leq T} G_n\right] = 1 \times \frac{1}{2} + \sum_{n=2}^{\infty}(1 - 2^{n-1}) \times 2^{-n} = \frac{1}{2} + \sum_{n=2}^{\infty}(2^{-n} - 2^{-1}) = -\infty.$$

(e) For $1 \leq n < T$, we have $G_n - 2V_n = 1 - 2^n - 2 \times 2^{n-1} = 1 - 2^{n+1}$. Thus

$$G_n - 2V_n < -b \qquad \Longleftrightarrow \qquad n > \log(1 + b) - 1.$$

Let n_0 denote the smallest n for which this happens. Then $T' = T$ on the event $\{T \leq n_0\}$, and $T' = n_0$ on the event $\{T > n_0\}$.

There are at least possible two ways to continue.

Method 1: For this method we only need that n_0 is a finite deterministic number, but no explicit formula. One can directly see that

$$G_{T'} = (V \bullet S)_{n_0},$$

where S is the random walk $(S_n)_{1\leq n\leq n_0} = (\sum_{i=1}^{n} X_i)_{1\leq n\leq n_0}$.

Since $V_k(\omega)$ can be expressed as $V_k(\omega) = \varphi_k(X_1(\omega), \ldots X_{k-1}(\omega)$ with a function $\varphi_k : \{-1, +1\}^{k-1} \to \mathbb{R}$, we know that $V = (V_n)_{1\leq n\leq n_0}$ is a gambling system (see definition 1.3.13 in the lecture notes).

Therefore we can conclude by Theorem 1.3.17 that $\mathbb{E}[(V \bullet S)_{n_0}] = 0$ and thus $\mathbb{E}[G_{T'}] = 0$. Note that this argument could not be made for T, because T is not bounded (so we could not use a finite deterministic number n_0).

Method 2:

Thus,

$$\begin{aligned} \mathbb{E}[G_{T'}] &= \mathbb{E}\left[G_T 1_{\{T\leq n_0\}} + G_{n_0} 1_{\{T>n_0\}}\right] \\ &= 1 \times \mathbb{P}[T \leq n_0] + (1 - 2^{n_0}) \times \mathbb{P}[T > n_0] \\ &= 1 \times (1 - \mathbb{P}[T > n_0]) + (1 - 2^{n_0}) \times \mathbb{P}[T > n_0] \\ &= 1 - 2^{n_0}\mathbb{P}[T > n_0] \\ &= 0, \end{aligned}$$

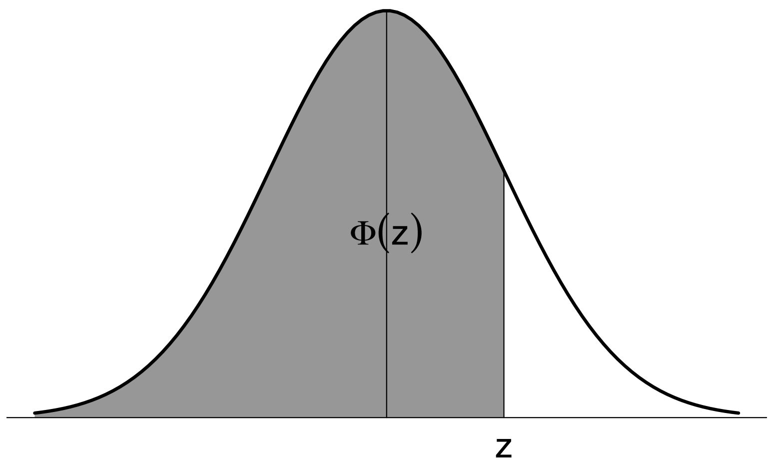
using that $\mathbb{P}[T > n_0] = \mathbb{P}[X_1 = \cdots = X_{n_0} = -1] = 2^{-n_0}$.

Note that the result (which is of course the same with both methods) is completely independent of the actual value of the bound b!

⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆

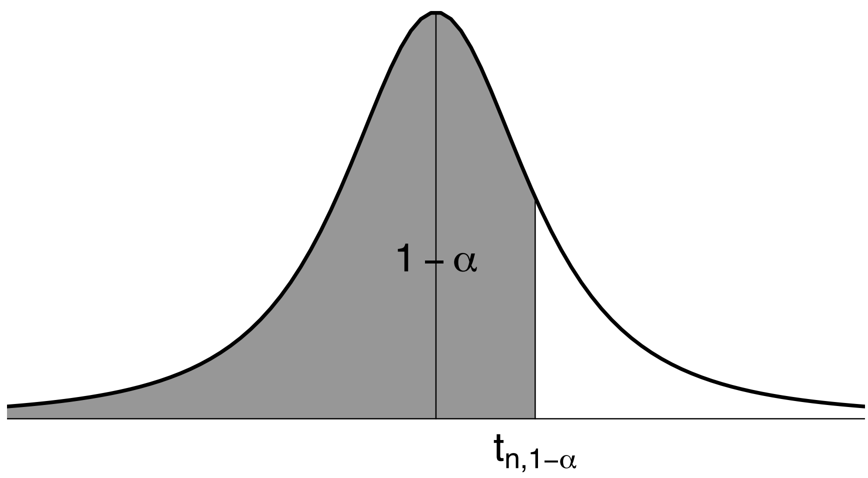
⋆⋆⋆ Good luck! ⋆⋆⋆

Tabellen



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

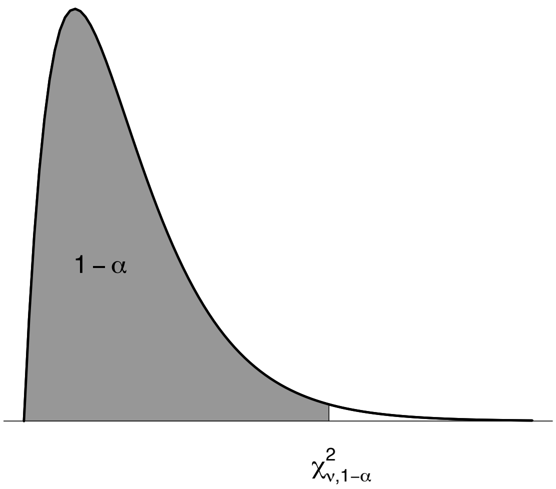
Tabelle der Standard-Normalverteilungsfunktion $\Phi(z) = P[Z \leq z]$ mit $Z \sim \mathcal{N}(0,1)$



1 − α
t_{n,1−α}

df	$t_{0.60}$	$t_{0.70}$	$t_{0.80}$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750
31	0.255	0.530	0.853	1.309	1.696	2.040	2.452	2.744
32	0.255	0.530	0.853	1.309	1.694	2.037	2.449	2.738
33	0.255	0.530	0.853	1.308	1.693	2.035	2.445	2.733
34	0.255	0.529	0.852	1.307	1.691	2.032	2.441	2.728
35	0.255	0.529	0.852	1.306	1.690	2.030	2.438	2.724
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576

Ausgewählte Quantile $t_{n,1-\alpha}$ der t-Verteilung; in der Tabelle ist $n = df$.
Für $df = \infty$ erhält man die Quantile $z_{1-\alpha}$ der Standard-Normalverteilung.



	$p = 0.90$	$p = 0.95$	$p = 0.975$	$p = 0.999$	$p = 0.9995$
$\nu = 1$	2.7055	3.8415	5.0239	10.8276	12.1157
$\nu = 2$	4.6052	5.9915	7.3778	13.8155	15.2018
$\nu = 3$	6.2514	7.8147	9.3484	16.2662	17.7300
$\nu = 4$	7.7794	9.4877	11.1433	18.4668	19.9974
$\nu = 5$	9.2364	11.0705	12.8325	20.5150	22.1053
$\nu = 6$	10.6446	12.5916	14.4494	22.4577	24.1028
$\nu = 7$	12.0170	14.0671	16.0128	24.3219	26.0178
$\nu = 8$	13.3616	15.5073	17.5345	26.1245	27.8680
$\nu = 9$	14.6837	16.9190	19.0228	27.8772	29.6658
$\nu = 10$	15.9872	18.3070	20.4832	29.5883	31.4198
$\nu = 11$	17.2750	19.6751	21.9200	31.2641	33.1366
$\nu = 12$	18.5493	21.0261	23.3367	32.9095	34.8213

Ausgewählte Quantile $\chi^2_{\nu,1-\alpha}$ der Chiquadrat-Verteilung; in der Tabelle ist $p = 1 - \alpha$.

n	k | p	0.5	0.6	0.7	0.8	0.9
20	0	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0002	0.0000	0.0000	0.0000	0.0000
	3	0.0013	0.0001	0.0000	0.0000	0.0000
	4	0.0059	0.0003	0.0000	0.0000	0.0000
	5	0.0207	0.0016	0.0000	0.0000	0.0000
	6	0.0577	0.0065	0.0003	0.0000	0.0000
	7	0.1316	0.0210	0.0013	0.0000	0.0000
	8	0.2517	0.0565	0.0051	0.0001	0.0000
	9	0.4119	0.1275	0.0171	0.0006	0.0000
	10	0.5881	0.2447	0.0480	0.0026	0.0000
	11	0.7483	0.4044	0.1133	0.0100	0.0001
	12	0.8684	0.5841	0.2277	0.0321	0.0004
	13	0.9423	0.7500	0.3920	0.0867	0.0024
	14	0.9793	0.8744	0.5836	0.1958	0.0113
	15	0.9941	0.9491	0.7625	0.3704	0.0432
	16	0.9987	0.9840	0.8929	0.5886	0.1330
	17	0.9998	0.9964	0.9645	0.7939	0.3231
	18	1.0000	0.9995	0.9924	0.9308	0.6083
	19	1.0000	1.0000	0.9992	0.9885	0.8784
	20	1.0000	1.0000	1.0000	1.0000	1.0000

The above is a table of the (cumulative) distribution function of the binomial distribution with $n = 20$. For all pairs of p and k, the table gives the value

$$\mathbb{P}[Y \leq k] \qquad \text{for } Y \sim \text{Bin}(n = 20, p).$$