WuS Exam Notes

Ruben Schenk

August 2021

1 Wahrscheinlichkeit

1.1 Grundbegriffe

Def: Der **Ereignisraum** $\Omega \neq \emptyset$ ist die Menge aller möglichen Ergebnisse eines Zufallsexperiment. Die Element $\omega \in \Omega$ heissen **Elementarereignisse** des Experiments.

Def: Die **Potenzmenge** von Ω, $\mathcal{P}(\Omega)$ oder 2^Ω, ist die Menge aller Teilmengen von Ω. Ein **prinzipielles Ereignis** ist eine Teilmenge $A \subseteq \Omega$. Die Klasse aller beobachtbaren Ereignisse ist $\mathcal{F}$.

Def: Ein **Wahrscheinlichkeitsmass** ist eine Abbildung $\mathrm{P} : \mathcal{F} \to [0, 1]$, welche die folgenden Axiome erfüllt. Für $A \in \mathcal{F}$ nennen wir $\mathrm{P}[A] \in [0, 1]$ die **Wahrscheinlichkeit**, kurz ***WS***, dass A eintritt. Die geforderten Axiome sind:

1. $\mathrm{P}[A] \geq 0$ für alle Ereignisse $A \in \mathcal{F}$
2. $\mathrm{P}[\Omega] = 1$
3. $\mathrm{P}[\dot{\bigcup}_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} \mathrm{P}[A_i]$, sofern die $A_i \in \mathcal{F}$ paarweise disjunkt sind.

Bmk: $\dot{\bigcup}$ steht für die Vereinigung paarweiser disjunkten Mengen.
Bmk: Aus den obigen Axiomen lassen sich einige grundlegende **Rechenregeln** ableiten:

- $\mathrm{P}[A^C] = 1 - \mathrm{P}[A]$
- $\mathrm{P}[\emptyset] = 0$
- Für $A \subseteq B$ gilt $\mathrm{P}[A] \leq \mathrm{P}[B]$
- Für beliebige, nicht unbedingt disjunkte, A, B gilt die allgemeine **Additionsregel:** $\mathrm{P}[A \cup B] = \mathrm{P}[A] + \mathrm{P}[B] - \mathrm{P}[A \cap B]$.

1.2 Diskrete Wahrscheinlichkeit

Def: Ist $\Omega = \{\omega_1, \omega_2, ..., \omega_N\}$ endlich mit $|\Omega| = N$ und $\mathcal{F} = 2^\Omega$, und sind $\omega_1, \omega_2, ..., \omega_N$ alle gleich wahrscheinlich, also $p_1 = p_2 = \cdots = p_N = \frac{1}{N}$, so heisst Ω ein **Laplace-Raum** und P die **diskrete Gleichverteilung** auf Ω. Für beliebige $A \subseteq \Omega$ ist

$$\mathrm{P}[A] = \frac{\text{Anzahl der Elementarereignisse in A}}{\text{Anzahl der Elementarereignisse in } \Omega} = \frac{|A|}{|\Omega|}.$$

1.3 Bedingte Wahrscheinlichkeit

Def: Seien A, B Ereignisse und $\mathrm{P}[A] > 0$. Die **bedingte Wahrscheinlichkeit** von B unter der Bedingung, dass A eintritt (kurz "*gegeben A*") wird definiert durch:

$$\mathrm{P}[B|A] := \frac{\mathrm{P}[B \cap A]}{\mathrm{P}[A]}.$$

Aus obiger Definition folgt direkt die sogenannte **Multiplikationsregel:** Für beliebige Ereignisse A, B ist $\mathrm{P}[A \cap B] = \mathrm{P}[B|A] \cdot \mathrm{P}[A]$.

Satz der totalen Wahrscheinlichkeit: Sei $A_1, ..., A_n$ eine Zerlegung von Ω (in paarweise disjunkte Ereignisse), d.h. $\dot{\bigcup}_{i=1}^{n} A_i = \Omega$. Für beliebige Ereignisse B gilt dann:

$$\mathrm{P}[B] = \sum_{i=1}^{n} \mathrm{P}[B|A_i] \cdot \mathrm{P}[A_i]$$

Satz (von Bayes): Für zwei Ereignisse A und B und $\mathrm{P}[B] \neq 0$ lautet der Satz von Bayes wie folgt:

$$\mathrm{P}[A|B] = \frac{\mathrm{P}[B|A] \cdot \mathrm{P}[A]}{\mathrm{P}[B]} = \frac{\mathrm{P}[B|A] \cdot \mathrm{P}[A]}{\mathrm{P}[B|A] \cdot \mathrm{P}[A] + \mathrm{P}[B|A^C] \cdot \mathrm{P}[A^C]}.$$

1.4 Unabhängigkeit von Ereignissen

Def: Zwei Ereignisse A, B heissen **stochastisch unabhängig**, falls

$$\mathrm{P}[A \cap B] = \mathrm{P}[A] \cdot \mathrm{P}[B].$$

Ist $\mathrm{P}[A] = 0$ oder $\mathrm{P}[B] = 0$, so sind A und B immer unabhängig. Für $P[A] \neq 0$ $(\mathrm{P}[B] \neq 0)$ gilt:

$$A,\, B \text{ unabhängig} \iff \mathrm{P}[B|A] = P[B] \quad (\mathrm{P}[A|B] = \mathrm{P}[A]).$$

Def: Die Ereignisse $A_1, ..., A_n$ heissen **stochastisch unabhängig**, wenn für jede endliche Teilfamilie die Produktformel gilt, d.h. für $m \in \mathcal{N}$ und $\{k_1, ..., k_m\} \subseteq \{1, ..., n\}$ gilt immer:

$$\mathrm{P}\left[\bigcap_{i=1}^{m} A_{k_i}\right] = \prod_{i=1}^{m} \mathrm{P}[A_{k_i}]$$

2 Diskrete Zufallsvariablen und Verteilungen

2.1 Grundbegriffe

Def: Eine **diskrete Zufallsvariable**, kurz ***ZV***, auf Ω ist eine Funktion $X : \Omega \to \mathbb{R}$. Die **Verteilung** von X ist gegeben durch:

$$\mu_X(B) := \mathrm{P}[X \in B] := \mathrm{P}[\{\omega \in \Omega : X(\omega) \in B\}] \quad \forall B \subseteq \mathcal{W}(X).$$

Die **Verteilungsfunktion**, kurz ***VF***, von X ist die Abbildung $F_X : \mathbb{R} \to [0, 1]$, definiert durch

$$t \to F_X(t) := \mathrm{P}[X \leq t] := \mathrm{P}[\{\omega : X(\omega) \leq t\}]$$

Def: Die **Gewichtsfunktion** oder **diskrete Dichte**, von X ist die Funktion $p_X : \mathcal{W}(X) \to [0, 1]$, die definiert ist durch

$$p_X(x_k) := \mathrm{P}[X = x_k] = \mathrm{P}[\{\omega : X(\omega) = x_k\}] \quad (\text{für } k = 1, 2, ...).$$

Bmk: Ist Ω endlich oder abzählbar mit $\mathcal{F} = 2^\Omega$, so ist *jede* Funktion $X : \Omega \to \mathbb{R}$ eine diskrete Zufallsvariable.

Def: Für jede Teilmenge $A \subseteq \Omega$ ist die **Indikatorfunktion** I_A von A definiert durch:

$$I_A(\omega) := \begin{cases} 1 & \text{für } \omega \in A \\ 0 & \text{für } \omega \in A^C \end{cases}$$

Eigenschaften der Dichte und Verteilungsfunktion:

- die Verteilungsfunktion F_X ist vollständig durch die Dichte p_X festgelegt: $F_X(t) = \mathrm{P}[X \leq t] = \sum_{x_k \leq t} \mathrm{P}[X = x_k] = \sum_{x_k \leq t} p_X(x_k)$
- Die Verteilung beschreibt das stochastische Verhalten einer Zufallsvariable. Das ist das WHS-Mass μ_X, das durch $\mu_X(B) := \mathrm{P}[X \in B]$ definiert ist. Ist X eine diskrete ZV, so ist μ_X die **diskrete Verteilung**. Damit kann man die Verteilung μ_X und die Gewichtsfunktion p_X miteinander identifizieren: der Unterschied besteht darin, dass μ_X als Argumente Teilmengen von $\mathcal{W}(X)$ hat, p_X hingegen Elemente von $\mathcal{W}(X)$. Es gilt:

$$\mu_X(B) = \mathrm{P}[X \in B] = \sum_{x_k \in B} p_X(x_k) \quad \text{für } B \subseteq \mathcal{W}(X)$$

2.2 Erwartungswerte

Def: Sei X eine diskrete Zufallsvariable mit Gewichtsfunktion p_X. Dann definieren wir den **Erwartungswert** von X als

$$\mathrm{E}[X] := \sum_{x_k \in \mathcal{W}(X)} x_k p_X(x_k),$$

sofern die Reihe absolut konvergiert ($\sum_{x_k \in \mathcal{W}(X)} |x_k| p_X(x_k) < \infty$). Ist die Reihe nicht konvergent, so existiert der Erwartungswert nicht. Des weiteren gilt:

$$\mathrm{E}[X] = \sum_{\omega_i \in \Omega} X(\omega_i) \mathrm{P}[\{\omega_i\}] = \sum_{\omega_i \in \Omega} p_i X(\omega_i).$$

Satz: Sei X eine diskrete ZV mit Gewichtsfunktion p_X, und sei $Y = g(X)$ für eine Funktion $g : \mathbb{R} \to \mathbb{R}$. Sofern die Reihe absolut konvergiert gilt:

$$\mathrm{E}[Y] = \mathrm{E}[g(X)] = \sum_{x_k \in \mathcal{W}(X)} g(x_k) p_X(x_k).$$

Satz: Seien X und Y diskrete ZV's, für die jeweils der Erwartungswert existiert. Dann gilt:

1. **Monotonie:** Ist $X \leq Y$ (d.h. $X(\omega) \leq Y(\omega)$ für alle ω), so gilt auch $\mathrm{E}[X] \leq \mathrm{E}[Y]$.
2. **Linearität:** Für beliebige $a, b \in \mathbb{R}$ gilt $\mathrm{E}[aX + b] = a\mathrm{E}[X] + b$.
3. Falls X nur Werte in $\mathbb{N}_0$ annimmt, so gilt

$$\mathrm{E}[X] = \sum_{j=1}^{\infty} \mathrm{P}[X \geq j] = \sum_{l=0}^{\infty} \mathrm{P}[X > l]$$

Def: Sei X eine diskrete Zufallsvariable. Ist $\mathrm{E}[X^2] < \infty$, so heisst

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

die **Varianz** von X, und $\sqrt{\mathrm{Var}[X]}$ heisst die **Standardabweichung** von X. Manchmal schreibt man auch $\mathrm{sd}(X) := \sqrt{\mathrm{Var}[X]}$ oder $\sigma(X) := \sqrt{\mathrm{Var}[X]}$.

Lemma: Sei X eine diskrete ZV mit $\mathrm{E}[X]^2 < \infty$, und sei $Y = aX + b$. Dann gilt:

1. $\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$
2. $\mathrm{Var}[Y] = \mathrm{Var}[aX + b] = a^2\mathrm{Var}[X]$

2.3 Gemeinsame Verteilungen, unabhängige Zufallsvariablen

Def: Seien $X_1, ..., X_n$ (diskrete oder auch beliebige) Zufallsvariablen. Die **gemeinsame Verteilungsfunktion** von $X_1, ..., X_n$ ist die Abbildung $F : \mathbb{R}^n \to [0, 1]$, definiert durch

$$(x_1, ..., x_n) \to F(x_1, ..., x_n) := \mathrm{P}[X_1 \leq x_1, ..., X_n \leq x_n].$$

Sind $X_1, ..., X_n$ diskrete ZV's, so definiert man ihre **gemeinsame Gewichtsfunktion** $p : \mathbb{R}^n \to [0, 1]$ durch

$$p(x_1, ..., x_n) := \mathrm{P}[X_1 = x_1, ..., X_n = x_n].$$

Def: Haben X und Y die gemeinsame Verteilungsfunktion F, so ist die Funktion $F_X : \mathbb{R} \to [0, 1]$,

$$x \to F_X(x) := \mathrm{P}[X \leq x] = \mathrm{P}[X \leq x, Y < \infty] = \lim_{y\to\infty} F(x, y)$$

die Verteilungsfunktion der **Randverteilung**, kurz ***RV***, von X. Analog ist $F_Y : \mathbb{R} \to [0, 1]$,

$$y \to F_Y(y) := \mathrm{P}[Y \leq y] = \mathrm{P}[X < \infty, Y \leq y] = \lim_{x\to\infty} F(x, y)$$

die Verteilungsfunktion der Randverteilung von Y.

Def: Sind X und Y diskrete Zufallsvariablen mit $\mathcal{W}(Y) = \{y_1, y_2, ...\}$ und gemeinsamer Gewichtsfunktion $p(x, y)$, so ist die **Gewichtsfunktion der Randverteilung** von X gegeben durch $p_X : \mathcal{W}(X) \to [0, 1]$,

$$\begin{aligned} x \to p_X(x) = \mathrm{P}[X = x] &= \sum_{y_j\in\mathcal{W}(Y)} \mathrm{P}[X = x, Y = y_j] \\ &= \sum_{y_j\in\mathcal{W}(Y)} p(x, \ , y_j) \quad \text{für } x \in \mathcal{W}(X). \end{aligned}$$

Def: Die Zufallsvariablen $X_1, ..., X_n$ heissen **unabhängig**, falls gilt

$$F(x_1, ..., x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n),$$

d.h. die gemeinsame Verteilungsfunktion F ist das Produkt der Verteilungsfunktionen F_{X_i} der Randverteilungen.

2.4 Funktionen von mehreren Zufallsvariablen

Satz: Seien $X_1, ..., X_n$ diskrete ZV's mit endlichen Erwartungswert $\mathrm{E}[X_1], ..., \mathrm{E}[X_n]$. Sei $Y = a + \sum_{l=1}^{n} b_l X_l$ mit Konstanten $a, b_1, ..., b_n$. Dann gilt:

$$\mathrm{E}[Y] = a + \sum_{l=1}^{n} b_l \mathrm{E}[X_l].$$

Def: Die Grösse

$$\begin{aligned} \mathrm{Cov}(X_1, X_2) &:= \mathrm{E}[X_1X_2] - \mathrm{E}[X_1]\mathrm{E}[X_2] \\ &= E[(X_1 - \mathrm{E}[X_1])(X_2 - \mathrm{E}[X_2])] \end{aligned}$$

heisst die **Kovarianz** von X_1 und X_2, und damit haben wir allgemein die Formel:

$$\mathrm{Var}[X_1 + X_2] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2] + 2\mathrm{Cov}(X_1, X_2).$$

Def: Die **Korrelation** von X und Y ist definiert durch

$$\rho(X, Y) := \begin{cases} \frac{\mathrm{Cov}(X, Y)}{\sigma(X)\sigma(Y)}, & \text{falls } \sigma(X)\sigma(Y) > 0 \\ 0, & \text{sonst} \end{cases}$$

Ist $\mathrm{Cov}(X, Y) = 0$, so nennt man X und Y **unkorreliert.**

Satz: Seien $X_1, ..., X_n$ diskrete Zufallsvariablen mit endlichen Erwartungswerten $\mathrm{E}[X_1], ..., \mathrm{E}[X_n]$. Falls $X_1, ..., X_n$ unabhängig sind, so ist

$$\mathrm{E}\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} \mathrm{E}[X_i].$$

Insbesondere sind dann $X_1, ..., X_n$ paarweise unkorreliert. und es gilt

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i].$$

sofern $\mathrm{Var}[X_1], ..., \mathrm{Var}[X_n]$ alle existieren und endlich sind.
Bmk: Es gelten die Implikationen

$$\text{unabhängig} \implies \text{paarweise unabhängig} \implies \text{unkorreliert}$$

Beide umgekehrten Implikationen sind in der Regel falsch.

2.5 Bedingte Verteilungen

Def: Seien X und Y diskrete Zufallsvariablen mit gemeinsamer Gewichtsfunktion $p(x, y)$. Die **bedingte Gewichtsfunktion** von X, gegeben dass $Y = y$, ist definiert durch

$$p_{X|Y}(x|y) := \mathrm{P}[X = x|Y = y] = \frac{\mathrm{P}[X = x, Y = y]}{\mathrm{P}[Y = y]} = \frac{p(x, y)}{p_Y(y)}$$

für $p_Y(y) > 0$ und 0 sonst.

3 Wichtige diskrete Verteilungen

3.1 Diskrete Gleichverteilung

Def: Die **diskrete Gleichverteilung** auf einer endlichen Menge $\mathcal{W} = \{x_1, ..., x_N\}$ gehört zu einer Zufallsvariable X mit Wertebereich $\mathcal{W}$ und Gewichtsfunktion

$$p_X(x_k) = \mathrm{P}[X = x_k] = \frac{1}{N} \quad \text{für } k = 1, ..., N.$$

Das typische ist hier also, dass alle Elementarereignisse mit der gleichen Wahrscheinlichkeit auftreten. Beispiele: Ergebnis eines Würfelwurfs oder Münzwurfs, zufällige Auswahl einer dreistelligen Zahl, oder zufällige Permutationen von $\{1, ..., N\}$.

3.2 Unabhängige 0-1-Experimente

Wir betrachten eine Folge gleichartiger Experimente die nur in Erfolg oder Misserfolg enden können (z.B. das Würfeln einer 6, das Werfen von Kopf, usw.). Wir betrachten die Ereignisse

$$A_i = \{\text{Erfolg beim } i\text{-ten Experiment}\}.$$

Wir nehmen dabei an, dass (1) die A_i unabhängig sind und (2) dass $\mathrm{P}[A_i] = p$ für alle i ist. Dabei heisst $p \in [0, 1]$ der **Erfolgsparameter**. Mit $Y_i = I_{A_i}$ bezeichnen wir die **Indikatorfunktion** des Ereignisses A_i, also ist:

$$Y_i(\omega) = \begin{cases} 1 & \text{für } \omega \in A_i, \\ 0 & \text{für } \omega \notin A_i, \end{cases}$$

und nimmt nur die Werte 0 oder 1 an, mit $\mathrm{P}[Y_i = 1] = \mathrm{P}[A_i] = p$. Die Zufallsvariablen Y_i bilden daher eine **Folge unabhängiger 0-1-Experimente** mit Erfolgsparameter p.

3.3 Bernoulli-Verteilung

Def: Bezeichnen wir mit X das Ergebnis eines einzigen 0-1-Experiment, so hat X eine **Bernoulli-Verteilung** mit Parameter p. Wir schreiben $X \sim Be(p)$. Es gilt:

- Wertebereich: $\mathcal{W}(X) = \{0, 1\}$
- Gewichtsfunktion: $p_X(x) = p^x(1-p)^{1-x}$ für $x \in \mathcal{W}(X)$
- Erwartungswert: $\mathrm{E}[X] = p$
- Varianz: $\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = p(1-p)$

3.4 Binomialverteilung

Def: Die **Binomialverteilung** mit Parametern n und p beschreibt die *Anzahl der Erfolge* bei n unabhängigen 0-1-Experimenten mit Erfolgsparameter p. Sei die Zufallsvariable X gegeben durch $X = \sum_{i=1}^n I_{A_i} = \sum_{i=1}^n Y_i$, dann hat diese eine Binomialverteilung. Wir schreiben $X \sim Bin(n, p)$. Es gilt:

- Wertebereich: $\mathcal{W}(X) = \{0, 1, 2, ..., n\}$
- Gewichtsfunktion: $p_X(k) = \mathrm{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$ für $k = 0, 1, ..., n$
- Erwartungswert: $\mathrm{E}[X] = \sum_{i=1}^n \mathrm{E}[Y_i] = np$
- Varianz: $\mathrm{Var}[X] = \sum_{i=1}^n \mathrm{Var}[Y_i] = np(1-p)$

3.5 Geometrische Verteilung

Def: Betrachten wir eine unendliche Folge von unabhängigen 0-1-Experimenten mit Erfolgsparameter p. Sei X die Wartezeit auf den ersten Erfolg, also

$$X = \inf\{i \in \mathbb{N} : A_i \text{ tritt ein}\} = \inf\{i \in \mathbb{N} : Y_i = 1\}$$

d.h. X ist der Index der ersten Eins in der zufälligen 0-1-Folge. Dann hat X eine **geometrische Verteilung** mit Parameter p. Wir schreiben $X \sim Geom(p)$. Es gilt:

- Wertebereich: $\mathcal{W}(X) = \mathbb{N}$
- Gewichtsfunktion: $p_X(k) = \mathrm{P}[X = k] = p(1-p)^{k-1}$ für $k = 1, 2, 3, ...$
- Erwartungswert: $\mathrm{E}[X] = \sum_{l=0}^\infty \mathrm{P}[X > l] = \sum_{l=0}^\infty (1-p)^l = \frac{1}{1-(1-p)} = \frac{1}{p}$
- Varianz: $\mathrm{Var}[X] = \mathrm{E}[X(X-1)] + \mathrm{E}[X] - (\mathrm{E}[X])^2 = \frac{1-p}{p^2}$

3.6 Negativbinomiale Verteilung

Def: Betrachten wir eine unendliche Folge von unabhängigen 0-1-Experimenten mit Erfolgsparameter p, so können wir für $r \in \mathbb{N}$ auch die Wartezeit auf den r-ten Erfolg betrachten. Dann hat X eine **negativbinomiale Verteilung** mit Parametern r und p. Wir schreiben kurz $X \sim NB(r, p)$. X lässt sich schreiben als:

$$X = \inf\left\{k \in \mathbb{N} : \sum_{i=1}^k I_{A_i} = r\right\} = \inf\left\{k \in \mathbb{N} : \sum_{i=1}^k Y_i = r\right\}.$$

Es gilt:

- Wertebereich: $\mathcal{W}(X) = \{r, r+1, r+2, ...\}$
- Gewichtsfunktion: $p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$ für $k = r, r+1, ...$
- Erwartungswert: $\mathrm{E}[X] = \sum_{i=1}^r \mathrm{E}[X_i] = \frac{r}{p}$
- Varianz: $\mathrm{Var}[X] = \sum_{i=1}^r \mathrm{Var}[X_i] = \frac{r(1-p)}{p^2}$

wobei die Zufallsvariablen $X_1, ..., X_r \sim Geom(p)$ und ihre Summe $X := X_1 + \cdots X_r \sim NB(r, p)$.

3.7 Hypergeometrische Verteilung

Def: In einer Urne seien n Gegenstände, davon r vom Typ 1 und $n - r$ vom Typ 2. Man zieht ohne Zurücklegen m der Gegenstände. Die Zufallsvariable X beschreibt die Anzahl der Gegenstände vom Typ 1 in dieser Stichprobe. Dann hat X eine **hypergeometrische Verteilung** mit Parametern $n \in \mathbb{N}$ und $m.r \in \{1, ..., n\}$. Es gilt:

- Wertebereich: $\mathcal{W}(X) = \{0, 1, ..., \min(m, r)\}$
- Gewichtsfunktion:

$$p_X(k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}} \quad \text{für } k \in \mathcal{W}(X)$$

3.8 Poisson-Verteilung

Def: Die **Poisson-Verteilung** mit Parameter $\lambda \in (0, \infty)$ ist eine Verteilung auf der Menge $\mathbb{N}_0$ mit Gewichtsfunktion

$$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{für } k = 0, 1, 2, ...$$

Wir schreiben kurz $X \sim \mathcal{P}(\lambda)$. Die Poissonverteilung eignet sich zur Modellierung von seltenen Ereignissen. Sei z.B. X_n für jedes n eine Zufallsvariable mit $X_n \sim Bin(n, p_n)$ und $np_n = \lambda$. Lassen wir nun $n \to \infty$ gehen, so geht $p_n = \frac{\lambda}{n} \to 0$, und X_n beschreibt die Anzahl Erfolge bei sehr vielen Versuchen mit sehr kleiner Erfolgswahrscheinlichkeit, also bei seltenen Ereignissen. In diesem Sinn ist also die Poisson-Verteilung ein Grenzwert von Binomialverteilungen bei geeigneter Skalierung der Parameter. Es gilt:

- Erwartungswert: $\mathrm{E}[X] = \sum_{k=0}^\infty k p_X(k) = \lambda \sum_{j=0}^\infty e^{-\lambda} \frac{\lambda^j}{j!} = \lambda$
- Varianz: $\mathrm{Var}[X] = \mathrm{E}[X(X-1)] + \mathrm{E}[X] - (\mathrm{E}[X])^2 = \lambda$

4 Allgemeine Zufallsvariablen

4.1 Grundbegriffe

Def: Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum, also Ω ein Grundraum, $\mathcal{F} \subseteq 2^\Omega$ eine σ-Algebra von beobachtbaren Ereignissen und P ein Wahrscheinlichkeitsmass auf $\mathcal{F}$. Eine **Zufallsvariable** (***ZV***) auf Ω ist eine messbare Funktion $X : \Omega \to \mathbb{R}$. Die **Verteilungsfunktion** (***VF***) von X ist dann eine Abbildung $F_X : \mathbb{R} \to [0, 1]$,

$$t \to F_X(t) := \mathrm{P}[X \leq t] := \mathrm{P}[\{\omega : X(\omega) \leq t\}]$$

Jede Verteilungsfunktion F_X hat die folgende Eigenschaften:

1. F_X ist wachsend und rechtsstetig. Das bedeutet, dass $F_X(s) \leq F_X(t)$ für $s \leq t$ gilt und $F_X(u) \to F_X(t)$ für $u \to t$ mit $u > t$.
2. $\lim_{t\to-\infty} F_X(t) = 0$, $\lim_{t\to+\infty} F_X(t) = 1$
3. Für *stetige Verteilungen* gilt: $\forall x : \mathrm{P}[X = x] = 0$

Das stochastische Verhalten einer Zufallsvariable X wird durch ihre **Verteilung** beschrieben. Das ist dasjenige Wahrscheinlichkeitsmass μ_X auf $\mathbb{R}$, definiert durch:

$$\mu_X(B) := \mathrm{P}[X \in B]$$

Def: Eine Zufallsvariable X mit Verteilungsfunktion $F_X(t) = \mathrm{P}[X \leq t]$ heisst **absolut stetig mit Dichte(funktion)** $f_x : \mathbb{R} \to [0, \infty)$, falls gilt

$$F_X(t) = \int_{-\infty}^t f_X(s)\,\mathrm{d}s \quad \text{für alle } t \in \mathbb{R}$$

Eine Dichtefunktion f_X hat, analog zur Gewichtsfunktion im diskreten Fall, stets folgende Eigenschaften:

1. $f_X \geq 0$, und $f_X = 0$ ausserhalb von $\mathcal{W}(X)$
2. $\int_{-\infty}^\infty f(s)\,\mathrm{d}s = 1$

4.2 Wichtige stetige Verteilungen

4.3 Gleichverteilung

Def: Die **Gleichverteilung** auf einem Intervall $[a, b]$ ist ein Modell für die zufällige Wahl eines Punktes in $[a, b]$. Für die zugehörige Zufallsvariable X gilt:

- Wertebereich: $\mathcal{W}(X) = [a, b]$
- Dichtefunktion:

$$f_X(t) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq t \leq b, \\ 0 & \text{sonst} \end{cases}$$

- Verteilungsfunktion:

$$F_X(t) = \begin{cases} 0 & \text{für } t < a, \\ \frac{t-a}{b-a} & \text{für } a \leq t \leq b, \\ 1 & \text{für } t > b \end{cases}$$

Wir schreiben kurz $X \sim \mathcal{U}(a, b)$, dabei steht $\mathcal{U}$ für uniform. Ein wichtiger Spezialfall ist $\mathcal{U}(0, 1)$.

4.4 Exponentialverteilung

Def: Die **Exponentialverteilung** mit Parameter $\lambda > 0$ ist das stetige Analogon der geometrischen Verteilung. Für die zugehörige Zufallsvariable X gilt:

- Wertebereich: $\mathcal{W}(X) = [0, \infty)$
- Dichte:

$$f_X(t) = \begin{cases} \lambda e^{-\lambda t} & \text{für } t \geq 0, \\ 0 & \text{für } t < 0 \end{cases}$$

Wir schreiben $X \sim Exp(\lambda)$.
Bmk: Wie die geometrische Verteilung hat die Exponentialverteilung die Eigenschaft der "**Gedächtnislosigkeit**" in dem Sinne, dass gilt

$$\mathrm{P}[X > t + s | X > s] = \mathrm{P}[X > t]$$

Berechnung: Für $X \sim Exp(\lambda)$:

$$\begin{aligned} \mathrm{P}[X > t + s \,|\, X > s] &= \frac{\mathrm{P}[X > t + s]}{\mathrm{P}[X > s]} = \frac{1 - F_X(t + s)}{1 - F_X(s)} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} \\ &= 1 - F_X(t) = \mathrm{P}[X > t] \end{aligned}$$

4.5 Normalverteilung

Def: Die **Normalverteilung**, auch kurz ***NV*** oder Gauss-Verteilung genannt, hat zwei Parameter $\mu \in \mathbb{R}$ und $\sigma^2 > 0$. Für die zugehörige Zufallsvariable X gilt:

- Wertebereich: $\mathcal{W}(X) = \mathbb{R}$
- Dichtefunktion:

$$f_X(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \quad \text{für } t \in \mathbb{R}$$

Wir schreiben $X \sim \mathcal{N}(\mu, \sigma^2)$.
Ein wichtiger Spezialfall ist die **Standard-Normalverteilung** mit $\mu = 0$, $\sigma^2 = 1$, also $\mathcal{N}(0, 1)$. Die Dichte wird mit $\varphi(t)$ und die Verteilungsfunktion mit $\Phi(t)$ bezeichnet. Weder für F_X noch für Φ gibt es eine geschlossenen Ausdruck, aber das Integral

$$\Phi(t) = \int_{\infty}^{t} \varphi(s)\,\mathrm{d}s = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}s^2}\,\mathrm{d}s$$

ist ***tabelliert***. Ist $X \sim \mathcal{N}(\mu, \sigma^2)$, so ist $F_X(t) = \Phi(\frac{t-\mu}{\sigma})$.

4.6 Erwartungswerte

Ist X stetig mit Dichte $f_X(x)$, so gilt

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x,$$

sofern das Integral absolut konvergiert, ansonsten existiert der Erwartungswert nicht.
Bsp: Bei einer Gleichverteilung auf $[a, b]$: Für $X \sim \mathcal{U}(a, b)$ ist

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x = \int_a^b x \frac{1}{b-a}\,\mathrm{d}x = \cdots = \frac{a+b}{2}$$

Bsp: Bei einer Normalverteilung: Für $X \sim \mathcal{N}(\mu, \sigma^2)$ ist

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}\,\mathrm{d}x = \cdots = \mu$$

Bsp: Cauchy-Verteilung: Eine **Cauchy-verteilte** Zufallsvariable X hat den Wertebereich $\mathcal{W}(X) = \mathbb{R}$ mit der Dichtefunktion $f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}$ für $x \in \mathbb{R}$. Die zugehörige Verteilungsfunktion ist $F_X(x) = \frac{1}{2} + \frac{1}{\pi}\arctan(x)$ für $x \in \mathbb{R}$. Es existiert kein Erwartungswert.

Satz: Sei X eine Zufallsvariable und $Y = g(X)$ eine weitere Zufallsvariable. Ist X steig mit Dichte f_X, so ist

$$\mathrm{E}[Y] = \mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x,$$

sofern das Integral absolut konvergiert.

4.7 Gemeinsame Verteilungen, unabhängige Zufallsvariablen

Def: Die **gemeinsame Verteilungsfunktion** von n Zufallsvariablen $X_1, .., X_n$ ist die Abbildung $F : \mathbb{R}^n \to [0, 1]$

$$(x_1, ..., x_n) \to F(x_1, ..., x_n) := \mathrm{P}[X \leq x_1, ..., X_n \leq x_n].$$

Falls die gemeinsame Verteilungsfunktion F von $X_1, ..., X_n$ sich schreiben lässt als

$$F(x_1, ..., x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1, ..., t_n)\,\mathrm{d}t_n \cdots \mathrm{d}t_1$$

für eine Funktion $f : \mathbb{R}^n \to [0, \infty)$, so heisst $f(x_1, ..., x_n)$ die **gemeinsame Dichte** von $X_1, ..., X_n$.

Def: Haben X und Y die gemeinsame Verteilungsfunktion F, so ist die Funktion $F_X : \mathbb{R} \to [0, 1]$,

$$x \to F_X(x) := \mathrm{P}[X \leq x] = \mathrm{P}[X \leq x, Y < \infty] = \lim_{y\to\infty} F(x, y)$$

die Verteilungsfunktion der **Randverteilung** (RV) von X.

Def: Die Zufallsvariablen $X_1, ..., X_n$ heissen **unabhängig**, falls gilt

$$F(x_1, ..., x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$$

d.h. die gemeinsame Verteilungsfunktion F ist das Produkt der Verteilungsfunktionen F_{X_i} der Randverteilungen.

4.8 Momente

Def: Für $n \in \mathbb{N}$ und falls $\mathrm{E}[|X|^n] < \infty$ ist, kann man das **n-te Zentralmoment** von X definieren als

$$\mathrm{E}[(X - \mathrm{E}[X])^n].$$

Def: Sei X eine Zufallsvariable. Die **absoluten Momente** von X sind definiert als $\mathrm{E}[|X|^p$ für $p > 0$. Die **Momente** von X sind definiert als $m_n := \mathrm{E}[X^n]$ für $n \in \mathbb{N}$, unter der Voraussetzung, dass $\mathrm{E}[|X|^n] < \infty$. Ist X diskret mit Gewichtsfunktion p_X und Wertebereich $\mathcal{W}(X)$, so gilt

$$\mathrm{E}[|X|^p] = \sum_{x\in\mathcal{W}(X)} |x|^p p_X(x).$$

Ist X absolutstetig mit Dichtefunktion f_X, so gilt

$$\mathrm{E}[|X|^p] = \int_{-\infty}^{\infty} |x|^p f_X(x)\,\mathrm{d}x.$$

4.9 Funktionen und Transformationen von Zufallsvariablen

Satz: Sei F eine stetige und streng monoton wachsende Verteilungsfunktion, mit Umkehrfunktion F^{-1}. Ist $X \sim \mathcal{U}(0, 1)$ und $Y = F^{-1}(X)$, so hat Y gerade die Verteilungsfunktion F.

5 Ungleichungen und Grenzwertsätze

Idee: In vielen Situationen taucht die *Summe von vielen gleichartigen Zufallsvariablen* auf. Wir möchten wisse, wie sich diese Summe etwa verhält, und untersuchen deshalb ihre Asymptotik, wenn die Anzahl der Summanden gegen unendlich geht.

5.1 Ungleichungen

Prop. (Markov-Ungleichung): Sei X eine Zufallsvariable und ferner $g : \mathcal{W}(X) \to [0, \infty)$ eine wachsende Funktion. Für jedes $c \in \mathbb{R}$ mit $g(c) > 0$ gilt dann

$$\mathrm{P}[X \geq c] \leq \frac{\mathrm{E}[g(X)]}{g(c)}.$$

Korr. (Chebyshev-Ungleichung): Sei Y eine Zufallsvariable mit endlicher Varianz. Für jedes $b > 0$ gilt dann

$$\mathrm{P}[|Y - \mathrm{E}[Y]| \geq b] \leq \frac{\mathrm{Var}[Y]}{b^2}.$$

5.2 Das Gesetz der grossen Zahlen

Satz (schwaches Gesetz der grossen Zahlen): Sei $X_1, X_2, ...$ eine Folge von unabhängigen Zufallsvariablen, die alle den gleichen Erwartungswert $E[X_i] = \mu$ und die gleiche Varianz $\mathrm{Var}[X_i] = \sigma^2$ haben. Sei

$$\overline{X}_n = \frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Dann konvergiert $\overline{X}_n$ für $n \to \infty$ in Wahrscheinlichkeit/stochastisch gegen $\mu = E[X_i]$, d.g.

$$P[|\overline{X}_n - \mu| > \epsilon] \xrightarrow[n\to\infty]{0} \quad \text{für jedes } \epsilon > 0.$$

Satz (starkes Gesetz der grossen Zahlen): Sei $X_1, X_2, ...$ eine Folge von unabhängigen Zufallsvariablen, die alle dieselbe Verteilung haben, und ihr Erwartungswert $E[X_i] = \mu$ sein endlich. Für

$$\overline{X}_n = \frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

gilt dann

$$\overline{X}_n \xrightarrow[n\to\infty]{} \mu \quad \text{P-fastsicher (P-f.s.)},$$

d.h.

$$P\left[\{\omega \in \Omega : \overline{X}_n(\omega) \xrightarrow[n\to\infty]{\mu}\}\right] = 1.$$

Bmk: Man hat es oft mit Zufallsvariablen zu tun, die unabhängig sind und alle dieselbe Verteilung haben. Der Kürze halber nennt man solche Zufallsvariablen ***i.i.d.*** (independent identically distributed).

5.3 Der Zentrale Grenzwertsatz

Satz (zentral Grenzwertsatz, ZGS): Sei $X_1, X_2, ...$ eine Folge von i.i.d. Zufallsvariablen mit $E[X_i] = \mu$ und $\mathrm{Var}[X_i] = \sigma^2$. Für die Summe

$$S_n = \sum_{i=1}^{n} X_i$$

gilt dann

$$\lim_{n\to\infty} P\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right] = \Phi(x) \quad \text{für alle } x \in \mathbb{R},$$

wobei Φ die Verteilungsfunktion der $\mathcal{N}(0, 1)$-Verteilung ist.

5.4 Grosse Abweichungen & Chernoff-Schranken

Def: Die **momenterzeugende Funktion** einer Zufallsvariable X ist

$$M_X(t) := E[e^{tX}] \quad \text{für } t \in \mathbb{R},$$

das ist immer wohldefiniert in $[0, \infty]$, kann aber $+\infty$ werde.

Satz: Seien $X_1, ..., X_n$ i.i.d. Zufallsvariablen, für welche die momenterzeugende Funktion $M_X(t)$ für alle $t \in \mathbb{R}$ endlich ist. Für jedes $b \in \mathbb{R}$ gilt dann

$$P[S_n \ge b] \le \exp\left(\inf_{t\in\mathbb{R}} (n \log M_X(t) - tb)\right).$$

Satz: Seien $X_1, ..., X_n$ unabhängig mit $X_i \sim Be(p_i)$ und $S_n = \sum_{i=1}^{n} X_i$. Sei ferner

$$\mu_n := E[S_n] = \sum_{i=1}^{n} p_i$$

und $\delta > 0$. Dann gilt

$$P[S_n \ge (1+\delta)\mu_n] \le \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^{\mu_n}.$$

6 Schätzer

6.1 Grundbegriffe

Def: Sei $X_1, ..., X_n$ eine Stichprobe, für die wir ein Modell suchen. Wir haben also einen Parameterraum Θ und für jedes $\vartheta \in \Theta$ einen Wahrscheinlichkeitsraum $(\Sigma, \mathcal{F}, P_\vartheta)$. Meistens ist $\Theta \subseteq \mathbb{R}^m$, und wir suchen dann für die Parameter $\vartheta_1, ..., \vartheta_m$ **Schätzer** $T_1, ..., T_m$ aufgrund unserer Stichprobe. Solche Schätzer sind Zufallsvariablen der Form $T_j = t_j(X_1, ..., X_n)$, wobei wir die **Schätzfunktionen** $t_j : \mathbb{R}^n \to \mathbb{R}$ noch geeignet finden müssen. Einsetzten von Daten $x_i = X_i(\omega)$, $i = 1, ..., n$, liefert dann **Schätzwerte** $T_j(\omega) = t_j(x_1, ..., x_n)$ für ϑ_j, $j = 1, ..., m$.

6.2 Maximum-Likelihood-Methode

Im jedem folgenden Modell P_ϑ sind $X_1, ..., X_n$ entweder diskret mit gemeinsamer Gewichtsfunktion $p_{\vec{X}}(x_1, ..., x_n; \vartheta)$ oder stetig mit gemeinsamer Dichtefunktion $f_{\vec{X}}(x_1, ..., x_n; \vartheta)$. Meistens sind die X_i unter P_ϑ i.i.d., dann ist also die gemeinsame Gewichtsfunktion

$$p_{\vec{X}}(x_1, ..., x_n; \vartheta) = \prod_{i=1}^{n} p_X(x_i; \vartheta)$$

bzw. die gemeinsame Dichtefunktion

$$f_{\vec{X}}(x_1, ..., x_n; \vartheta) = \prod_{i=1}^{n} f_X(x_i; \vartheta).$$

Anschaulich ist $p_{\vec{X}}(x_1, ..., x_n; \vartheta) = P_\vartheta[X_1 = x_1, ..., X_n = x_n]$ gerade die Wahrscheinlichkeit im Modell P_ϑ, dass unsere Stichprobe $X_1, ..., X_n$ die Werte $x_1, ..., x_n$ liefert, und $f_{\vec{X}}(x_1, ..., x_n; \vartheta)$ ist das übliche stetige Analogon.

Def: Die **Likelihood-Funktion** ist

$$L(x_1, ..., x_n; \vartheta) := \begin{cases} p_{\vec{X}}(x_1, ..., x_n; \vartheta) & \text{im diskreten Fall,} \\ f_{\vec{X}}(x_1, ..., x_n; \vartheta) & \text{im stetigen Fall.} \end{cases}$$

Die Funktion $\log L(x_1, ..., x_n; \vartheta)$ heisst **log-Likelihood-Funktion**. Sie hat gegenüber der Likelihood-Funktion den Vorteil, dass sie im i.i.d.-Fall durch eine Summe (statt ein Produkt) gegeben und damit zum Rechnen oft wesentlich einfacher ist.

Def: Der **Maximum-Likelihood-Schätzer (ML-Schätzer)** T_{ML} für ϑ wird dadurch definiert, dass er

$$\vartheta \to L(X_1, ..., X_n; \vartheta)$$

als Funktion von ϑ maximiert, d.h.

$$T_{\mathrm{ML}} = t_{\mathrm{ML}}(X_1, ..., X_n) \in \arg\max_{\vartheta\in\Theta} L(X_1, ..., X_n; \vartheta).$$

Bsp: *Bernoulli-Verteilung:* Im Modell P_p seien $X_1, ..., X_n$ i.i.d. $\sim Be(p)$. Hier ist $\vartheta = p$ und der ML-Schätzer für ϑ ist gegeben durch

$$T = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}_n.$$

Bsp: *Normalverteilung:* Im Modell P_ϑ seien $X_1, ..., X_n$ i.i.d. $\sim \mathcal{N}(\mu, \sigma^2)$. Der ML-Schätzer für $\vartheta = (\mu, \sigma^2)$ ist gegeben durch

$$T_1 = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}_n,$$

$$T_2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - (\overline{X}_n)^2.$$

Der Schätzer $T = (T_1, T_2)$ im obigen Beispiel ist ganz allgemein auch der sogenannte **Momentenschätzer** für

$$(E_\vartheta[X], \mathrm{Var}_\vartheta[X])$$

in jedem Modell P_ϑ, wo $X_1, ..., X_n$ i.i.d. sind.

Für eine Funktion $h : \Theta \to \mathbb{R}^d$ wollen wir die d Grössen $h_j(\vartheta)$, $j = 1, ..., d$, schätzen. Um den **Momentenschätzer** für $h(\vartheta)$ zu bestimmen, gehen wir in der Regel wie folgt vor:

1. Für $j = 1, ..., d$ berechnen wir in jedem Modell P_ϑ das j-te Moment $m_j(\vartheta) = E_\vartheta[X^j]$ als Funktion des Parameters ϑ.

2. Für $j = 1, ..., d$ definieren wir das j-te **empirische Mittel** oder **Stichprobenmittel** als

$$\tilde{m}_j(x_1, ..., x_n) := \frac{1}{n} \sum_{i=1}^{n} x : i^j.$$

3. Wir betrachten das System von d Gleichungen $\tilde{m}_j(x_1, ..., x_n) = m_j(\vartheta)$, und fassen das auf als ein Gleichungssystem für die d Unbekannten $h_1(\vartheta), ..., h_d(\vartheta)$. Falls eine eindeutige Lösung existiert, so nennen wir diese $t_{MM}(x_1, ..., x_n)$. Der Momentenschätzer für $h(\vartheta)$ ist dann definiert durch $T_{MM} := t_{MM}(X_1, ..., X_n)$.

6.3 Verteilungsaussagen

Satz: Seien $X_1, ..., X_n$ i.i.d. $\sim \mathcal{N}(\mu, \sigma^2)$. Dann gilt:
1. $\overline{X}_n$ ist normalverteilt gemäss $\mathcal{N}(\mu, \frac{1}{n}\sigma^2)$, und damit gilt $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$.
2. $\frac{n-1}{\sigma^2} S^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$ ist χ^2-verteilt mit $n-1$ Freiheitsgraden.
3. $\overline{X}_n$ und S^2 sind unabhängig.
4. Der Quotient

$$\frac{\overline{X}_n - \mu}{S/\sqrt{n}} = \frac{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}}{S/\sigma} = \frac{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{1}{n-1} \frac{n-1}{\sigma^2} S^2}}$$

ist t-verteilt mit $n-1$ Freiheitsgraden.

7 Tests

7.1 Grundbegriffe

Wie im letzten Abschnitt betrachten wir eine Stichprobe $X_1, ..., X_n$ sowie eine Familie von Wahrscheinlichkeiten P_ϑ mit $\vartheta \in \Theta$. Wir haben nun eine Vermutung, wo in Θ der richtige Parameter ϑ liegen könnte, wollen diese aber mit Hilfe der Daten überprüfen. Das Grundproblem ist also, eine Entscheidung zwischen zwei konkurrierenden Modellklassen zu treffen - der **Hypothese** $\Theta_0 \subseteq \Theta$ und der **Alternative** $\Theta_A \subseteq \Theta$, wobei $\Theta_0 \cap \Theta_A = \emptyset$. Meist schreibt man das als:

$$\text{Hypothese } H_0 : \vartheta \in \Theta_0,$$
$$\text{Alternative } H_A : \vartheta \in \Theta_A.$$

Def: Ein **Test** ist allgemein eine Entscheidungsregel, die zu gegebenen Daten $X_1, ..., X_n$ einen von zwei Werten liefert. Dabei nimmt man als die zwei Werte 0 und 1, und man interpretiert den Wert 1 als die Entscheidung, die Hypothese H_0 abzulehnen.

Konkreter sieht die Entscheidungsregel meist folgendermassen aus. Man hat eine Abbildung $t : \mathbb{R}^n \to \mathbb{R}$, $(x_1, ..., x_n) \to t(x_1, ..., x_n)$, und einen **kritischen Bereich** oder **Verwerfungsbereich** $K \subseteq \mathbb{R}$.

Die Zufallsvariable $T = t(X_1, ..., X_n)$ heisst dann **Teststatistik**. Die Entscheidungsregel ist $I_{\{t(x_1,...,x_n) \in K\}}$, d.h. man verwirft die Hypothese genau dann, wenn der realisierte Wert $t(x_1, ..., x_n) = t(X_1(\omega), ..., X_n(\omega)) = T(\omega)$ im Verwerfungsbereich K liegt.

Die Entscheidung bei einem Test kann auf zwei verschiedene Arten falsch herauskommen:

1. *Fehler 1. Art:* Die Hypothese wird zu unrecht abgelehnt. Das passiert für $\vartheta \in \Theta_0$ und $T \in K$. Deshalb heisst $P_\vartheta[T \in K]$ für $\vartheta \in \Theta_0$ die Wahrscheinlichkeit für einen Fehler 1. Art.
2. *Fehler 2. Art:* Die Hypothese wird zu Unrecht akzeptiert. Das passiert für $\vartheta \in \Theta_A$ und $T \notin K$, und deshalb heisst $P_\vartheta[T \notin K] = 1 - P_\vartheta[T \in K]$ für $\vartheta \in \Theta_A$ die Wahrscheinlichkeit für einen Fehler 2. Art.

Weil ein gleichzeitige Minimieren auf Θ_0 und Maximieren auf Θ_A der Fehlerwahrscheinlichkeit in der Regel grundsätzlich nicht möglich ist, hat sich folgendes zweistufiges Vorgehen durchgesetzt:

1. Man wählt zuerst ein **Signifikanzniveau** $\alpha \in (0, 1)$ und sorgt zunächst für
$$\sup_{\vartheta \in \Theta_0} P_\vartheta[T \in K] \leq \alpha,$$
d.h. man kontrolliert die Wahrscheinlichkeit für einen Fehler 1. Art durch α.
2. Anschliessend versucht man, die **Macht** der Tests, die Funktion
$$\beta : \Theta_A \to [0, 1], \quad \vartheta \to \beta(\vartheta) := P[T \in K],$$
möglichst gross zu bekommen. Das heisst, man versucht die Grösse $1 - \beta(\vartheta) = P_\vartheta[T \notin K]$ für $\vartheta \in \Theta_A$, also die Wahrscheinlichkeit für einen Fehler 2. Art, möglichst klein zu bekommen.

Bmk: Bei einer diskreten Teststatistik T kann ein vorgegebenes Niveau α in der Regel nicht eingehalten werden, d.h. es ist unmöglich, einen kritischen Bereich K mit $P_\vartheta[T \in K] = \alpha$ zu finden. Einen Ausweg bietet ein sogenannter **randomisierter Test**. Man wählt dazu $\gamma \in [0, 1]$ so, dass gilt $\gamma P_{\vartheta_0}[T > c] + (1 - \gamma) P_{\vartheta_0}[T > c + 1] = \alpha$ und entscheidet wie folgt: Ist $T > c$, so verwirft man H_0 mit Wahrscheinlichkeit γ, d.h. H_0 wird abgelehnt, falls erstens $T > c$ gilt und zweitens eine unabhängige $\mathcal{U}(0, 1)$-verteilte Zufallsvariable einen Wert $\leq \gamma$ realisiert.

7.2 Konstruktion von Tests

Der folgende systematische Ansatz führt in vielen Situationen zu einem optimalen Test. Sei $L(x_1, ..., x_n; \vartheta)$ die Likelihood-Funktion. Für diskrete X_i ist also im Modell P_ϑ die Wahrscheinlichkeit $P_\vartheta[X_1 = x_1, ..., X_n = x_n]$, die beobachteten Werte $x_1, ..., x_n$ zu erhalten. Für $\vartheta_0 \in \Theta_0$ und $\vartheta_A \in \Theta_A$ betrachten wir den **Likelihood-Quotienten**

$$R(x_1, ..., x_n; \vartheta_0, \vartheta_A) := \frac{L(x_1, ..., x_n; \vartheta_A)}{L(x_1, ..., x_n; \vartheta_0)}.$$

Ist der Quotient gross, dann sind die Beobachtungen $x_1, ..., x_n$ als Resultate im Modell P_{ϑ_A} deutlich wahrscheinlicher als im Modell P_{ϑ_0}.

Satz (Neyman-Pearson-Lemma): Sei $\Theta_0 = \{\vartheta_0\}$ und $\Theta_A = \{\vartheta_A\}$. Sei $T := R(X_1, ..., X_n; \vartheta_0, \vartheta_A)$ und $L := (c, \infty)$ sowie $a^* := P_{\vartheta_0}[T \in K] = P_{\vartheta_0}[T > c]$. Der **Likelihood-Quotienten-Test** mit Teststatistik T und kritischem Bereich K ist dann im folgenden Sinn optimal: Jeder andere Test mit Signifikanzniveau $\alpha \leq \alpha^*$ hat kleinere Macht bzw. eine grössere Wahrscheinlichkeit für einen Fehler 2. Art.

7.3 Der p-Wert

Wir haben folgende Situation: Sei $X_1, ..., X_n$ eine Stichprobe vom Umfang n. Wir wollen eine Hypothese $H_0 : \vartheta = \vartheta_0$ gegen eine Alternative $H_A : \vartheta \in \Theta_A$ testen. Wir nennen $T = t(X_1, ..., X_n)$ die Teststatistik und K den kritischen Bereich zum fixierten Niveau α. Seien $x_i = X_{(}\omega)$, $i = 1, ..., n$, die realisierten Daten. Bei einem Test zum Niveau α wird dann also die Hypothese abgelehnt, wenn $T(\omega) = t(x_1, ..., x_n)$ einen Wert in K ergibt, und wir haben $P_{\vartheta_0}[T \in K] \leq \alpha$.

Def: Alternativ zum obigen Vorgehen kann man auch den sogenannten **p-Wert** berechnen. Das ist die Wahrscheinlichkeit, unter der Hypothese H_0 einen mindest so extremen Wert für T zu bekommen wie den Wert $T(\omega)$ aus der Stichprobe. "Extrem" bezieht sich dabei auf die Alternative - ist also beispielsweise $H_A : \vartheta > \vartheta_0$ und entspricht das einem grossen Wert von T, so ist

$$\text{p-Wert}(\omega) = P_{\vartheta_0}[T > t_0]|_{t_0 = t(X_1(\omega), ..., X_n(\omega))},$$

d.h. wir berechnen die Wahrscheinlichkeit $t_0 \to P_{\vartheta_0}[T > t_0]$ als Funktion von t_0 und setzen dann für t_0 den zufälligen Wert $T(\omega)$ ein.

7.4 z-Test

Test für den Erwartungswert einer Normalverteilung mit bekannter Varianz der Grundgesamtheit. Seien also $X_1, \ldots, X_n \sim \mathcal{N}(\theta, \sigma^2)$-verteilt (i.i.d.) für bekanntes $\sigma > 0$.

- **Hypothese:** $H_0 : \theta = \theta_0$
- **Teststatistik:**

$$T = \frac{\overline{X}_n - \theta_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1) \qquad \text{unter } P_{\theta_0}$$

- **Kritische Bereiche** (zum Signifikanzniveau $\alpha \in (0, 1)$) kann au Tabelle abgelesen werden:

Alternative H_A	Kritischer Bereich
$\theta < \theta_0$	$(-\infty, z_\alpha)$
$\theta > \theta_0$	$(z_{1-\alpha}, \infty)$
$\theta \neq \theta_0$	$(-\infty, z_{\alpha/2}) \cup (z_{1-\alpha/2}, \infty)$

Dabei bezeichnet z_α das α-Quantil der Standardnormalverteilung. Man findet es, indem man in der Tabelle der

Standardnormalverteilung nach $\Phi^{-1}(\alpha)$ sucht. Aus Symmetriegründen gilt $z_\alpha = -z_{1-\alpha}$.

$$\Phi(z_\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_\alpha} e^{-x^2/2}\, dx = \alpha$$

Rezept Fehler 2. Art berechnen:
Nehme an: einseitiger z-Test, $T = \frac{\overline{X_n}-\mu_0}{\sigma/\sqrt{n}}$, $\mu_0 = 70$
$H_0 : \mu = \mu_0$; $H_A : \mu < \mu_0$.
Kritischer Bereich mit 5%-Niveau: $K = (-\infty, -1.645)$
Objective: Fehler 2. Art finden für $\mu_A = 69.5$. Wir nehmen an, dass $T = \frac{\overline{X_n}-\mu_A}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ unter P_{μ_A}

$$\begin{aligned}
\text{Fehler 2. Art} &= P_{\mu_A}[T \notin K] \\
&= P_{\mu_A}[T > -1.645] \\
&= P_{\mu_A}\left[\frac{\overline{X_n}-\mu_0}{\sigma/\sqrt{n}} > -1.645\right] \\
&= P_{\mu_A}\left[\frac{\overline{X_n}-\mu_A}{\sigma/\sqrt{n}} > \frac{\mu_0-\mu_A}{\sigma/\sqrt{n}} - 1.645\right] \\
&= 1 - P_{\mu_A}\left[\frac{\overline{X_n}-\mu_A}{\sigma/\sqrt{n}} \leq \frac{\mu_0-\mu_A}{\sigma/\sqrt{n}} - 1.645\right] \\
&= 1 - \Phi\left(\frac{\mu_0-\mu_A}{\sigma-\sqrt{n}} - 1.645\right) \quad \text{weil} \sim \mathcal{N}(0,1)
\end{aligned}$$

7.5 t-Test

Test für den Erwartungswert einer Normalverteilung mit unbekannter Varianz. Seien also $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$-verteilt (i.i.d.) für unbekanntes $\sigma > 0$.

- **Hypothese:** $H_0 : \mu = \mu_0$. Formal präziser wäre $\Theta_0 = \{\theta = (\mu_0, \sigma),\ \sigma > 0\}$
- **Teststatistik:** $T = \frac{\overline{X_n}-\mu_0}{S/\sqrt{n}} \sim t_{n-1}$ unter P_{μ_0}, wobei $S^2 :=$ empirische Stichprobenvarianz
- **Kritische Bereiche** (zum Signifikanzniveau $\alpha \in (0,1)$) kann aus Tabelle abgelesen werden:

Alternative H_A	Kritischer Bereich
$\mu < \mu_0$	$(-\infty, t_{n-1,\alpha})$
$\mu \neq \mu_0$	$(-\infty, t_{n-1,\alpha/2}) \cup (t_{n-1,1-\alpha/2}, \infty)$

Dabei bezeichnet $t_{m,\alpha}$ das α-Quantil der t_m-Verteilung. Aus Symmetriegründen gilt $t_{m,\alpha} = -t_{m,1-\alpha}$:

$$\int_{-\infty}^{t_{m,\alpha}} f_m(x)\, dx = \alpha$$

wobei f_m die Dichte der t_m Verteilung ist. Diesen Wert erhält man aus einer Tabelle zur t-Verteilung.

7.6 Gepaarte Zweistichproben-Tests für Normalverteilungen

Seien $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ Zufallsvariablen, so dass (X_i, Y_i) natürliche Paare bilden. Bezeichnen wir nun $Z_i := X_i - Y_i$.

- *bekannte Varianz:* Falls $Z_1, \ldots, Z_n \sim \mathcal{N}(\theta, \sigma^2)$ (i.i.d.) für bekanntes $\sigma > 0$, dann kann z-Test analog zu Kapitel 8.3 angewendet werden.
- *unbekannte Varianz:* Falls $Z_1, \ldots, Z_n \sim \mathcal{N}(\mu, \sigma^2)$ (i.i.d.) für unbekanntes $\sigma > 0$, dann kann t-Test analog zu Kapitel 8.4 angewendet werden.

7.7 Ungepaarte Zweistichproben-Tests für Normalverteilungen

Seien $X_1, \ldots, X_n \sim \mathcal{N}(\mu_X, \sigma_X^2)$ (i.i.d.) und $Y_1, \ldots, Y_m \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ (i.i.d.), so dass alle X_i, Y_j unabhängig.

7.7.1 Normalverteilungen mit bekannten Varianzen

Seien also σ_X, σ_Y bekannt.

- **Hypothese:** $H_0 : \mu_X - \mu_Y = \mu_0$ (bspw. $\mu_0 = 0$)
- **Teststatistik:**

$$T = \frac{\overline{X}_n - \overline{Y}_m - \mu_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim \mathcal{N}(0,1) \qquad \text{für } P_{\mu_0}$$

Die kritischen Bereiche zum Signifikanzniveau sind analog zur Tabelle aus Kapitel 8.3.

7.7.2 Normalverteilungen mit unbekannten aber gleichen Varianzen

Sei also $\sigma_X = \sigma_Y = \sigma$ für $\sigma > 0$ unbekannt.

- **Hypothese:** $\mu_X - \mu_Y = \mu_0$ (bspw. $\mu_0 = 0$)
- **Teststatistik:**

$$T = \frac{\overline{X}_n - \overline{Y}_m - \mu_0}{S\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2} \qquad \text{unter } P_{\mu_0}$$

- **Kritische Bereiche:** analog zu Tabelle aus Kapitel 8.4, jedoch ist nun die Anzahl der Freiheitsgrade $n+m-2$ und nicht mehr $n-1$.

Dabei benutzen wir für die Varianz ein gewichtetes Mittel aus den Stichprobenvarianzen S_X, S_Y, definiert als

$$S^2 := \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$$

8 Konfidenzbereiche

Def: Formal ist ein realisierter **Konfidenzbereich** für ϑ zu Daten $x_1, \ldots, x_n$ eine Menge $C(x_1, \ldots, x_n) \subseteq \Theta$. In den meisten Fällen ist das ein Intervall, dessen Endpunkte von $x_1, \ldots, x_n$ abhängen. Ersetzen wir die Daten $x_1, \ldots, x_n$ durch die sie generierenden Zufallsvariablen, so ist $\tilde{C}(\omega) := C(X_1, \ldots, X_n)$ also eine zufällige Teilmenge von Θ, mit Realisierung $\tilde{C}(\omega) = C(X_1(\omega), \ldots, X_n(\omega))$ bei einem festen ω. Ein solches C heisst Konfidenzbereich zum Niveau $1-\alpha$, falls gilt:

$$P_\vartheta[C(X_1, \ldots, X_n) \ni \vartheta] \geq 1 - \alpha \quad \text{für alle } \vartheta \in \Theta,$$

d.h. in jedem Modell erwischt man den Parameter mit grosser Wahrscheinlichkeit.

9 Nice To Know

9.1 Ableitung, Integration

- **Summenregel** $(f(x) + g(x))' = f'(x) + g'(x)$
- **Produktregel** $(f(x) \cdot g(x))' = f'(x)g(x) + f(x)g'(x)$
- **Quotientenregel** $\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x)-f(x)g'(x)}{g^2(x)} (g \neq 0)$
- **Kettenregel** $(f(g(x)))' = (f \circ g)' = f'(g(x))g'(x)$
- **Partielle Integration:** $\int_a^b f'(x) \cdot g(x)dx = [f(x)g(x)]_a^b - \int_a^b f(x)g'(x)$
- **Substitution:** $\int_{\phi(a)}^{\phi(b)} f(x)dx = \int_a^b f(\phi(t))\phi'(t)dt$
- $a+c, b+c \in I$ $\int_a^b f(t+c)dt = \int_{a+c}^{b+c} f(x)dx$
- $ca, cb \in I$**:** $\int_a^b f(ct)dt = \frac{1}{c} f(x)dx$
- **Logarithmus:** $\int \frac{f'(t)}{f(t)}dt = \log(|f(x)|)$, bzw. $\int_a^b \frac{f'(t)}{f(t)}dt = \log(f(|b|)) - \log(f(|a|))$

9.2 Schätzer Rezepte

9.2.1 Maximum Likelihood Schätzer

- Likehood-Funktion L bestimmen
- Falls Zufallsvariablen i.i.d., dann $\log L$ bestimmen
- $\log L$ (oder L) maximieren: ableiten von $\log L$ (oder L) und gleich 0 setzen.
- $\Rightarrow$ Funktion, die Parameter schätzt

9.2.2 Momentenschätzer mit zentralen/rohen Momenten

k-tes **rohes** Moment: $\mathrm{E}[X^k]$
k-tes empirisches ˜: $\frac{1}{n}\sum_{i=1}^n x_i^k$
k-tes **zentrales** Moment: $\mathrm{E}[(X - \mathrm{E}[X])^k]$
1. theoretisches ˜= $\mathbb{E}[X]$, 2. theoretisches ˜= $\mathrm{Var}[X]$)
1. empirisches ˜= $\overline{X}$, 2. empirisches ˜= $\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$, k-tes empirisches ˜: $\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^k$

9.3 Zusammenhänge Tests

Zusammenhänge α (Signifikanzniveau), Fehler 1. Art, β (Macht), Fehler 2. Art $(1-\beta)$

- α grösser $\iff$ Fehler 1. Art grösser $\iff$ Fehler 2. Art kleiner $\iff$ Macht grösser
- α kleiner $\iff$ Fehler 1. Art kleiner $\iff$ Fehler 2. Art

grösser $\iff$ Macht kleiner

9.4 Beispiel p-Wert

Beispiel: Hintergrundfarbe einer Webseite ändern und schauen, ob sich die Besuchsdauer von Nutzern verändern.

$$\mu = 20$$

- Nullhypothese H_0: $\mu = 20$ nach der Änderung
  Alternative H_A: $\mu > 20$ nach der Änderung
- Signifikanzniveau: $\alpha = 0.05$
- Stichprobe: $n = 100$, $\overline{X} = 25$, (σ)
- p-Wert: $P[\overline{X} \geq 25 \mid H_0$ ist wahr$]$
- Falls p-Wert $< \alpha$: H_0 verwerfen (und H_A akzeptieren)
  Falls p-Wert $\geq \alpha$: H_0 nicht verwerfen (keine Aussage)

Der p-Wert ist **nicht** $P[H_0$ ist wahr | Stichprobe$]$

9.5 Begriffe Tests

Modell: z.B. Unter P_φ sind die X_i i.i.d. $\sim$ Poi(λ), $i = 1, ..., 6$, λ unbekannt.
Teststatistik: Hilfsfunktion bei statistischen Tests. Kann zum Beispiel mittels Likelihood-Quotienten-Vorgehen gefunden werden.
Punkte; Beispiel jeweils in Klammern:

- Modell (*Unter P_ϑ sind die X_i i.i.d. $\sim$ Poi...*)
- Nullhypothese ($H_0 : p = ...$)
- Alternativhypothese ($H_A : p < ...$)
- Teststatistik ($T =< R >$ *also Likelihood-Quotienten verwenden*)
- Verteilung der Teststatistik unter der Nullhypothese ($H_0 : T \sim$ *Bin...*)
- Verwerfungsbereich ($K = [a, b]$, $P_{\vartheta_0}[T \in K] \leq 5\%$...)
- beobachteter Wert der Teststatistik ($t = T(\omega) = 6$)
- Testentscheid (*Nullhypothese wird nicht verworfen...*)
- eventuell p-Wert

Wichtig:
Falls beobachtetes Ergebnis im Verwerfungsbereich: H_0 wird abgelehnt, H_A wird angenommen.
Falls beobachtetes Ergebnis nicht im Verwerfungsbereich: H_0 wird nicht abgelehnt (keine Aussage über Annahme!), keine Aussage über H_A
p-Wert: kleinstes Niveau, auf dem der Test die Nullhypothese noch verwirft.
Auch: Falls $H_0 : p = 123$, $H_A : p < 123$ Mit Statistik $P_{H_0}[T \leq$ Beobachteter Wert$]$. p-Wert ist so wie die Signifikanz des Testresultats.

9.6 Summenformeln

1. $\sum_{i=1}^n i = \frac{n(n+1)}{2}$
2. $\sum_{i=0}^n k^i = \frac{1-k^{n+1}}{1-k}$
3. $\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$
4. $\sum_{i=1}^n i^3 = \frac{n^2(n+1)^2}{4}$
5. $(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k$

10 Exercises

10.1 Aufgabe 1 - Multiple Choice

1) Münze (fair) wird 2x geworfen. Für $i = 1, 2$ sei K_i das Ereignis, dass beim $i-$ten Wurf Kopf erscheint. Dann gilt:

1. **K_1 und K_2 unabhängig.**
2. K_1 und K_2 disjunkt.
3. K_1 und K_2 unabhängig und $K_1 \cup K_2 = \Omega$

$K_1 = \{(K, K), (K, Z)\}$ und $K_2 = \{(K, K), (Z, K)\}$, daher sind 2 und 3 falsch. 1 richtig da $P[K_1 \cap K_2] = P[K_1] \cdot P[K_2]$.
2) A und B Ereignisse mit $P[A] > 0$ und $A \cap B = \emptyset$. $P[B|A] = 0$ ist

1. **wahr.**
2. nicht wahr.

$P[B|A] = \frac{P[A \cap B]}{P[A]} = \frac{0}{P[A]} = 0$.
3) Untersuchung einer Bodenprobe auf Stoffe X und Y. A : "X in Probe", B : "Y in Probe". Es gilt $P[A] = 0.1$ und $P[B] = 0.3$. Wir wissen, dass WHS, dass die Probe weder X noch Y enthält, ist 0.7. Dann:

1. $P[A \cap B] = 0.03$
2. $\mathbf{P[A \cap B] = 0.1}$
3. Zu wenig Infos für $P[A \cap B]$

Es gilt $P[A \cap B] = P[A] + P[B] - P[A \cup B]$. Es gilt $P[A \cup B] = 1 - P[(A \cup B)^C] = 1 - 0.7 = 0.3$. Nun gilt $P[A \cap B] = 0.1 + 0.3 - 0.3 = 0.1$
4) Sei X eine ZV mit $E[X] = 2, \text{Var}[X] = 3$. Dann gilt:

1. $E[X^2] = 4$
2. $\mathbf{E[X^2] = 7}$
3. Zu wenig Infos um $E[X^2]$ zu berechnen.

Es gilt $\text{Var}[X] = E[X^2] - (E[X])^2 \Rightarrow E[X^2] = \text{Var}[X] + (E[X])^2 = 3 + 4 = 7$.
5) Wie oft einen fairen Würfel im Schnitt werfen bis eine 3 kommt?

1. **6 mal**
2. 6^6 mal
3. $6!$ mal

Sei X die Anzahl versuche, bis eine 3 erscheint. Dann ist $X \sim$ *Geom*$(\frac{1}{6})$. Es gilt $E[X] = \frac{1}{p} = 6$.
6) Sei X eine ZV mit Verteilungsfunktion

$$F_X(x) = \begin{cases} 1 - e^{-cx}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

für ein $c > 0$. Der **Median** (d.h. $F_X^{-1}(\frac{1}{2})$) beträgt:

1. $\frac{1}{2}$
2. $\mathbf{\frac{1}{c} \ln 2}$
3. $1 - e^{-\frac{1}{2}c}$

$y = 1 - e^{-cx} \Rightarrow \ln y = \ln(1 - e^{-cx}) = \frac{\ln 1}{\ln e^{-cx}} = \frac{\ln 1}{-cx \ln e} = \frac{\ln 1}{-cx} \Rightarrow -cx \ln y = \ln 1 \Rightarrow x = \frac{\ln 1}{-c \ln y} = -\frac{1}{c} \ln(1 - y) = F_X^{-1}(y)$. Dann ist $F_X^{-1}(\frac{1}{2}) = -\frac{1}{c} \ln 1 - \frac{1}{2} = -\frac{1}{c} \ln \frac{1}{2} = -\frac{1}{c} \ln 2^{-1} = \frac{1}{c} \ln 2$.
7) Seien 10% der Menschen Raucher. X sei die Anzahl Raucher unter 100 zufällig gewählten Personen. Wie gross ist die Standardabweichung von X?

1. **3**
2. $\sqrt{10}$
3. 10

Es gilt $X \sim Bin(100, 0.1)$. Die **Standardabweichung** ist gegeben durch $\sqrt{\text{Var}[X]}$, und $\text{Var}[X]$ für $X \sim Bin(n, p)$ ist $np(1 - p) = 100 \cdot 0.1(1 - 0.1) = 9$ und daher ist $\sqrt{\text{Var}[X]} = 3$.
8) X und Y zwei ZV die **lognormalverteilt** und unabhängig sind, d.h. $\ln X$ und $\ln Y$ sind normalverteilt. Dann gilt:

1. **XY ist lognormalverteilt**
2. XY ist normalverteilt
3. e^{x+y} ist normalverteilt

Da $\ln XY = \ln X + \ln Y$ gilt, ist $\ln XY$ als Summe unabhängiger normalverteilten ZV ebenfalls normalverteilt und daher ist XY lognormalverteilt.
9) Die stetige Zufallsvariable X habe die kumulative Verteilungsfunktion

$$F(x) = 1 - e^{-2x}, \quad x \geq 0.$$

Welche der folgenden Aussagen ist richtig?

1. Die Dichte von X ist $f(x) = 2xe^{-2x}$, $x \geq 0$
2. $P[X \geq 0.5] = e^{-0.5}$
3. $P[X = 1] = 1 - e^{-2}$
4. **Für $\mathbf{Y = 2X}$ gilt $\mathbf{P[Y \leq 2] = 1 - e^{-2}}$**

1. Falsch. Dichte ist (Verteilungsfunktion)'. $(1 - e^{-2x})' = (-e^{-2x})' = -1 \cdot (-2x)' \cdot (e^{-2x}) = 2e^{-2x}$. 2. Falsch. $P[X \geq 0.5] = 1 - P[X < 0.5] = 1 - F(0.5) = 1 - (1 - e^{-2 \cdot 0.5}) = e^{-1}$. 3. Falsch. Für stetige Verteilungen gilt $P[X = x] = 0$ für alle x. Dies ist weil $P[X = x] = P[x \geq X \geq x]$, was einem Integral von x nach x gleicht. 4. Richtig. $P[Y \leq 2] = P[2X \leq 2] = P[X \leq 1] = 1 - e^{-2 \cdot 1} = 1 - e^{-2}$.
10) Bei der Berechnung des realisierten p-Wertes eines Tests

1. **muss man wissen, ob der Test einseitig oder zweiseitig ist.**
2. braucht man die konkreten Daten nicht.
3. muss man das Signifikanzniveau α kennen.

1. Man muss wissen, was es bedeutet, dass die Teststatistik T einen extremeren Wert als den beobachteten Wert $t = T(\omega)$ annimmt. Bei zweiseitigen Tests betrachtet man daher das Ereignis $\{|T| < (>) \, t\}$, während dies für einseitige Tests $\{T < (>) \, t\}$ ist. **11)** Seien A und B zwei Ereignisse. Die Aussage $P[A^C \cap B^C] = 1 - P[A \cap B]$ ist:

1. wahr
2. **nicht wahr**

2. Es gilt $P[A^C \cap B^C] = 1 - P[(A^C \cap B^C)^C)] = P[A \cup B]$.
12) Seien A und B zwei Ereignisse. Das Additionsgesetz $P[A \cup B] = P[A] + P[B]$ gilt, falls:

1. A und B unabhängig sind
2. **A und B disjunkt sind.**
3. A eine Teilmenge von B ist.

2. Die ganze Formel ist $P[A \cup B] + P[A \cap B] = P[A] + P[B]$, wenn nun A und B disjunkt sind, dann ist $P[A \cap B] = 0$ und die obige Formel gilt.
12) Wir betrachten einen Labortest für eine Krankheit. Sei A das Ereignis, dass eine getestete Person die Krankheit hat. Sei B das Ereignis, dass der Test positiv ist. Man weiss $P[B|A] = 0.99$ und $P[B^C|A^C] = 0.995$. Der Anteil der Bevölkerung, der die Krankheit hat, ist 0.1%. Wie gross ist die WHS, dass eine zufällig gewählte Person mit einem positiven Test tatsächlich die Krankheit hat?

1. $P[A|B] \simeq 0.065$
2. $\mathbf{P[A|B] \simeq 0.165}$
3. $P[A|B] \simeq 0.560$

2. Nach dem Satz von Bayes gilt:

$$P[A|B] = \frac{P[B|A] \cdot P[A]}{P[B|A] \cdot P[A] + P[B|A^C] \cdot P[A^C]}$$
$$= \frac{0.99 \cdot 0.001}{0.99 \cdot 0.001 + (1 - 0.995) \cdot 0.999} \simeq 0.165$$

13) Seien X und Y unkorrelierte ZV mit Varianzen $\mathrm{Var}[X] = 2$ und $\mathrm{Var}[Y] = 3$. Wie gross ist die Varianz von $Z = 7X - 2Y$?

1. 86
2. **110**
3. 8

2. Es gilt $\mathrm{Var}[aX + b] = a^2\mathrm{Var}[X]$. Daher ist $\mathrm{Var}[7X - 2Y] = 7^2\mathrm{Var}[X] + (-2)^2\mathrm{Var}[Y] = 49 \cdot 2 + 4 \cdot 3 = 110$.
14) Ein fairer Würfel wird 1800mal geworfen. Die Zufallsvariable X sei die Anzahl der Würfe mit Augenzahl 6. Wie gross ist die Varianz von X?

1. 1800
2. 300
3. **250**

3. Es gilt, dass $x \sim Bin(1800, \frac{1}{6})$ ist. Daher ergibt sich $\mathrm{Var}[X] = 1800 \cdot \frac{1}{6} \cdot \frac{5}{6} = 250$.
15) Seien $X \sim \mathcal{N}(0, 1)$ und $Y \sim Unif(0, 1)$ ZV mit $Cov(X, Y) = 0$. Was ist korrekt?

1. X und Y sind unabhängig
2. $\mathbf{E[XY] = 0}$
3. Keine der Aussagen trifft zu.

2. Da $X \sim \mathcal{N}(0, 1)$ ist, gilt $E[X] = 0$. Mit $0 = Cov(X, Y) = E[XY] - E[X]E[Y]$ gilt $E[XY] = 0$. Die beiden ZV sind zwar unkorreliert, müssen aber nicht unbedingt unabhängig sein.
16) Seien X und Y unabhängig und identisch $Poi(\frac{9}{2})$ verteilt. Wie gross ist die Standardabweichung der Zufallsvariablen $X + Y$?

1. **3**
2. $\frac{9}{2}$
3. $\sqrt{\frac{9}{2}}$

1. Da X und Y sind, gilt $\mathrm{Var}[X+Y] = \mathrm{Var}[X]+\mathrm{Var}[Y] = \frac{9}{2}+\frac{9}{2} = 9$. Daher ist $\sqrt{\mathrm{Var}[X + Y]} = \sqrt{9} = 3$.

10.2 Aufgabe 2 - χ^2-Verteilung

Eine Zufallsvariable Y heisst χ^2-verteilt mit Freiheitsgrad $v \in \mathbb{N}$ ($Y \sim \chi^2_v$), falls

$$Y = \sum_{k=1}^{v} Z_k^2,$$

wobei $Z_1, \ldots, Z_v$ i.i.d $\mathcal{N}(0, 1)$-verteilt sind.
a) Zeige, dass $\mathbf{E}[Y] = v$ und $\mathbf{Var}[Y] = 2v$ gilt.
Für ganzzahlige Werte von v kann Y geschrieben werden als $Y = \sum_{k=1}^{v} Z_k^2$ mit $Z_1, \ldots, Z_v$ i.i.d. $\mathcal{N}(0, 1)$. Daraus folgt einerseits:

$$E[Y] = \sum_{k=1}^{v} E[Z_k^2] = v \cdot 1 = v.$$

Anderseits ist wegen der Unabhängigkeit und identischen Verteilung der Z_k:

$$E[Y^2] = E[(Z_1^2 + Z_2^2 + \cdots + Z_v^2)^2] = vE[Z_1^4] + v(v-1)(E[Z_1^2])^2,$$

oder alternativ

$$\mathrm{Var}[Y] = \mathrm{Var}\left[\sum_{k=1}^{v} Z_k^2\right] = \sum_{k=1}^{v} \mathrm{Var}[Z_1^2] = v(E[Z_1^4] - E[Z_1^2]^2).$$

Mit $E[Z_1^2] = 1$ und

$$E[Z_1^4] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} Yx^4 e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} Yx^3 xe^{-x^2/2} dx$$
$$= \frac{1}{\sqrt{2\pi}} \left(-x^3 e^{-x^2/2}\right)\Big|_{-\infty}^{\infty} + \frac{3}{\sqrt{2\pi}} \int_{-\infty}^{\infty} Yx^2 e^{-x^2/2} dx = 3$$

erhalten wir schliesslich $E[Y^2] = 3v + v(v-1) = v^2 + 2v$ und daraus $\mathrm{Var}[Y] = (v^2 + 2v) - v^2 = 2v$, beziehungsweise

$$\mathrm{Var}[Y] = v(3-1) = 2v.$$

b) Gib mit Hilfe der Chebyshev-Ungleichung eine untere Schranke für die Wahrscheinlichkeit

$$P\left[\left|\frac{Y}{v} - 1\right| \leq \frac{3}{4}\right]$$

an. Wie lautet der konkrete Wert der Schranke für $v = 12$?
Die Chebyshev-Ungleichung lautet $P[|Y - v| > c] \leq \frac{\mathrm{Var}[Y]}{c^2} = \frac{2v}{c^2}$. Somit folgt:

$$P\left[\left|\frac{Y}{v} - 1\right| \leq \frac{3}{4}\right] = P\left[\left|\frac{Y-v}{v}\right| \leq \frac{3}{4}\right] = 1 - P\left[|Y - v| > \frac{3v}{4}\right]$$
$$\geq 1 - \frac{2v}{9v^2/16} = 1 - \frac{32}{9v}.$$

Für $v = 12$ erhalten wir somit:

$$P\left[\left|\frac{Y}{v} - 1\right| \leq \frac{3}{4}\right] = \frac{19}{27}.$$

c) Leite eine Annäherung für die obige Wahrscheinlichkeit mit Hilfe des zentralen Grenzwertsatzes her. Wie lautet der konkrete Wert der Approximation für $v = 12$?
Die $Y_i = Z_i^2$ sind i.i.d. $\sim \chi_1^2$, also ist $E[Y_i = 1]$ und $\mathrm{Var}[Y_i] = 2$, $i = 1, \ldots, v$. Der zentrale Grenzwertsatz für $Y = \sum_{i=1}^{v} Y_i$ liefert

$$Z = \frac{Y - vE[Y_i]}{\sqrt{v\mathrm{Var}[Y_i]}} = \frac{Y - v}{\sqrt{2v}} \sim^{approx.} \mathcal{N}(0, 1).$$

Eine Approximation der Wahrscheinlichkeit ergibt sich daher durch

$$P\left[\left|\frac{Y}{v} - 1\right| \leq \frac{3}{4}\right] = P\left[\left|\frac{Y-v}{v}\right| \leq \frac{3}{4}\right] = P\left[\left|\frac{Y-v}{\sqrt{2v}}\right| \leq \frac{3}{4}\sqrt{\frac{v}{2}}\right]$$
$$\simeq \varphi\left(\frac{3}{4}\sqrt{\frac{v}{2}}\right) - \varphi\left(-\frac{3}{4}\sqrt{\frac{v}{2}}\right).$$

Für $v = 12$ erhalten wir

$$P\left[\left|\frac{Y}{v} - 1\right| \leq \frac{3}{4}\right] \simeq \varphi\left(\frac{3}{4}\sqrt{6}\right) - \varphi\left(-\frac{3}{4}\sqrt{6}\right)$$
$$= 2\varphi\left(\frac{3}{4}\sqrt{6}\right) - 1$$

und mit Hilfe der Abkürzung $\sqrt{6} \simeq \sqrt{6.25} = 2.5$ folgt $\frac{3}{4}\sqrt{6} \simeq \frac{15}{8} = 1.875$, wodurch wir den konkreten Wert wie folgt erhalten:

$$P\left[\left|\frac{Y}{v} - 1\right| \leq \frac{3}{4}\right] \simeq 2\varphi(1.875) - 1 \simeq 2 \cdot 0.9696 - 1 = 0.9392.$$

10.3 Aufgabe 3 - Verteilung und Dichte

Peter und Hans sitzen pro Woche die zufälligen Zeiten P und H über ihren Stochastikserien. Peters Zeitaufwand P kann durch eine Uniform(0, 1)-verteilte Zufallsvariable beschrieben werden. Hans braucht gegeben Peters Zeit P die zufällige Zeit H, die uniform auf $(0, P)$-verteilt ist.
a) Bestimme die gemeinsame Dichte von H und P, und skizzieren den Bereich, auf dem die Dichte positiv ist.
Da $P \sim \mathcal{U}(0, 1)$, gilt $f_P(p) = 1_{[0,1]}(p)$. Da $H|P \sim \mathcal{U}(0, P)$, gilt $f_P(h|p) = \frac{1}{p}1_{[0,p]}(h)$ wobei für beide Dichtefuntkionen gilt, dass $1_{[a,b]} := 1$, falls $a \leq x \leq b$, 0 sonst. Somit verschwindet $f_{H|P}$ ausserhalb des Dreiecks

$$D := \{(p, h) \,|\, 0 < h < p \leq 1\}.$$

Für die gemeinsame Dichte von P und H ergibt sich:

$$f_{P,H}(p, h) = f_{H|P}(h|p) f_P(p) = \frac{1}{p}1_{[0,p]}(p)$$

Der Bereich D, wo $f_{H,P}$ positiv ist, bildet ein Dreieck von der x-Achse von 0 bis 1 und der Geraden $x = y$, ebenfalls von 0 bis 1.
b) Bestimmen Sie die Randdichte von H.

Die Randdichte von H berechnet sich für $0 < h \leq 1$ als Integral über die gemeinsame Dichte, also

$$f_H(h) = \int_{-\infty}^{\infty} f_{P,H}(p,h)dp = \int_0^1 \frac{1}{p} 1_{[0,p]}(h)dp = \int_h^1 \frac{1}{p}dp = -\log h.$$

c) Bestimme den Erwartungswert von H. Vergleiche diesen mit dem Erwartungswert von P.
Mit partieller Integration folgt

$$\begin{aligned}\mathrm{E}[H] &= -\int_0^1 h \log h \, dh = -\frac{h^2}{2} \log h\Big|_0^1 + \int_0^1 \frac{h^2}{2}\frac{1}{h}dh \\ &= 0 + \int_0^1 \frac{h^2}{2}\frac{1}{h}dh = \frac{h^2}{4}\Big|_0^1 = \frac{1}{4}.\end{aligned}$$

Anderseits ist $\mathrm{E}[P] = \frac{1}{2}$, also $\mathrm{E}[H] = \frac{\mathrm{E}[P]}{2}$.
d) Bestimme die Kovarianz zwischen H und P. Sind H und P unabhängig?
Man hat

$$\begin{aligned}\mathrm{E}[PH] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} ph f_{P,H}(p,h)\, dh\, dp \quad = \int_0^1 dp \int_0^p h\, dh\, dp \\ &= \int_0^1 \frac{p^2}{2}dp = \frac{1}{6},\end{aligned}$$

und somit

$$\mathrm{Cov}(P,\, H) = \mathrm{E}[PH] - \mathrm{E}[P]\mathrm{E}[H] = \frac{1}{6} - \frac{1}{2}\frac{1}{4} = \frac{1}{24}.$$

Da $\mathrm{Cov}(P,\, H) \neq 0$ ist, sind H und P somit nicht unabhängig.

10.4 Aufgabe 4 - Test

Unterhalb einer Kläranlage wurden 16 unabhängige Wasserproben aus einem Fluss entnommen und jeweils deren Ammoniumkonzentration X_i (angegeben in μgNH_4-N$/l$) mit einem Messgerät bestimmt. Der Mittelwert der Proben ergab $\overline{x} = 204.2$. Wir wollen wissen, ob mit diesem Experiment eine Überschreitung des Grenzwertes von 200 μgNH_4-N$/l$ nachgewiesen werden kann (auf dem 5%-Niveau).
a) Angenommen die Standardabweichung sei im Voraus bekannt und betrage 10 μgNH_4-N$/l$. Finde einen geeigneten statistischen Test um zu prüfen, ob eine Grenzwertüberschreitung nachgewiesen werden kann. Wie lauten die Modellannahmen?
Da die Standardabweichung der Messungen bekannt ist, entscheiden wir uns für einen z-Test. Wir nehmen an, dass die Ammoniumkonzentrationen X_i unabhängig und identisch normalverteilt sind für $i = 1, ..., 16$, mit $\sigma^2 = 100$ und μ unbekannt.
b) Führe den Test aus Punkt a) durch. Gibt dazu folgendes explizit an: H_0 und H_A, die Teststatistik, den realisierten Wert der Teststatistik, den Verwerfungsbereich sowie das Testergebnis.

- $H_0 : X_i$ i.i.d., $\sim \mathcal{N}(\mu_0,\, \sigma^2)$ mit $\mu_0 = 200$, $\sigma = 10$
- $H_A : X_i$ i.i.d., $\sim \mathcal{N}(\mu,\, \sigma^2)$ mit $\mu > 200$, $\sigma = 10$ (einseitig)
- Teststatistik: $Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,\, 1)$ unter H_0
- Verwerfungsbereich: Wir suchen $\mathcal{K} = \{z_{1-\alpha} \leq z\}$, also der Bereich, wo H_0 verworfen wird. Aus der Normalverteilungstabelle erhalten wir 1.64 für 0.9495 und 1.65 für .9505 (in unserem Fall ist $z_{1-\alpha} = 1 - 0.05 = 0.95$), daher ist unser Verwerfungsbereich $\mathcal{K} = [1.645,\, \infty)$
- Wert der Teststatistik: $z = \frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{204.2-200}{10/\sqrt{16}} = \frac{4.2}{2.5} = 1.68$
- Testentscheid: Da $1.68 \in \mathcal{K}$ wird die Nullhypothese H_0 verworfen. Die Überschreitung des Grenzwertes ist auf dem 5%-Niveau signifikant.

c) Wir wahrscheinlich ist es, dass man mit 16 unabhängigen Wasserproben eine Grenzwertüberschreitung nachweisen kann, wenn die wahre Ammoniumkonzentration tatsächlich über dem Grenzwert und zwar bei 205 μgNH_4-N$/l$ liegt?
Wir fragen nach der Wahrscheinlichkeit, dass $T \in \mathcal{K}$ mit $\mu_A = 205\,\mu$gNH_4-N$/l$. Mit anderen Worten kann H_0 verworfen werden, falls der Mittelwert der Messungen $\overline{X} \geq 204.2$ (oder grösser 204.1) ist. Wir gehen dazu wieder zu einer standardisierten normalverteilten Zufallsvariable über. Mit $\mu_A = 205$ und $\sigma 10$ erhält man

$$\begin{aligned}\mathrm{P}_{\mu_A}[\overline{X} > 204.1] &= \mathrm{P}_{\mu_A}\left[\frac{\overline{X}-\mu_A}{\sigma/\sqrt{n}} > \frac{204.1-\mu_A}{\sigma/\sqrt{n}}\right] \\ &= \mathrm{P}_{\mu_A}\left[\frac{\overline{X}-\mu_A}{\sigma/\sqrt{n}} > -0.36\right] = \mathrm{P}[Z > -0.36].\end{aligned}$$

Diese Wahrscheinlichkeit ist wegen der Symmetrie der Normalverteilung gleich $\mathrm{P}[Z \leq 0.36]$. Aus der Tabelle der Normalverteilungen entnehmen wir, dass dies gleich 0.6406 ist. Die gesuchte WHS (oder auch die Macht des Tests an der Stelle $\mu = \mu_A = 205$) ist also ungefähr 64%.
d) Wir wahrscheinlich ist es, dass man mit 16 unabhängigen Wasserproben fälschlicherweise eine Grenzüberschreitung nachweist, obwohl die wahre Ammoniumkonzentration bei 200 μgNH_4-N$/l$ liegt und den Grenzwert somit genau einhält?
Das ist genau das Niveau des Tests und war mit 5% in der Aufgabenstellung gegeben.

10.5 Aufgabe 5 - Schätzer und p-Wert

Die WHS, dass an einem beliebigen Tag in einer Grossstadt erhöhte (d.h. einen kritischen Wert überschreitende) Schadstoffemissionen gemessen werden, sei p. Zusätzlich nehmen wir an, dass die Emissionshöhen an unterschiedlichen Tagen unabhängig sind.
a) Sei in einem Zeitraum von n Tagen X_n die Anzahl derjenigen Tage, an denen erhöhte Emissionswerte gemessen werden. Welche Verteilung hat X_n? Wie lauten die Parameter der Verteilung?
$X_n \sim Bin(n,\, p)$. Genau gesagt ist $X_n = \sum_{i=1}^n B_i$, wobei $B_i \sim Ber(p)$ und i.i.d. und daher $X \sim Bin(n,\, p)$.
b) Innerhalb der letzten 345 Tage wurden an 207 Tagen erhöhte Emissionswerte gemessen. Finde zuerst einen vernünftigen Schätzer $\hat{p}$ für p und bestimme unter einer Normalapproximation ein $99\% - Vertrauensintervall$ für $\hat{p}$. Welche realisierten Ergebnisse findet man für Schätzwert und Vertrauensintervall?
Ein Schätzer für p ist $\hat{p} = \frac{1}{n}X_n = \frac{207}{345} = \frac{69}{115} = \frac{3}{5} = 0.6$. Laut **Normalapproxiamtion** ist $X_n \sim^{approx} \mathcal{N}(np,\, np(1-p))$, also

$$\hat{p} \sim^{approx} \mathcal{N}\left(p,\, \frac{1}{n}p(1-p)\right)$$

und damit ein approximatives Vertrauensintervall gegeben durch

$$\hat{p} \pm \phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\hat{p}\,(1-\hat{p})\,\frac{1}{n}},$$

wobei $\alpha = 0.01$. Weiter ergibt sich das realisierte Intervall mit $\hat{p}(\omega) = \frac{1}{n}X_n(\omega)$, $X_n(\omega) = 207$ und $n = 345$ durch

$$0.6 \pm \phi^{-1}(0.995)\sqrt{0.6 \cdot 0.4 \cdot \frac{1}{345}}.$$

Mit dem Wer $\phi^{-1}(0.995) = 2.575$ erhält man mit Hilfe der Approximationen

$$0.6 \pm 2.575\sqrt{0.6 \cdot 0.4 \cdot \frac{1}{345}} = \cdots = [0.5485,\, 0.6515]$$

als realisiertes approximatives 99%-Vertrauensintervall für p.
c) Neue Massnahmen werden implementiert, von denen man sich eine Reduktion von p auf 0.4 erhofft. 100 Tage nach der Implementierung soll daher versucht werden, statistisch zu belegen, dass p auf unter 0.4 gefallen ist. Formuliere die für die Belegung der Aussage geeigneten Null- und Alternativhypothesen und bestimme mittels Normalapproximation den maximalen approximativen Verwerfungsbreich für das Signifikanzniveau 0.05.
Zur Belegung der Aussage wählt man folgende Hypothesen:

- $H_0 : p = 0.4$
- $H_A : p < 0.4$

Der Verwerfungsbereich ist $\mathcal{K} = \{x : x \leq c\}$, wobei $c \in \mathbb{N}_0$ so gross wie möglich mit $\mathrm{P}_{0.4}[X_{100} \leq c] \leq 0.05$ ist. Zur Berechnung von $\mathrm{P}[X_{100} \leq c]$ verwendet man die Normalapproxiamtion. Unter Berücksichtigung der Kontinuitätskorrektur ergibt sich:

$$\begin{aligned}\mathrm{P}_{0.4}[X_{100} \leq c] &= \mathrm{P}_{0.4}\left[\frac{X_{100}-40}{\sqrt{0.4 \cdot 0.6 \cdot 100}} \leq \frac{c-40}{\sqrt{0.4 \cdot 0.6 \cdot 100}}\right] \\ &\simeq \phi\left(\frac{c-40+0.5}{\sqrt{0.4 \cdot 0.6 \cdot 100}}\right).\end{aligned}$$

Somit benötigen wir $\phi\left(\frac{c-39.5}{\sqrt{0.4 \cdot 0.6 \cdot 100}}\right) \leq 0.05$, beziehungsweise

$$1 - \phi\left(\frac{c-39.5}{\sqrt{0.4 \cdot 0.6 \cdot 100}}\right) = \phi\left(\frac{39.5-c}{\sqrt{0.4 \cdot 0.6 \cdot 100}}\right) \geq 0.95,$$

oder

$$c \leq 39.5 - \phi^{-1}(0.95)\sqrt{0.4 \cdot 0.6 \cdot 100}^{\phi^{-1}(0.95)} \simeq 39.5 - 1.645\sqrt{25}.$$

Somit ist der maximale approximative Verwerfungsbereich $\{x : x \leq 31\}$.