

Vorlesungsmitschrift

STOCHASTIK

Mitschrift von

Falk-Jonatan Strube

Vorlesung von

Prof. Dr. Fabian Schwarzenberger

6. Juli 2017

INHALTSVERZEICHNIS

1	Wahrscheinlichkeitsrechnung	6
1.1	Zufallsexperimente, Ereignisse und Wahrscheinlichkeiten	6
1.1.1	Zufallsexperimente und Ereignisse	6
1.1.2	Definition der Wahrscheinlichkeit	8
1.1.3	Laplacesches Modell	9
1.1.4	Kombinatorik	10
1.1.5	Bedingte Wahrscheinlichkeit	11
1.1.6	Unabhängigkeit	16
1.2	Zufallsvariablen	19
1.2.1	Grundlagen	19
1.2.2	Diskrete und stetige Zufallsvariable	21
1.2.3	Erwartungswert und Varianz	26
1.2.4	Kovarianz und Unabhängigkeit	28
1.3	Spezielle Verteilungen	31
1.3.1	Spezielle diskrete Verteilungen	31
1.3.1.1	Bernoulli Verteilung	31
1.3.1.2	Binomialverteilung	33
1.3.1.3	Diskrete Gleichverteilung	34
1.3.1.4	Hypergeometrische Verteilung	34
1.3.1.5	Geometrische Verteilung	35
1.3.1.6	Poisson-Verteilung	37
1.3.2	Spezielle stetige Verteilungen	38
1.3.2.1	Stetige Gleichverteilung	38
1.3.2.2	Normalverteilung	39
1.3.2.3	Exponentialverteilung	42
1.3.2.4	χ^2 -Verteilung	43
1.3.2.5	t -Verteilung	44
1.3.2.6	F -Verteilung	45
1.4	Grenzwertsätze	45
1.4.1	Gesetz der Großen Zahlen	46
1.4.2	Der zentrale Grenzwertsatz	47
2	Statistik	51
2.1	Deskriptive Statistik	51
2.1.1	Grundbegriffe	51
2.1.2	Eindimensionales Datenmaterial	54
2.1.2.1	Stichprobenfunktionen	54
2.1.2.2	Aufbereitung statistischer Daten am Beispiel	56
2.1.3	Zweidimensionales Datenmaterial	56
2.1.3.1	Anschauung am Beispiel	56
2.1.3.2	Stichprobenfunktionen für 2-dimensionale Merkmale	64
2.2	Schätztheorie	66
2.2.1	Einführung	66
2.2.2	Punktschätzer	67
2.2.2.1	Eigenschaften von Punktschätzern	67



2.2.3	Bereichsschätzer	68
2.2.3.1	Problemstellung	68
2.2.3.2	Konfidenzintervalle bei Normalverteilung	70
2.2.3.3	Konfidenzintervall für eine unbekannte Wahrscheinlichkeit . . .	70
2.2.4	Bootstrapping	71
2.3	Testtheorie	73
2.3.1	Grundidee	73
2.3.2	Parametertests	75
2.3.2.1	Tests für μ und σ unter Normalverteilung	75
2.3.2.2	Tests für 2 unabhängige Stichproben unter Normalverteilung . .	76
2.3.2.3	Tests für unbekannte Wahrscheinlichkeiten	77
2.3.3	Parameterfreie Tests	78
2.3.3.1	chi-quadrat-Unabhängigkeitstest	78



VORBEMERKUNG

Lernraum: Dienstag 17:00 S327, S329



STOCHASTIK

WAS IST STOCHASTIK

Stochastik...

- ... kommt etymologisch aus dem Griechischem; Bedeutung: „Kunst des Vermutens“
- ... beschäftigt sich mit der Beschreibung und dem Untersuchen von zufälligen Ereignissen (z.B. Lotto, Wurf eines Würfels, Lebensdauer einer Glühbirne, ...)
- ... beinhaltet die Teilgebiete
 - Wahrscheinlichkeitsrechnung:
Zu Grunde liegende Gesetzmäßigkeit des Zufalls bekannt. Frage nach Wahrscheinlichkeiten „interessanter“ Ereignisse

Bsp. Würfel: Jede Seite fällt mit Wahrscheinlichkeit $\frac{1}{6}$.
Wie groß ist die Wahrscheinlichkeit, dass unter 10 Würfeln mindestens 4 mal 6 kommt?
 - Statistik:
Zur Grunde liegende Gesetzmäßigkeit des Zufalls ist unbekannt. Idee: Nutze Stichproben/Daten um diese Gesetzmäßigkeiten zu erkennen.

Bsp.: Gesamtproduktion 100 000 Teile, Stichprobe von 100 Teilen enthält 2 defekte.
Kann davon ausgegangen werden, dass die Fehlerquote von 1% nicht eingehalten wird?



1 WAHRSCHEINLICHKEITSRECHNUNG

1.1 ZUFALLSEXPERIMENTE, EREIGNISSE UND WAHRSCHEINLICHKEITEN

1.1.1 ZUFALLSEXPERIMENTE UND EREIGNISSE

Erster wichtiger Begriff:

1.1-1 Definition Ein ZUFALLSEXPERIMENT ist ein Vorgang

- der beliebig oft unter gleichartigen Bedingungen wiederholt werden kann und
- dessen Ergebnis nicht mit Sicherheit vorhergesagt werden kann

$\Omega :=$ Ergebnismenge (oder Ergebnisraum) ist die Menge aller möglichen Ergebnisse

1.1-2 Bemerkung Drei wichtige Fälle

- Ω endlich, d.h. $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$
- Ω abzählbar unendlich, d.h. $\Omega = \{\omega_1, \omega_2, \dots\}$ (Ereignisse lassen sich mit den natürlichen Zahlen aufzählen)¹
- Ω überabzählbar unendlich, d.h. $\Omega = \mathbb{R}$ oder $\Omega[0, 1)$

1.1-3 Beispiel

- Würfel: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Anzahl der defekten Glühbirnen in einer Stichprobe von 100 Stück: $\Omega = \{0, 1, 2, \dots, 100\}$
- Anzahl der Anrufe im Call-Center zwischen 8:00 und 9:00
 - (1) Möglichkeit 1: $\Omega = \{0, 1, 2, \dots\} = \mathbb{N}_0$
 - (2) Möglichkeit 2: $\Omega = \{\omega_1, \omega_2, \dots, \omega_{100}\}$ mit
$$w_i = \begin{cases} i \text{ Anrufe, falls } i \leq 99 \\ 100 \text{ oder mehr Anrufe, falls } i = 100 \end{cases}$$
- Downloadzeit einer Datei: $\Omega = (0, \infty)$

Wir interessieren uns oft nicht allein für das Eintreten von einem $w \in \Omega$, sondern dafür ob ein w aus einer gewissen Teilmenge aus Ω eingetreten ist (z.B. sind weniger als 3 Glühbirnen defekt). Daher:

¹ zu natürlichen Zahlen (in dieser VL): $\mathbb{N} = \{1, 2, 3, \dots\}$, $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$



1.1-4 Definition Ein ZUFÄLLIGES EREIGNIS A ist eine Teilmenge des Ergebnisraums Ω .
 Spezielle Ereignisse:

- $A = \emptyset$... das unmögliche Ereignis ($\omega \in \emptyset$ tritt nie ein)
- $A = \Omega$... das sichere Ereignis ($\omega \in \Omega$ tritt immer ein)
- $A = \{\omega\}$... Elementarereignis (für ein $\omega \in \Omega$)
- $\bar{A} = \Omega \setminus A$... Gegenereignis zu A

Sprechweise: „Das Ereignis A tritt ein“, falls ein $\omega \in A$ beobachtet wird.

1.1-5 Beispiel (Würfel)

$A = \{\text{„gerade Zahl fällt“}\}$

$\Rightarrow A = \{2, 4, 6\} \subseteq \Omega = \{1, \dots, 6\}$

Gegenereignis: $\bar{A} = \{1, 3, 5\}$

1.1-6 Bemerkung Da Ereignisse Teilmengen von Ω sind, lassen sich alle Rechenoperationen für Mengen anwenden. Seien $A, B \subseteq \Omega$.

- $A \subseteq B$... A ist Teilereignis von B
- $A = B$, gleiche Ereignisse
- Durchschnitt: $A \cap B$, „ A und B “ (beide Ereignisse treten gleichzeitig ein)
- Vereinigung: $A \cup B$, „ A oder B “ (entweder A oder B treten ein)
- Differenz: $A \setminus B$, „ A ohne B “ (A tritt ein, B aber nicht)
- Negation/Gegenereignis: $\bar{A} = \Omega \setminus A$ (A tritt nicht ein)
- gilt $A \cap B = \emptyset$, so heißen A und B UNVEREINBAR/DISJUNKT.

1.1-7 Beispiel (Würfel)

$\Omega = \{1, \dots, 6\}$,

$A = \{2, 4, 6\}$, $B = \{2, 3, 5\}$, $C = \{1, 3\}$

Bestimme: $A \cup B$, $A \cap B$, $A \cap C$, $C \cup \bar{C}$

$A \cup B = \{2, 3, 4, 5, 6\}$

$A \cap B = \{2\}$

$A \cap C = \emptyset$

$C \cup \bar{C} = \Omega$

1.1-8 Satz (Rechenregeln) Es seien A , B und C Ereignisse. Dann gilt:

- $A \cap B = B \cap A$ $A \cup B = B \cup A$ (Kommutativgesetz)
- $A \cap (B \cap C) = (A \cap B) \cap C$
 $A \cup (B \cup C) = (A \cup B) \cup C$ (Assoziativgesetze)
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
 $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (Distributivgesetze)
- $\overline{A \cap B} = \bar{A} \cup \bar{B}$
 $\overline{A \cup B} = \bar{A} \cap \bar{B}$ (De Morgansche Regeln)
- aus $A \subseteq B$ folgt $\bar{B} \subseteq \bar{A}$ und $A \setminus B = A \cap \bar{B}$



1.1-9 Definition Sei Ω eine Menge. Ein Mengensystem $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ heißt σ -Algebra, falls gilt

- $\Omega \in \mathcal{A}$
- $A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$
- $A_1, A_2, A_3, \dots \in \mathcal{A} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$

(Sprich: die Menge, alle Komplemente und die Schnitte und Vereinigungen aller Teilmengen müssen in $\mathcal{A}$ liegen [Mächtigkeit der σ -Algebra ist bei einer endlichen Grundmenge immer eine 2er-Potenz!])

1.1-10 Bemerkung Sei $\mathcal{A}$ eine σ -Algebra auf Ω . Dann gilt:

- $\emptyset \in \mathcal{A}$
- $A, B \in \mathcal{A} \Rightarrow A \setminus B \in \mathcal{A}$
- $A_1, A_2, A_3, \dots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

1.1-11 Beispiel (Würfel)

- $\mathcal{A} = \{\{1\}, \{2\}, \emptyset, \{1, 2\}, \{3, 4, 5, 6\}, \{2, 3, 4, 5, 6\}, \{1, 3, 4, 5, 6\}, \{1, 2, 3, 4, 5, 6\}\}$ ist eine σ -Algebra über $\Omega = \{1, \dots, 6\}$
- $\mathcal{A} = \{A \mid A \subseteq \Omega\} = \mathcal{P}(\Omega)$ ist auch ein σ -Algebra

1.1-12 Bemerkung Besteht Ω aus n Elementen, so enthält $\mathcal{P}(\Omega)$ genau 2^n Elemente.

1.1.2 DEFINITION DER WAHRSCHEINLICHKEIT

Ziel: Ordne zufälligem Ereignis A eine Wahrscheinlichkeit $\mathbb{P}(A)$ zu, die die Chance beurteilt, dass A eintritt.

1.1-13 Definition (Kolmogorov 1933)

Gegeben sei eine Ereignismenge Ω und eine σ -Algebra $\mathcal{A}$. Eine Funktion $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ heißt WAHRSCHEINLICHKEITSMASS AUF $(\Omega, \mathcal{A})$, falls

- (1) $\mathbb{P}(\Omega) = 1$
- (2) für paarweise disjunkte $A_i \in \mathcal{A}$, $i = 1, 2, \dots$ (d.h. $A_i \cap A_j = \emptyset$ falls $i \neq j$) gilt $\mathbb{P}(A_1 \cup A_2 \cup \dots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots$

Weitere Bezeichnungen:

- $\mathbb{P}(A)$... Wahrscheinlichkeit des Ereignisses A
- $(\Omega, \mathcal{A}, \mathbb{P})$... Wahrscheinlichkeitsraum / Wahrscheinlichkeitsmodell



1.1-14 Bemerkung (Allgemeines Vorgehen, Vereinfacht Darstellung)

- (1) Theoretische Untersuchungen (Kombinatorik, physikalische Gesetze); Beobachtung der relativen Häufigkeit (deskriptive Statistik)
- (2) Schritt (1) liefert für gewisse Grundereignisse die Wahrscheinlichkeiten (exakt oder zumindest näherungsweise)
- (3) Bestimmen der Wahrscheinlichkeiten für alle interessierenden Ereignisse (mittels Rechenregeln, siehe später)

1.1-15 Satz Seien A, B, C sowie $A_1, A_2, \dots, A_n$ zufällige Ereignisse. Dann gilt:

- (a) $\mathbb{P}(\emptyset) = 0$
- (b) $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$
- (c) $A \subseteq B \Rightarrow \mathbb{P}(A) \subseteq \mathbb{P}(B)$
- (d) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
 $\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C)$
(Gut zu veranschaulichen durch Venn-Diagramme)

Es gilt sogar der Additionssatz:

1.1-16 Satz Seien $A_1, \dots, A_n$ zufällige Ereignisse. Dann gilt:

$$\mathbb{P}(A_1 \cup \dots \cup A_n) = \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{i < j} \mathbb{P}(A_i \cap A_j) + \sum_{i < j < k} \mathbb{P}(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} \mathbb{P}(A_1 \cap \dots \cap A_n)$$

1.1.3 LAPLACESCHES MODELL

Nun wollen wir ein spezielles, einfaches, aber oft sehr nützliches WK-Maß einführen.

1.1-17 Definition Ein WK-Modell $(\Omega, \mathcal{A}, \mathbb{P})$ heißt LAPLACESCHES MODELL, falls

$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ endlich ist, $\mathcal{A} = \mathcal{P}(\Omega)$ und $\mathbb{P}(\{\omega_1\}) = \mathbb{P}(\{\omega_2\}) = \dots = \mathbb{P}(\{\omega_n\}) = \frac{1}{n}$ gilt.

1.1-18 Bemerkung für beliebiges $A \in \mathcal{A}$ gilt im Laplaceschen Modell:

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{m}{n}, \text{ wobei } m = |A| \text{ die Anzahl der Elemente in } A \text{ ist (und } |\Omega| = n).$$

Also $\mathbb{P}(A) = \frac{\text{Anzahl der günstigen Elementarereignisse}}{\text{Anzahl der möglichen Elementarereignisse}}.$

Man sagt auch: $\mathbb{P}$ ist dann die diskrete Gleichverteilung auf Ω .

1.1-19 Beispiel

- (a) (fairer Würfel) Wie groß ist die WK eine Zahl größer 4 zu würfeln?

$$\Omega = \{1, 2, 3, 4, 5, 6\}, A = \{5, 6\} \text{ und es gilt } \mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \dots = \mathbb{P}(\{6\}) = \frac{1}{6}$$

$$\text{Daher: } \mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{2}{6} = \frac{1}{3}$$



(b) (2 faire Würfel) Wie groß ist die WK mit 2 Würfeln mindestens eine 11 zu würfeln?

$$\Omega = \{(1, 1), (1, 2), \dots, (1, 6) \\ \vdots \\ (6, 1), (6, 2), \dots, (6, 6)\}$$

$A = \{(6, 5), (5, 6), (6, 6)\}$ und es gilt $\mathbb{P}(\{(i, j)\}) = \frac{1}{36}$ für beliebiges $i, j \in \{1, \dots, 6\}$. Also liegt Laplace Modell vor.

Daher gilt: $\mathbb{P}(A) = \frac{3}{36} = \frac{1}{12}$.

Um in Laplace-Modellen die Größe (Mächtigkeit) von Ereignissen zu bestimmen, sind oft spezielle „Abzähltricks“ sinnvoll. Diese liefert die Kombinatorik.

1.1.4 KOMBINATORIK

Fragestellung: Wie viele Möglichkeiten gibt es aus einer n -elementigen Menge k Elemente auszuwählen? Dabei sind die Spielregeln zu klären:

- Spielt die Reihenfolge eine Rolle?
- Dürfen Elemente mehrfach ausgewählt werden (mit Zurücklegen oder ohne)?

1.1-20 Satz In einer Urne befinden sich n (voneinander unterscheidbare) Elemente. Wir ziehen k davon...

(a) ... mit Zurücklegen, unter Berücksichtigung der Reihenfolge, dann gibt es

$$\bar{v}_n^k = n^k$$

viele Möglichkeiten (Variation von n Elementen zur k -ten Klasse mit Wiederholungen).

(b) ... ohne Zurücklegen, unter Berücksichtigung der Reihenfolge, dann gibt es

$$v_n^k = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-(k-1)) = \frac{n!}{(n-k)!}$$

viele Möglichkeiten (Variation von n Elementen zur k -ten Klasse ohne Wiederholungen).

(c) ... mit Zurücklegen, ohne Berücksichtigung der Reihenfolge, dann gibt es

$$\bar{c}_n^k = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$$

viele Möglichkeiten (Kombination von n Elementen zur k -ten Klasse mit Wiederholungen).

(d) ... ohne Zurücklegen, ohne Berücksichtigung der Reihenfolge, dann gibt es

$$c_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

viele Möglichkeiten (Kombination von n Elementen zur k -ten Klasse ohne Wiederholungen).



1.1-21 Bemerkungen

- $n! = n \cdot (n-1) \cdot \dots \cdot 2 \cdot 1$ mit $0! = 1$
- Spezialfall in (b): $n = k$, dann $v_n^k = n!$. Dies beschreibt die Anzahl der möglichen Anordnungen von n Elementen (Permutationen).
- Spezialfälle in (d):
 - $n = k$, dann $c_n^k = 1 = \binom{n}{n}$
 - $k = 0$, dann $c_n^0 = \binom{n}{0} = 1$
 - $k = 1$, dann $c_n^1 = \binom{n}{1} = n$

1.1-22 Beispiel

- (a) Wie viele mögliche Zieleinläufe gibt es beim 100m-Lauf mit 8 Teilnehmern?
 $8! = 40320$
- (b) Wie viele Möglichkeiten gibt es beim Lotto (6 aus 49)
 $\binom{49}{6} = 13\,983\,816$
- (c) Wie viele Möglichkeiten gibt es ein Nummernschild der Art „DD-Buchstabe Buchstabe Ziffer Ziffer Ziffer“ zu konstruieren?
 $26^2 \cdot 10^3 = 676\,000$
- (d) Wie viele Möglichkeiten gibt es 5 (nicht unterscheidbare) Äpfel auf 3 Kinder aufzuteilen?
 $\binom{3+5-1}{5} = \binom{7}{1} = 21$

1.1.5 BEDINGTE WAHRSCHEINLICHKEIT

Frage: Wie verändert sich die Wahrscheinlichkeit eines Ereignisses, falls ich Zusatzwissen mit einfließen lasse?

1.1-23 Beispiel HIV Prävalenz liegt weltweit bei 0,8%, also:

$\mathbb{P}_1(\{\text{zufällig ausgewählte Person ist HIV-positiv}\}) = 0,008$

Modell 1: $\Omega = \{0, 1\}$, $\mathbb{P}_1(\{1\}) = 0,008$, $\mathbb{P}_1(\{0\}) = 0,992$

Zusatzwissen: ausgewählte Person ist Europäer und Prävalenz in Europa: = 0,2%, also:

$\mathbb{P}_2(\{\text{zufällig ausgewählte Person ist HIV-positiv}\}) = 0,002$

Modell 2: $\Omega = \{0, 1\}$, $\mathbb{P}_2(\{1\}) = 0,002 = 1 - \mathbb{P}(\{0\})$

Problem/Frage:

- Wie kombiniert man beide Modelle?
- Wir wollen nicht mit 2 verschiedenen $\mathbb{P}$ s rechnen.
- WK für HIV positiv unter Nicht-Europäern?



1.1-24 Beispiel

- Von insgesamt 800 Schülern besitzen 440 ein Smartphone.
- Unter den Smartphone-Besitzern sind 60% männlich.
- Unter den Nicht-Smartphone-Besitzern sind 35% männlich.
- Unter allen 800 Schülern wird ein Smartphone verlost.

Fragen:

- Wie groß ist die Wahrscheinlichkeit, dass der Gewinner bereits ein Smartphone besitzt?
- Wie groß ist die WK, dass der Gewinner bereits ein Smartphone besitzt, wenn man schon weiß, dass ein Mädchen gewonnen hat?

1.1-25 Definition Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wk-Raum und seien $A, B \subset \Omega$ Ereignisse mit $\mathbb{P}(B) > 0$. Dann definieren wir

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

und nennen $\mathbb{P}(A|B)$ die Wahrscheinlichkeit von A bedingt auf B .

Interpretation: „Wie groß ist die Wk, von A , wenn ich schon weiß, dass B eingetreten ist?“

1.1-26 Beispiel (Smartphone, s.o.)

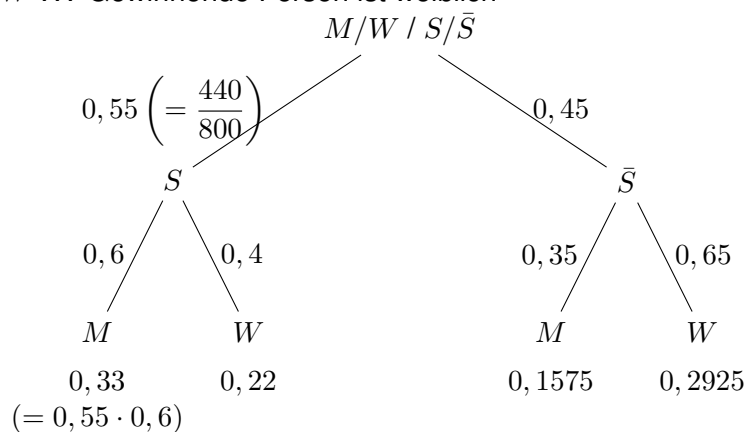
$\Omega = \{(S, M), (\bar{S}, M), (S, W), (\bar{S}, W)\}$

S ... Gewinnende Person besitzt Smartphone

$\bar{S}$... Gewinnende Person besitzt kein Smartphone

M ... Gewinnende Person ist männlich

W ... Gewinnende Person ist weiblich



gegeben:

$$\mathbb{P}(\{(S, M)\}) = 0,33$$

$$\mathbb{P}(\{(S, W)\}) = 0,22$$

$$\mathbb{P}(\{(\bar{S}, M)\}) = 0,1575$$

$$\mathbb{P}(\{(\bar{S}, W)\}) = 0,2925$$

Antwort auf Fragen:

- 0,55 (klar)



- (b) Intuition: Wir wissen, dass nur noch die Stränge mit „W“ interessieren. Die Stränge ohne „W“ sollten wir „streichen“. Wie groß ist die WK der Kombination (S,W) im Vergleich zu allen, wo W vorkommt? Also:

$$\frac{\mathbb{P}(\{(S, W)\})}{\mathbb{P}(\{(S, W), (\bar{S}, W)\})} = \frac{0,22}{0,22 + 0,2925} = 0,4293$$

Was hat das mit der bedingten WK aus Def. 1.1.25 zu tun?

$A := \{\text{Person besitzt Smartphone}\} = \{(S, M), (S, W)\}$

$B := \{\text{Person ist weiblich}\} = \{(S, W), (\bar{S}, W)\}$

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(\{(S, W)\})}{\mathbb{P}(\{(S, W), (\bar{S}, W)\})} = \dots = 0,4293$$

1.1-27 Satz (Rechnen mit bedingten WK)

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein WK-Raum und $A, A_1, A_2, B \in \mathcal{A}$ Ereignisse mit $\mathbb{P}(B) > 0$. Dann gilt:

- $\mathbb{P}(B|B) = 1, \mathbb{P}(\emptyset|B) = 0$
- Falls A und B disjunkt, gilt $\mathbb{P}(A|B) = 0$
- $\mathbb{P}(\bar{A}|B) = 1 - \mathbb{P}(A|B)$
- $\mathbb{P}(A_1 \cup A_2|B) = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B) - \mathbb{P}(A_1 \cap A_2|B)$
- Falls $B \subseteq A$, so gilt $\mathbb{P}(A|B) = 1$
- Falls $A \subseteq B$, so gilt $\mathbb{P}(A|B) = \mathbb{P}(A)$

1.1-28 Beispiel Auf einer E-Mail Adresse kommen im Schnitt 80% Spam-Mails und 20% gute Mails.

Eine „gute“ Mail enthalte mit 2% WK das Wort „Viagra“. In einer Spam-Mail liegt dieser Anteil bei 60%. Berechnen Sie die WK, dass eine Spam-Mail vorliegt, falls man weiß, dass das Wort „Viagra“ enthalten ist.

Lösung:

$A = \{\text{Mail enthält „Viagra“}\}$

$\bar{A} = \{\text{Mail enthält kein „Viagra“}\}$

$B = \{\text{Mail ist Spam}\}$

$\bar{B} = \{\text{Mail ist kein Spam}\}$

4-Felder-Tafel:

	B : Spam	$\bar{B}$ kein Spam	
A , mit Viagra	$0,8 \cdot 0,6 = 0,48$	$0,2 \cdot 0,002 = 0,004$	0,484
$\bar{A}$, ohne Viagra	0,32	0,196	0,516
	0,8	0,2	1

Gesucht ist $\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{0,48}{0,484} = 0,9917$ Auch interessant ist die WK, dass die Mail

kein Spam ist, wenn man schon weiß, dass „Viagra“ nicht enthalten ist. $\mathbb{P}(\bar{B}|\bar{A}) = \frac{0,196}{0,516} = 0,3798$

1.1-29 Satz (Multiplikationssatz)

Seien A und B Ereignisse mit $\mathbb{P}(A) > 0, \mathbb{P}(B) > 0$. Dann gilt:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B|A) = \mathbb{P}(B) \cdot \mathbb{P}(A|B)$$



Sind $A_1, \dots, A_n$ Ereignisse mit $\mathbb{P}\left(\bigcap_{i=1}^{n-1} A_i\right) > 0$, dann gilt sogar:

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \cdot \dots \cdot \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1})$$

1.1-30 Beispiel In einer Los-Trommel befinden sich 20 Lose. Jemand zieht 3 nacheinander. Es gibt insgesamt 5 Gewinnlose. Wie groß ist die WK, dass alle 3 gezogenen Lose Gewinnlose sind?

$A_k = \{\text{Gewinn beim } k\text{-ten Los}\}, k = 1, 2, 3$

Gesucht: $\mathbb{P}(A_1 \cap A_2 \cap A_3)$

Satz 1.1.29 liefert:

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2)$$

$$\mathbb{P}(A_1) = \frac{5}{20} = \frac{1}{4} \quad (5 \text{ Günstige in } 20 \text{ Losen})$$

$$\mathbb{P}(A_2|A_1) = \frac{4}{19}$$

$$\mathbb{P}(A_3|A_1 \cap A_2) = \frac{3}{18}$$

$$\Rightarrow \mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{1}{4} \cdot \frac{4}{19} \cdot \frac{3}{18} = \frac{1}{114} = 0,0087$$

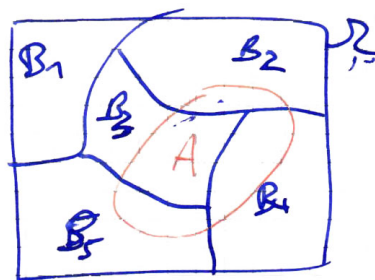
1.1-31 Satz (Formel der totalen WK)

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein WK-Raum und seien $B_1, \dots, B_n \in \mathcal{A}$ mit

- $\bigcup_{i=1}^n B_i = \Omega$
- $B_i \cap B_j = \emptyset$ für $i \neq j$
- $\mathbb{P}(B_i) > 0$ für alle $i = 1, \dots, n$

Dann gilt:

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)$$



1.1-32 Beispiel (Prävalenz von HIV)

- HIV-Prävalenz weltweit: 0,8%
- HIV-Prävalenz in Europa: 0,2%
- es gibt 7 Mrd. Menschen auf der Erde
- es gibt 740 Mio Menschen in Europa

Gesucht:

- WK, dass zufällig ausgewählter Europäer HIV-positiv ist.
- WK, dass zufällig ausgewählter Nicht-Europäer HIV-positiv ist.

Lösung:

$E := \{\text{ausgewählte Person ist Europäer}\}$

$P := \{\text{ausgewählte Person ist HIV positiv}\}$

Wir wissen:

$$\mathbb{P}(P) = 0,008, \mathbb{P}(E) = \frac{74}{700} \approx 0,1057$$

$$\mathbb{P}(P|E) = 0,002.$$

Wir wollen wissen:

- $\mathbb{P}(\bar{P}|E) = 1 - \mathbb{P}(P|E) = 1 - 0,002 = 0,998$
- $\mathbb{P}(P|\bar{E}) = \mathbb{P}(B|\bar{E}) \cdot \mathbb{P}(E) + \mathbb{P}(P|E) \cdot \mathbb{P}(\bar{E})$ (mit $B_1 = \bar{E}$ und $B_2 = E$)
Umstellen liefert:
$$\mathbb{P}(P|\bar{E}) = \frac{\mathbb{P}(B) - \mathbb{P}(P|E) \cdot \mathbb{P}(E)}{\mathbb{P}(\bar{E})} = \frac{0,008 - 0,002 \cdot 0,1057}{1 - 0,1057} = 0,008709$$

1.1-33 Satz (Formel von Bayes)

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ WK-Räume und seien $B_1, \dots, B_n \in \mathcal{A}$ mit

- $\bigcup_{i=1}^n B_i = \Omega$
- $B_i \cap B_j = \emptyset$ für $i \neq j$
- $\mathbb{P}(B_i) > 0$ für alle $i = 1, \dots, n$

Dann gilt für beliebige $A \in \mathcal{A}$ mit $\mathbb{P}(A) > 0$ und beliebiges $j \in \{1, \dots, n\}$:

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j) \cdot \mathbb{P}(B_j)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_j) \cdot \mathbb{P}(B_j)}{\sum_{i=1}^n \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)}$$

Formel von Bayes dreht also die Bedingung um.

1.1-34 Beispiel (Ziegenproblem)

In einer Spielshow steht der Kandidat vor 3 verschlossenen Türen. Eine Tür verbirgt den Hauptgewinn, ein Auto. Hinter den beiden anderen Türen sind Ziegen. Der Kandidat zeigt auf eine der Türen, der Spielleiter (der weiß, wo das Auto steht) öffnet gemäß der Spielregeln eine der beiden anderen Türen um eine Ziege zu präsentieren.

Der Kandidat darf nun seine Wahl ändern. Sollte er das tun?



Lösung:

Wir legen uns fest, dass der Kandidat Tor 1 gewählt hat und Moderator Tor 3 öffnet (ohne Beschränkung der Allgemeinheit (oBdA): sonst Umnummerieren).

Ergebnismenge: $\Omega = \{(i, j) \mid i, j = 1, 2, 3\}$ mit $(i, j) \dots$ Gewinn ist hinter Tor i , Moderator öffnet Tor j .

Definiere die Ereignisse



$G_i := \{\text{Gewinn hinter Tor } i\} = \{(i, 1), (i, 2), (i, 3)\}$ und
 $M_j := \{\text{Moderator öffnet Tor } j\} = \{(1, j), (2, j), (3, j)\}$

Wir wissen:

$$\mathbb{P}(G_i) = \frac{1}{3} \text{ für alle } i = 1, 2, 3$$

$$\mathbb{P}(M_3|G_1) = \frac{1}{2}$$

$$\mathbb{P}(M_3|G_2) = 1$$

$$\mathbb{P}(M_3|G_3) = 0$$

Gesucht: $\mathbb{P}(G_2|M_3)$

$$\begin{aligned} \mathbb{P}(G_2|M_3) &= \frac{\mathbb{P}(M_3|G_2) \cdot \mathbb{P}(G_2)}{\mathbb{P}(M_3|G_1) \cdot \mathbb{P}(G_1) + \mathbb{P}(M_3|G_2) \cdot \mathbb{P}(G_2) + \mathbb{P}(M_3|G_3) \cdot \mathbb{P}(G_3)} \\ &= \frac{1 \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} \\ &= \frac{2}{3} \end{aligned}$$

Dieses scheinbare Paradoxon ist gut zu veranschaulichen, wenn man sich nicht 3 sondern 100 Tore vorstellt. Wenn man eines der 100 auswählt und der Moderator von den restlichen 99 Toren 98 öffnet, ist offensichtlich, dass die Wahrscheinlichkeit zu gewinnen höher ist, wenn man das Tor wechselt. Die gesamten Wahrscheinlichkeiten der geöffneten Tore „sammeln“ sich hinter dem nicht geöffneten, nicht ausgewählten Tor.

1.1-35 Beispiel (Zuverlässigkeit diagnostischer Tests)

Betrachten eines Tests zum Diagnostizieren einer Krankheit. Dieser kann entweder „positiv“ oder „negativ“ sein.

Gegebene Ereignisse:

$P := \{\text{Test positiv}\} \dots$ Test tippt darauf, dass Krankheit vorliegt.

$\bar{P} := \{\text{Test negativ}\} \dots$ Test tippt darauf, dass Krankheit nicht vorliegt.

$K := \{\text{Person ist krank}\}$

$\bar{K} := \{\text{Person ist nicht krank}\}$

- Sensitivität $:= \mathbb{P}(P|K)$ (WK, dass Test „positiv“ anzeigt, wenn man tatsächlich auch krank ist. D.h. richtig-positiver Test)
- Spezifität $:= \mathbb{P}(\bar{P}|\bar{K})$ (WK, dass Test „negativ“ anzeigt, wenn man tatsächlich gesund ist. D.h. richtig-negativer Test)

	krank	gesund
Test positiv	richtig-positiv	falsch-positiv
Test negativ	falsch-negativ	richtig-negativ

Problem: Typischerweise sind Sensitivität und Spezifität gegeben, aber eigentlich interessieren uns $\mathbb{P}(K|P)$ oder $\mathbb{P}(K|\bar{P})$.

1.1.6 UNABHÄNGIGKEIT

Wir untersuchen die Frage, ob sich Ereignisse gegenseitig beeinflussen.

1.1-36 Definition Zwei Ereignisse $A, B \in \mathcal{A}$ heißen (stochastisch) unabhängig, wenn

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$



Die Ereignisse $A_1, \dots, A_n$ heißen paarweise (stochastisch) unabhängig, wenn

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j)$$

für alle $i \neq j$.

Die Ereignisse $A_1, \dots, A_n$ heißen (stochastisch) unabhängig (in ihrer Gesamtheit), wenn

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdot \dots \cdot \mathbb{P}(A_{i_k})$$

für jede beliebige Auswahl von k ($2 \leq k \leq n$) der n Ereignisse.

1.1-37 Bemerkung

(1) $A_1, \dots, A_n$ (in ihrer Gesamtheit) unabhängig $\Rightarrow A_1, \dots, A_n$ paarweise unabhängig. Rückrichtung gilt im Allgemeinen nicht (siehe Übung).

(2) Ist $\mathbb{P}(B) > 0$ so gilt A und B unabhängig $\Leftrightarrow P(A|B) = \mathbb{P}(A)$

Beweis:

$$\text{„}\Rightarrow\text{“ } \mathbb{P}(A|B) \stackrel{\text{Def.}}{=} \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A) \cdot \mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A)$$

$$\text{„}\Leftarrow\text{“ } \mathbb{P}(A \cap B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \cdot \mathbb{P}(B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

A und B unabhängig: Die WK für das Eintreten von A hängt nicht von dem Wissen, ob B bereits eingetreten ist, ab.

(3) A und B disjunkt und $\mathbb{P}(A) > 0$, $\mathbb{P}(B) > 0$.

$\Rightarrow A$ und B sind stochastisch ABHÄNGIG.

Sind A und B disjunkt, so sind sie abhängig!

denn: $\mathbb{P}(A \cap B) = 0 \neq \mathbb{P}(A) \cdot \mathbb{P}(B)$

(4) Sind A und B stochastisch unabhängig, so sind:

- A und $\bar{B}$ stochastisch unabhängig
- $\bar{A}$ und B stochastisch unabhängig
- $\bar{A}$ und $\bar{B}$ stochastisch unabhängig

Analog für mehr als zwei Ereignisse.

(5) $\emptyset$ und Ω sind zu jedem $A \in \mathcal{A}$ unabhängig.

1.1-38 Beispiel (Münze und Würfel)

Werfen faire Münze (Werte 0/1) und fairen Würfel (Werte 1, ..., 6). Untersuche auf Unabhängigkeit:

(1) $A = \{\text{Wer der Münze ist 1}\}$, $B = \{\text{Würfel} > 4\}$

(2) $A = \{\text{Wer der Münze ist 1}\}$, $C = \{(\text{Würfel} + \text{Münze}) > 4\}$

(3) $A = \{\text{Wer der Münze ist 1}\}$, $D = \{(\text{Würfel} + \text{Münze}) \in \{2, 3, 4\}\}$

Lösung:

$$\Omega = \{(0, 1), (0, 2), (0, 3), \dots, (1, 6)\}$$



$$(1) \mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{6}{12} = \frac{1}{2}, \quad \mathbb{P}(B) = \frac{|B|}{|\Omega|} = \frac{4}{12} = \frac{1}{3}$$

$$\mathbb{P}(A \cap B) = \frac{|A \cap B|}{|\Omega|} = \frac{2}{12} = \frac{1}{6}$$

$$\Rightarrow \mathbb{P}(A \cap B) = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} = \mathbb{P}(A) \cdot \mathbb{P}(B) \checkmark$$

$\Rightarrow A$ und B unabhängig.

$$(2) \mathbb{P}(C) = \frac{5}{12}$$

$$\mathbb{P}(A \cap C) = \frac{3}{12} \neq \frac{1}{2} \cdot \frac{5}{12} = \frac{5}{24} = \mathbb{P}(A) \cdot \mathbb{P}(C)$$

$\Rightarrow A$ und B nicht unabhängig, also abhängig.

$$(3) \mathbb{P}(D) = \frac{1}{2}$$

$$\mathbb{P}(A \cap D) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(A) \cdot \mathbb{P}(D) \checkmark$$

$\Rightarrow A$ und B unabhängig.

1.1-39 Beispiel Eine Maschine besteht aus 2 Bauteilen. Bauteil 1 ist mit WK 0,05 defekt, Bauteil 2 mit WK 0,02.

Wir nehmen, dass sie unabhängig voneinander ausfallen.

Frage: Wie groß ist die WK, dass mindestens 1 defekt ist?

Lösung: $\Omega = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$

$A = \{\text{Bauteil 1 defekt}\} = \{(0, 1), (1, 0)\}$

$B = \{\text{Bauteil 2 defekt}\} = \{(0, 0), (1, 0)\}$

$\mathbb{P}(A) = 0,05 \quad \mathbb{P}(B) = 0,02$

Gesucht: $\mathbb{P}(A \cup B)$

1. Variante: $\mathbb{P}(A \cup B) = 1 - \mathbb{P}(\overline{A \cup B}) = 1 - \mathbb{P}(\bar{A} \cap \bar{B}) = 1 - \mathbb{P}(\bar{A}) \cdot \mathbb{P}(\bar{B}) = 1 - 0,95 \cdot 0,98 = 0,069$

2. Variante: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 0,05 + 0,02 - 0,05 \cdot 0,02 = 0,069$

1.1-40 Satz Seien $A_1, \dots, A_n$ unabhängige Ereignisse. Dann gilt:

$$\mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_n) = 1 - \mathbb{P}(\bar{A}_1) \cdot \dots \cdot \mathbb{P}(\bar{A}_n)$$

Beweis:

$$\begin{aligned} \mathbb{P}(A_1 \cup \dots \cup A_n) &= 1 - \mathbb{P}(\overline{A_1 \cup \dots \cup A_n}) \\ &= 1 - \mathbb{P}(\bar{A}_1 \cap \dots \cap \bar{A}_n) \\ &= 1 - \mathbb{P}(\bar{A}_1) \cdot \dots \cdot \mathbb{P}(\bar{A}_n) \end{aligned}$$

1.1-41 Beispiel Drei Jäger schießen gleichzeitig und unabhängig voneinander auf Bambi.

Jäger 1 trifft mit WK 0,85, Jäger 2 mit 0,75 und Jäger 3 mit 0,2.

Mit welcher WK wird Bambi getroffen?

Lösung: $A = \{\text{Jäger } i \text{ trifft} \mid i = 1, 2, 3\}$

Gesucht: WK von $A_1 \cup A_2 \cup A_3$

$$\begin{aligned} \mathbb{P}(A_1 \cup A_2 \cup A_3) &= 1 - \mathbb{P}(\bar{A}_1) \cdot \mathbb{P}(\bar{A}_2) \cdot \mathbb{P}(\bar{A}_3) \\ &= 1 - 0,15 \cdot 0,25 \cdot 0,8 \\ &= 0,97 \end{aligned}$$



1.1-42 Beispiel Parallel- und Reihenschaltung

Ein System besteht aus mehreren Elementen.

- Die Zuverlässigkeit (WK in einem bestimmten Zeitintervall nicht auszufallen) sei für jedes Element bekannt.
- Die Elemente heißen IN REIHE GESCHALTET, wenn das System genau dann funktioniert, wenn alle Elemente funktionieren.
- Die Elemente heißen PARALLEL GESCHALTET, wenn das System genau dann funktioniert, wenn wenigstens eins der Elemente funktioniert.

$F := \{\text{System funktioniert in betrachtetem Zeitintervall}\}$

$F_i := \{\text{Element } i \text{ funktioniert in betrachtetem Zeitintervall}\}$

Wir betrachten eine Sicherungsanlage mit 2 Teilsystemen, die beide funktionieren müssen. Zur Erhöhung der Zuverlässigkeit sind bei Teilsystem 1 ein Element und bei Teilsystem 2 zwei Elemente zu den Hauptelementen als Reserve parallel geschaltet. Die Zuverlässigkeit der Elemente in TS 1 beträgt je 0,9 und die der Elemente in TS 2 je 0,8. Die Elemente arbeiten unabhängig voneinander.

Teilsystem 1 (mit parallelen Elementen E_1 [Hauptelement] und E_3) liegt in Reihe zu Teilsystem 2 (mit parallelen Elementen E_2 [Hauptelement], E_4 und E_5).

$F_k := \{\text{Element } E_k \text{ funktioniert in betrachtetem Zeitintervall}\}$

Lösung:

$$F = \underbrace{(F_1 \cup F_3)}_{=:F_{13}} \cap \underbrace{(F_2 \cup F_4 \cup F_5)}_{=:F_{245}}$$

$$\mathbb{P}(F) = \mathbb{P}(F_{13} \cap F_{245}) = \mathbb{P}(F_{13} \cdot F_{245})$$

$$\mathbb{P}(F_{13}) = 1 - \mathbb{P}(\bar{F}_1) \cdot \mathbb{P}(\bar{F}_3) = 1 - 0,1^2 = 0,99$$

$$\mathbb{P}(F_{245}) = 1 - \mathbb{P}(\bar{F}_2) \cdot \mathbb{P}(\bar{F}_4) \cdot \mathbb{P}(\bar{F}_5) = 1 - 0,2^3 = 0,992$$

$$\Rightarrow \mathbb{P}(F) = 0,99 \cdot 0,992 = 0,98208$$

1.2 ZUFALLSVARIABLEN

1.2.1 GRUNDLAGEN

1.2-1 Definition Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein WK-Raum. Eine Zufallsvariable (ZV) oder Zufallsgröße ist eine Funktion X , die jedem Elementarereignis eine reelle Zahl zuordnet (d.h. $X : \Omega \rightarrow \mathbb{R}$) und die Zusatzeigenschaft

$$\text{für jedes Intervall gilt: } \{\omega \in \Omega \mid X(\omega) \in I\} \in \mathcal{A}$$

erfüllt.

1.2-2 Bemerkung Manchmal benötigt man noch allgemeinere Definitionen von ZVen. Auch wenn X nicht zwingend Werte aus $\mathbb{R}$ annimmt (und eine ähnliche Bedingung wie in Definition 1.2-1 erfüllt) spricht man von ZVen.

Bsp.: X nimmt komplexe Zahlen, Farben, Geschlechter usw. an.

1.2-3 Bemerkung

- Oft interessiert man sich für WKen:

$$\mathbb{P}(\omega \in \Omega \mid X(\omega) \in I)$$

bzw. in Kurzschreibweise:

$$\mathbb{P}(X \in I) \quad \text{für } I \subseteq \mathbb{R}$$



- Die Bedingung aus Definition 1.2-1 stellt sicher, dass solche WKen berechnet werden können.
- Zur effektiven Beschreibung solcher WKen dient die Verteilungsfunktion.

1.2-4 Beispiel Gegeben: (idealer) Würfel mit gefärbten Seiten (ohne Zahlen).

Ergebnisraum $\Omega = \{\text{blau, grün, gelb, rot, schwarz, magenta}\}$

Um mit den zufälligen Ergebnissen „rechnen“ zu können, führen wir eine ZV $X : \Omega \rightarrow \mathbb{R}$ ein. Wir setzen $X(\text{blau}) := 1$, $X(\text{grün}) := 2$, $X(\text{gelb}) := 3$, $X(\text{rot}) := 4$, $X(\text{schwarz}) := 5$, $X(\text{magenta}) := 6$

Es gilt dann z.B.:

$$\mathbb{P}(X = 3) = \mathbb{P}(\underbrace{\{\omega \in \Omega \mid X(\omega) = 3\}}_{\{\text{gelb}\}}) = \frac{1}{6}$$

$$\mathbb{P}(X \leq 2) = \mathbb{P}(\underbrace{\{\omega \in \Omega \mid X(\omega) = 1 \vee X(\omega) = 2\}}_{\{\text{blau, grün}\}}) = \mathbb{P}(\{\text{blau}\}) + \mathbb{P}(\{\text{grün}\}) = \frac{2}{6} = \frac{1}{3}$$

1.2-5 Beispiel Alfons (A) und Britta (B) spielen ein Würfelspiel. Gewürfelt wird gleichzeitig. Das Ergebnis von Alfons' Würfel sagt, wie viel Euro Alfons von Britta bekommt. Das Ergebnis von Britt's Würfel sagt, wie viel Euro Britta von Alfons bekommt.

Gesucht:

- Passendes WK-Modell um die ZV, die den Nettogewinn von Alfons beschreibt.
- WK, dass Alfons (netto) mehr als 3€ Verlust hat.

Lösung:

- Gesucht: $(\Omega, \mathcal{A}, \mathbb{P})$:

$\Omega = \{(i, j) \mid i, j \in \{1, \dots, 6\}\} = \{1, \dots, 6\}^2$ (mit i = Britt's Würfel und j = Alfons' Würfel)

$\mathcal{A} = \mathcal{P}(\Omega)$

$\mathbb{P} \dots$ Gleichverteilung auf Ω (da Laplace-Experiment)

Alfons Gewinn ist für $(1, 4)$ gerade $4 - 1€$. Also definieren wir $X : \Omega \rightarrow \mathbb{R}$ mittels $X((i, j)) := j - i$.

- Gesucht ist $\mathbb{P}(X < -3)$:

$$\begin{aligned} \mathbb{P}(X < -3) &= \mathbb{P}(\{(i, j) \mid X((i, j)) < -3\}) \\ &= \mathbb{P}(\underbrace{\{(i, j) \mid j - i < -3\}}_{(1,5), (1,6), (2,6)}) \\ &= \frac{3}{36} = \frac{1}{12} \end{aligned}$$

1.2-6 Beispiel In einer Firma werden auf 3 verschiedenen Anlagen Sandwiches produziert. Wir modellieren die Anzahl der Sandwiches pro Tag mit $\Omega = \underbrace{\{(x, y, z) \mid x, y, z \in \mathbb{N}_0\}}_{\omega}$ mit x, y, z

jeweils Produktionsmenge Anlage 1, 2 und 3 ((200, 45, 120) heißt also 300 Sandwiches wurden in Anlage 1 produziert, 200 in Anlage 2 und 120 in Anlage 3).

Wir interessieren uns für die Gesamtproduktion. Definiere daher: $X : \Omega \rightarrow \mathbb{R}$ mittels $X((x, y, z)) = x + y + z$.

Frage: Mit welcher WK überschreitet die Gesamtproduktion eine gewisse Mindestanzahl m nicht?

Gesucht ist also $\mathbb{P}(X \leq m) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \leq m\})$.



1.2-7 Definition Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein WK-Raum und X eine ZV. Die Funktion

$$F_X : \mathbb{R} \rightarrow [0, 1], \quad F_X(x) := \mathbb{P}(X \leq x)$$

heißt Verteilungsfunktion (VF) von X .

$F_X(x)$ ist also die WK, dass X einen Wert kleiner oder gleich der Zahl x annimmt.

1.2-8 Beispiel A und B spielen immer noch mit den gleichen Regeln wie in Bsp 1.2-6.

Wie sieht F_X aus? Dazu die Wertetabelle:

$x \in$	$(-\infty, -5)$	$[-5, 4)$	$[-4, 3)$	$[-3, -2)$	$[-2, -1)$	$[-1, 0)$
$F_X(x)$	0	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$
$x \in$	$[0, 1)$	$[1, 2)$	$[2, 3)$	$[3, 4)$	$[4, 5)$	$[5, \infty)$
$F_X(x)$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$

Denn z.B. gilt:

$$F_X(-4) = \mathbb{P}(X \leq -4) = \mathbb{P}(\{(5, 1), (6, 1), (6, 2)\}) = \frac{3}{36}$$

$$F_X(3) = \mathbb{P}(X \leq 3) = 1 - \mathbb{P}(X > 3) = 1 - \mathbb{P}(\{(1, 5), (1, 6), (2, 6)\}) = 1 - \frac{3}{36} = \frac{33}{36}$$

Beachte: In F_X können alle reellen Zahlen eingesetzt werden (Gezeichnete Funktion geht also von $-\infty$ mit 0 bis -5 und macht dann unstetige Sprünge bis 5, wo es bis ∞ mit 1 weiter geht).

1.2-9 Lemma Sei X eine ZV und F_X die zugehörige VF. Dann gilt:

- $0 \leq F_X(x) \leq 1$
- $x_1 \leq x_2 \Rightarrow F_X(x_1) \leq F_X(x_2)$ (Monotonie)
- $\lim_{x \rightarrow -\infty} F_X(x) = 0, \lim_{x \rightarrow \infty} F_X(x) = 1$
- $\lim_{x \searrow x_0} F_X(x) = F_X(x_0)$ (rechts-stetig)
- $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$
- $\mathbb{P}(X > a) = 1 - F_X(a)$
- $\mathbb{P}(X = a) = F_X(a) - \lim_{x \nearrow a} F_X(x)$ (Sprunghöhe bei $x = a$)

1.2.2 DISKRETE UND STETIGE ZUFALLSVARIABLE

Wir betrachten 2 Klassen von ZVen: diskrete und stetige.

1.2-10 Definition Eine ZV heißt DISKRETE ZV, falls X nur endlich viele oder abzählbar viele oder abzählbar unendlich viele Werte annehmen kann (M abzählbar unendlich $\Leftrightarrow \exists$ eine bijektive Abbildung $\varphi : M \rightarrow \mathbb{N}$. Sprich: „man kann alle Elemente durchnummerieren“, bspw. rationale Zahlen usw.).



1.2-11 Bemerkung Sei X eine diskrete ZV welche nur die Werte $x_1, x_2, \dots$ annehmen kann.

(a) Wir nenne die Funktion

$$f : x_i \mapsto f(x_i) := p_i := \mathbb{P}(X = x_i)$$

Wahrscheinlichkeitsfunktion.

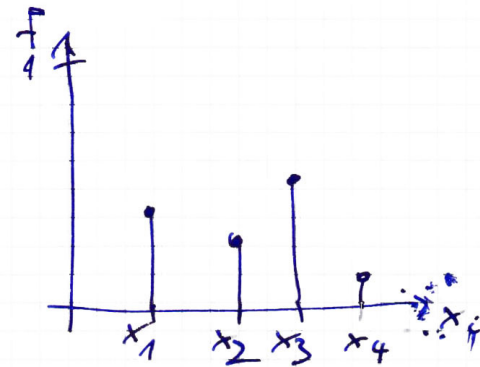
$$(b) \sum_i f(x_i) = \sum_i p_i = 1$$

$$(c) \mathbb{P}(a < X \leq b) = \sum_{i: a < x_i \leq b} f(x_i)$$

(d) Darstellung mit Verteilungstabelle:

Werte	x_1	x_2	x_3	$\dots$
WK $f(x_i)$	p_1	p_2	p_3	$\dots$

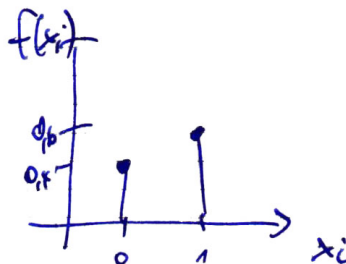
(e) Darstellung als Stabdiagramm:



1.2-12 Beispiel Betrachte unfaire Münze, die mit WK 0,6 auf Zahl fällt. Dann

- $\Omega = \{K, Z\}$
- $X(K) := 0, X(Z) := 1$
- $\mathbb{P}(X = 0) = 0,4 = 1 - \mathbb{P}(X = 1)$

Wert x_i	0	1
WK $f(x_i)$	0,4	0,6



1.2-13 Beispiel Betrachten Wurf mit 2 Würfeln: $\Omega = \{(i, j) \mid i, j = 1, \dots, 6\}$. Augensumme soll als ZV dargestellt werden: $X((i, j)) := i + j$ für $i, j = 1, \dots, 6$

Verteilungstabelle:

x_i	2	3	4	5	6	7	8	9	10	11	12
$f(x_i)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

$$(a) \mathbb{P}(X > 9) = \mathbb{P}(X = 10) + \mathbb{P}(X = 11) + \mathbb{P}(X = 12) = \frac{6}{36}$$

$$(b) \mathbb{P}(6 \leq X \leq 8) = \frac{16}{36}$$

$$(c) F_X(3) = \mathbb{P}(X \leq 3) = \frac{3}{36}$$

1.2-14 Definition Eine ZV heißt STETIGE ZV, falls ihre Verteilungsfunktion F_X stetig ist. In dieser Vorlesung betrachten wir nur stetige ZV, bei denen F_X differenzierbar ist.

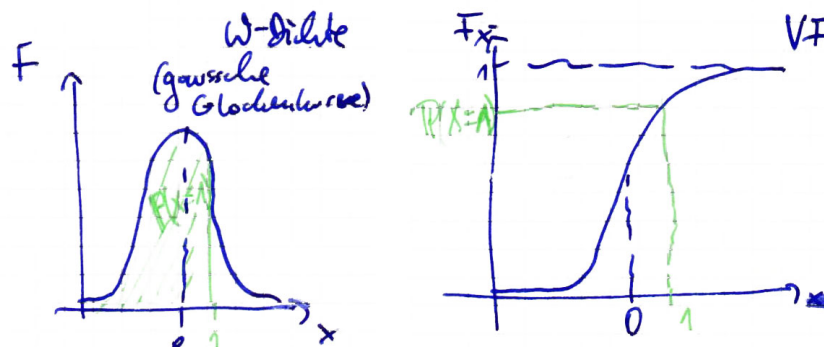
1.2-15 Definition Sei X eine ZV mit differenzierbarer Verteilungsfunktion F_X . Dann wird die Ableitung $f := F'_X$ (WAHRSCHEINLICHKEITS-)DICHTE(FUNKTION) genannt.

1.2-16 Bemerkung

- Ist f die Dichtefunktion zur VF F_X , so gilt

$$F_X(x) = \int_{-\infty}^x f(t) dt \quad (\text{Hauptsatz der Differential und Integralrechnung})$$

- Nicht jede stetige Funktion F kann als Integral einer Dichtefunktion geschrieben werden.
- Bilder:



$$\begin{aligned}
 \mathbb{P}(X \leq 1) &= F_X(1) \\
 &= \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \leq 1\}) \\
 &= \{\text{WK, dass } X \text{ einen Wert } \leq 1 \text{ annimmt}\} \\
 &= \{\text{Funktionswert von } F_X \text{ bei } X = 1\} \\
 &= \int_{-\infty}^1 f(t) dt \\
 &= \{\text{Flächeninhalt des markierten Bereichs der Dichtefunktion}\}
 \end{aligned}$$



1.2-17 Satz Sei X eine stetige ZV mit Dichtefunktion f . Dann gilt:

- f ist auf ganz $\mathbb{R}$ definiert.
- $f(x) \geq 0 \quad \forall x \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f(x) dx = 1$ (Dichtefunktion ist normiert)

Umkehrung des Satzes gilt auch:

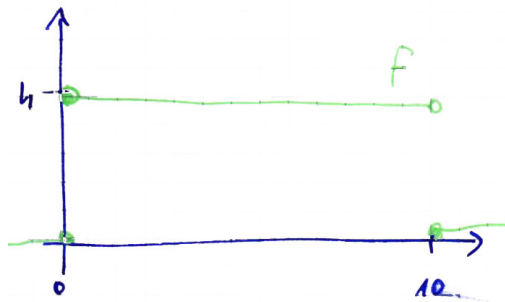
1.2-18 Satz Sei f eine integrierbare reelwertige Funktion. Dann gilt:

- f ist auf ganz $\mathbb{R}$ definiert.
- $f(x) \geq 0 \quad \forall x \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f(x) dx = 1$

Dann ist f die Dichte einer Zufallsvariablen X . Die zugehörige VF $F_X : \mathbb{R} \rightarrow [0, 1]$ ist dann

$$F_X(x) := \int_{-\infty}^{\infty} f(t) dt$$

1.2-19 Beispiel Wir kommen zu einer zufälligen Zeit an der Bushaltestelle an. Der Bus fährt alle 10 Minuten. Die ZV X beschreibe die Wartezeit auf den nächsten Bus, d.h. X kann alle Wert zwischen 0 und 10 annehmen wobei jede Wartezeit gleich-wahrscheinlich sein soll. Die zugehörige Dichtefunktion ist daher:



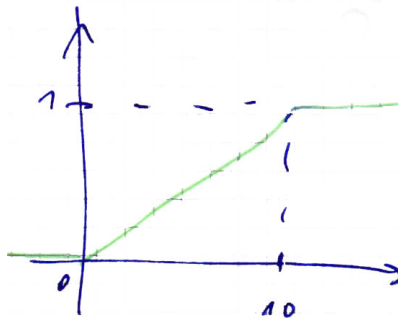
$$f(x) := \begin{cases} h & 0 < x < 10 \\ 0 & \text{sonst} \end{cases}$$

(a) Wie ist h zu wählen?

$$\text{Es muss gelten } 1 \stackrel{!}{=} \int_{-\infty}^{\infty} f(x) dx = \int_0^{10} h dx = [h \cdot x]_0^{10} = 10 \cdot h \Rightarrow h = \frac{1}{10}$$

(b) Wie sieht F_X aus?





$$F_X(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0 & x \leq 0 \\ \frac{1}{10}x & 0 < x < 10 \\ 1 & x \geq 10 \end{cases}$$

$$\text{Fall } x \leq 0: F_X(x) = \int_{-\infty}^x 0 dt = 0$$

$$\text{Fall } 0 < x < 10: F_X(x) = \int_{-\infty}^0 0 dt + \int_0^x \frac{1}{10} dt = \frac{1}{10}x$$

$$\text{Fall } x \geq 10: F_X(x) = \underbrace{\int_{-\infty}^0 0 dt}_0 + \underbrace{\int_0^{10} \frac{1}{10} dt}_1 + \underbrace{\int_{10}^x 0 dt}_0 = 1$$

1.2-20 Satz Sei X eine stetige ZV mit Dichte f und VF F . Die WK, dass X einen Wert im Intervall (a, b) (oder $[a, b]$, $[a, b)$, $(a, b]$) annimmt ist:

$$\begin{aligned} \mathbb{P}(a < X < b) &= \mathbb{P}(a \leq X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X \leq b) \\ &= F(b) - F(a) \\ &= \int_a^b f(x) dx \end{aligned}$$

Das entspricht den Flächeninhalt unter f zwischen a und b .

1.2-21 Bemerkung

- Für $a = b$ zeigt der Satz: Ist X stetig, so gilt

$$\mathbb{P}(X = a) = 0$$

für alle $a \in \mathbb{R}$

- Im Satz ist auch $a = -\infty$ und/oder $b = +\infty$ erlaubt, wobei

$$F(-\infty) := \lim_{x \rightarrow -\infty} F(x) = 0$$

$$F(\infty) := \lim_{x \rightarrow \infty} F(x) = 1$$



1.2-22 Beispiel (Fortsetzung von Beispiel 1.2.19)

- Wie groß ist die WK maximal 3 Minuten zu warten? Gesucht: $\mathbb{P}(X \leq 3)$
 $\mathbb{P}(X \leq 3) = F(3) = 0,1 \cdot 3 = 0,3$
- Wie groß ist die WK mindestens 2 Minuten zu warten? Gesucht: $\mathbb{P}(X \geq 2)$
 $\mathbb{P}(X \geq 2) = \mathbb{P}(2 \leq X < \infty) = F(\infty) - F(2) = 1 - 2 \cdot 0,1 = 0,8$
- Wie groß ist die WK zwischen 5 und 9 Minuten zu warten?
 $\mathbb{P}(5 \leq X \leq 9) = F(9) - F(5) = 9 \cdot 0,1 - 5 \cdot 0,1 = 0,4$

1.2.3 ERWARTUNGSWERT UND VARIANZ

1.2-23 Beispiel (Würfelspiel) A und B würfeln wieder: 2 Würfel. ZV X beschreibt die Augensumme. Spielregeln:

Ergebnis	WK	Bewertung
$\{X \leq 3\}$	1/12	A zahlt B 20,10€
$\{4 \leq X \leq 6\}$	1/3	A zahlt B 1,20€
$\{X = 7\}$	1/6	niemand zahlt etwas
$\{8 \leq X \leq 9\}$	1/4	B zahlt A 3,10€
$\{10 \leq X \leq 12\}$	1/6	B zahlt A 7,80€

$Y \dots$ Gewinn von A

Verteilungstabelle von Y :

Werte y_i	-20,1	-1,2	3,1	7,8	0
WK p_i	1/12	1/3	1/4	1/6	1/6

Frage: Ist das Spiel gerecht?

Mittlerer Gewinn: $-20,1 \cdot \frac{1}{12} + (-1,2) \cdot \frac{1}{3} + 0 \cdot \frac{1}{6} + 3,1 \cdot \frac{1}{4} + 7,8 \cdot \frac{1}{6} = 0$

Dies motiviert die folgende Definition:

1.2-24 Definition Sie X eine ZV. Der Erwartungswert $\mathbb{E}(X)$ und die Varianz $\text{Var}(X)$ sind wie folgt definiert:

(a) Falls X diskret ist, mit Verteilungstabelle:

Werte	x_1	x_2	x_3	$\dots$
WK	p_1	p_2	p_3	$\dots$

(bzw. mit WK-Funktion f), dann

$$\mathbb{E}(X) = \sum_i x_i \cdot p_i = \sum_i x_i \cdot f(x_i)$$

und

$$\text{Var}(X) = \sum_i (x_i - \mathbb{E}(X))^2 \cdot p_i = \sum_i (x_i - \mathbb{E}(X))^2 f(x_i)$$

(b) Falls X stetig ist mit Dicht f , dann

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

und

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 f(x) dx$$



1.2-25 Satz Sei X eine ZV und $a, b \in \mathbb{R}$. Dann:

- $\mathbb{E}(a + bX) = a + b\mathbb{E}(X)$ (Linearität des Erwartungswertes)
- $\text{Var}(a + bX) = b^2 \text{Var } X$
- $\text{Var}(X) = 0 \Leftrightarrow \exists a \in \mathbb{R} : \mathbb{P}(X = a) = 1$

1.2-26 Bemerkung

- $\sigma_X = \sqrt{\text{Var } X}$ wird STANDARDABWEICHUNG genannt.
- $\text{Var}(X)$ ist die mittlere quadratische Abweichung vom Erwartungswert und es gilt:

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

- Sei $g : \mathbb{R} \rightarrow \mathbb{R}$ eine beliebige Funktion, dann gilt:

(im diskreten) $\mathbb{E}(g(X)) = \sum_i g(x_i) \cdot f(x_i)$

(im stetigen) $\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f(x) dx$

z.B. für $g(x) := x^2$:

$$\mathbb{E}(X^2) = \sum_i x_i^2 f(x_i) \text{ bzw. } \mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$$

1.2-27 Beispiel $X \dots$ Zahl der Einsätze eines Havariedienstes an einem Tag. Erfahrung liefert:

x_i	0	1	2
$p_i = f(x_i)$	0,6	0,3	0,1

$$\mathbb{E}(X) = 0 \cdot 0,6 + 1 \cdot 0,3 + 2 \cdot 0,1 = 0,5$$

$$\mathbb{E}(X^2) = 0^2 \cdot 0,6 + 1^2 \cdot 0,3 + 2^2 \cdot 0,1 = 0,7$$

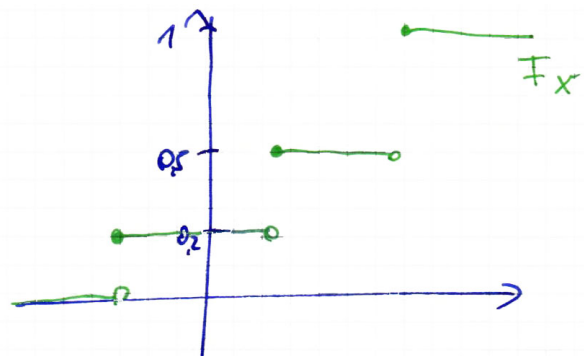
$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = 0,7 - 0,5^2 = 0,45$$

$$\sigma_X = \sqrt{0,45} = 0,671$$

Wir betrachten nun eine ZV X und fragen uns: Welchen Wert $m \in \mathbb{R}$ muss man wählen, damit

$\mathbb{P}(X \leq m) \geq \frac{1}{2}$ und $\mathbb{P}(X \geq m) \geq \frac{1}{2}$ gilt?

Problem: Antwort nicht eindeutig!



Hier erfüllt jedes $m \in [1, 3)$ diese Bedingung! Welchen dieser Werte wählen wir? Den kleinsten (und nennen ihn MEDIAN)!



1.2-28 Definition Ist F_X die Verteilungsfunktion einer ZV X , so heißt

$$F_X^{-1} : [0, 1] \rightarrow \mathbb{R}, \quad F_X^{-1}(\alpha) := \min\{x \in \mathbb{R} \mid F_X(x) \geq \alpha\}$$

die VERALLGEMEINERTE INVERSE VERTEILUNGSFUNKTION. Der Median m_X der Verteilungsfunktion F_X ist definiert als

$$m_X = F_X^{-1}(0,5).$$

Für gegebenes $\alpha \in (0, 1)$ heißt

$$q_\alpha = F_X^{-1}(\alpha)$$

das α -Quantil zur Verteilung F_X .

1.2-29 Bemerkung

- Daher ist der Median das 0,5-Quantil der Verteilung: $m_X = q_{0,5}$
- Im Allgemeinen gilt: Erwartungswert $\neq$ Median.
Beispiel:

x_i	1	3	6	7
$\mathbb{P}(X = x_i)$	0,2	0,3	0,1	0,4
$\mathbb{E}(X) = 4,5$				
$m_X = \min\{x \in \mathbb{R} \mid F_X(x) \geq 0,5\} = \min[3, \infty) = 3$				

1.2.4 KOVARIANZ UND UNABHÄNGIGKEIT

Betrachten nun mehrere ZVen gleichzeitig.
Fragen:

- Haben ZVen „Einfluss aufeinander“?
~~ Unabhängigkeit, Unkorreliertheit
- Kann man das Verhalten mehrerer ZVn gleichzeitig beschreiben?
~~ gemeinsame Verteilung

1.2-30 Definition Sind X und Y ZVen, so heißt

$$F : \mathbb{R}^2 \rightarrow [0, 1], \quad F(a, b) = \mathbb{P}(X \leq a, Y \leq b)$$

gemeinsame Verteilungsfunktion von X und Y .

- Sind beide ZVen diskret, wobei X die Werte $x_1, x_2, \dots$ und Y die Werte $y_1, y_2, \dots$ annehmen kann, dann heißt die Funktion f gegeben durch

$$f(x_i, y_j) := \mathbb{P}(X = x_i, Y = y_j)$$

GEMEINSAME VERTEILUNGSFUNKTION.

- Sind beide ZVen stetig und existiert eine Funktion $f : \mathbb{R}^2 \rightarrow [0, \infty)$ mit

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(s, t) dt ds$$

so heißt f GEMEINSAME (WAHRSCHEINLICHKEITS-)DICHTE(-FUNKTION) von X und Y .



1.2-31 Bemerkung Im diskreten Fall gilt:

$$F(x, y) = \sum_{i: x_i \leq x} \sum_{j: y_j \leq y} f(x_i, y_j)$$

- Die Definition 1.2.30 lässt sich auf beliebig viele ZVen erweitern.
- ZVen lassen sich zu einem (zufälligen) Vektor zusammenfassen:

1.2-32 Definition Sind $X_1, \dots, X_n$ ZVen so heißt

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

N-DIMENSIONALER ZUFALLSVEKTOR.

1.2-33 Beispiel (Zufallsvektor mit $n = 2$)

Seien X, Y diskrete ZVen gegeben durch

$X \dots$ Anzahl der technischen Durchsichten eines PKW eines bestimmten Typs zwischen 0 und 15.000 km.

$Y \dots$ Anzahl der Motorpannen dieses PKW zwischen 0 und 15.000 km.

Setzen $Z = \begin{pmatrix} X \\ Y \end{pmatrix}$

- Verteilungstabelle:

$\begin{pmatrix} X \\ Y \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 2 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 3 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 2 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 1 \end{pmatrix}$
$\mathbb{P}(X = x, Y = y)$	0,02	0,04	0,03	0,01	0,05	0,01	0,05	0,53	0,17

- gemeinsame Wahrscheinlichkeitsfunktion (in Matrixschreibweise):

$$P = (p_{ij}) \quad p_{ij} := f(x_i, y_i) = \mathbb{P}(X = x_i, Y = y_i)$$

$x \backslash y$	0	1	2	3	
0	0,02	0,04	0,04	0,01	0,1
1	0,05	0,1	0,05	0	0,2
2	0,53	0,17	0	0	0,7
	0,6	0,31	0,08	0,01	1

z.B. $f(0, 2) = 0,03$

- gemeinsame Verteilungsfunktion: Es gilt z.B. $F(2, 1) = \mathbb{P}(X \leq 2, Y \leq 1) = 0,02 + 0,04 + 0,05 + 0,1 + 0,53 + 0,17 = 0,91$ (entspricht dem „Rechteck“ der WK-Funktion in Matrixschreibweise, wo $x \leq 2$ und $y \leq 1$)
- Randverteilungen:

Verteilung von X

$$\mathbb{P}(X = x_i) = \sum_j p_{i,j} =: p_{i,\cdot}$$

x_i	0	1	2
$p_{i,\cdot}$	0,1	0,2	0,7

Verteilung von Y

$$\mathbb{P}(Y = y_i) = \sum_j p_{j,i} =: p_{\cdot,i}$$

y_i	0	1	2	3
$p_{\cdot,i}$	0,6	0,31	0,08	0,01



1.2-34 Bemerkung Mit der gemeinsamen Verteilung (Dicht, WK-Funktion) lassen sich z.B.

- (1) WKen berechnen und
- (2) Funktionen von ZVen untersuchen.

Seien X und Y ZVen mit gemeinsamer Dichte f , dann gilt z.B.:

$$(1) \mathbb{P}(X \in [x_1, x_2], Y \in [y_1, y_2]) = \mathbb{P}(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2)$$

$$\int_{x_1}^{x_2} \int_{y_1}^{y_2} f(s, t) dt ds$$

$$(2) \text{ und für beliebige } g: \mathbb{R}^2 \rightarrow \mathbb{R}:$$

$$E(g(X, Y)) = \int_{\mathbb{R}} \int_{\mathbb{R}} g(x, y) f(x, y) dy dx$$

(sofern die Integrale existieren)

Insbesondere:

$$E(X \cdot Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot y f(x, y) dy dx$$

Analoge Formeln gelten für diskrete ZVen mit der WK-Funktion f , z.B.

$$E(X \cdot Y) = \sum_i \sum_j x_i y_j f(x_i, y_j)$$

falls X die Werte $x_1, x_2, \dots$ annimmt und Y die Werte $y_1, y_2, \dots$.

1.2-35 Definition Seien X und Y zwei ZVen. Dann heißen

- (a) $\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y)) = \mathbb{E}(XY) - \mathbb{E}X \cdot \mathbb{E}Y$ die KOVARIANZ von X und Y .
- (b) $\varrho_{X,Y} := \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$ der KORRELATIONSKOEFFIZIENT.
- (c) X und Y UNKORRELIERT, wenn $\varrho_{X,Y} = 0$ (also wenn $\text{Cov}(X, Y) = 0$)

1.2-36 Definition Zwei ZVen X und Y heißen (stochastisch) unabhängig, falls für beliebige Intervalle $I_1, I_2 \subseteq \mathbb{R}$ gilt:

$$\mathbb{P}(X \in I_1, Y \in I_2) = \mathbb{P}(X \in I_1) \cdot \mathbb{P}(Y \in I_2)$$

Mehrere ZVen $X_1, X_2, \dots$ heißen (stochastisch) unabhängig, falls für jede Auswahl $X_{k_1}, \dots, X_{k_m}$ und beliebige Intervalle $I_1, \dots, I_m$ gilt:

$$\mathbb{P}(X_{k_1} \in I_1, \dots, X_{k_m} \in I_m) = \prod_{i=1}^m \mathbb{P}(X_{k_i} \in I_i)$$

1.2-37 Bemerkung

- X und Y stochastisch unabhängig $\Leftrightarrow$ für beliebige Intervalle I_1 und I_2 sind $\{x \in I_1\}$ und $\{y \in I_2\}$ stochastisch unabhängig.
- X und Y unabhängig $\Rightarrow \mathbb{E}(XY) = \mathbb{E}X \cdot \mathbb{E}Y$
- X und Y unabhängig $\Rightarrow X$ und Y unkorreliert
- ABER: X und Y unkorreliert $\nRightarrow X$ und Y unabhängig
- $\text{Cov}(X, X) = \text{Var}(X)$



1.2-38 Beispiel (Tetraeder-Würfel)

Ein Tetraeder ist mit den Zahlenpaaren $\Omega = \{(1, 2), (0, 2), (2, 1), (0, 0)\}$ beschriftet. Jede Seite ist gleich-wahrscheinlich. Die ZV X beschreibt die erste Zahl, Y die Zweite (im Zahlenpaar).

Dann:

$$\mathbb{P}(X = 0) = \frac{1}{2}, \mathbb{P}(X = 1) = \frac{1}{4}, \mathbb{P}(X = 2) = \frac{1}{4},$$

$$\mathbb{P}(Y = 0) = \frac{1}{4}, \mathbb{P}(Y = 1) = \frac{1}{4}, \mathbb{P}(Y = 2) = \frac{1}{2},$$

$$\mathbb{P}(X \cdot Y = 0) = \frac{1}{2}, \mathbb{P}(X \cdot Y = 2) = \frac{1}{2}$$

Und damit:

$$\mathbb{E}(X) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} = \frac{3}{4}$$

$$\mathbb{E}(Y) = \frac{5}{4}$$

$$\mathbb{E}(XY) = 1$$

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y) = 1 - \frac{3}{4} \cdot \frac{5}{4} = \frac{1}{16}$$

$\Rightarrow X$ und Y nicht unkorreliert

$\Rightarrow$ nicht unabhängig

$$\text{z.B. } \underbrace{\mathbb{P}(X = 1)}_{\frac{1}{4}} \cdot \underbrace{\mathbb{P}(Y = 0)}_{\frac{1}{4}} \neq \underbrace{\mathbb{P}(X = 1, Y = 0)}_0$$

1.2-39 Satz Seien X und Y ZVen und $a, b \in \mathbb{R}$. Dann gilt:

- $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$
- $\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2\text{Cov}(X, Y)$
- $\text{Cov}(aX + bY, Z) = a\text{Cov}(X, Z) + b\text{Cov}(Y, Z)$
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$

1.2-40 Bemerkung (zu $\rho_{X,Y}$)

- $\rho_{X,Y} \in [-1, 1]$
- $\rho_{X,Y}$ ist ein Maß für den linearen Zusammenhang zwischen X und Y
- Extremfälle:
$$\rho_{X,Y} = \begin{cases} +1 \\ -1 \end{cases} \Leftrightarrow Y = aX + b \text{ mit } \begin{cases} a > 0 \\ a < 0 \end{cases}$$

Hier gilt also jeweils ein streng linearer Zusammenhang.
- Die Gerade $y = a_0 + a_1x$ mit $a_1 = \frac{\sigma_X}{\sigma_Y} \rho_{X,Y}$, $a_0 = \mathbb{E}Y - a_1\mathbb{E}X$ heißt Regressionsgerade Y bezüglich X (beste lineare Näherung).

1.3 SPEZIELLE VERTEILUNGEN

1.3.1 SPEZIELLE DISKRETE VERTEILUNGEN

1.3.1.1 BERNOULLI VERTEILUNG

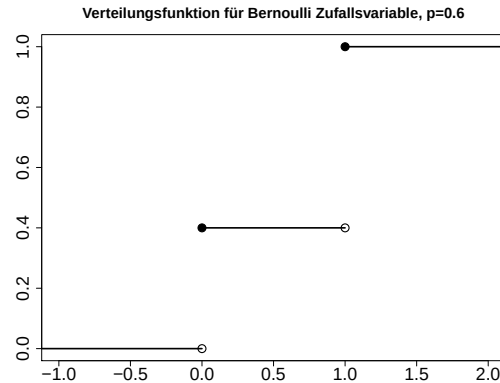
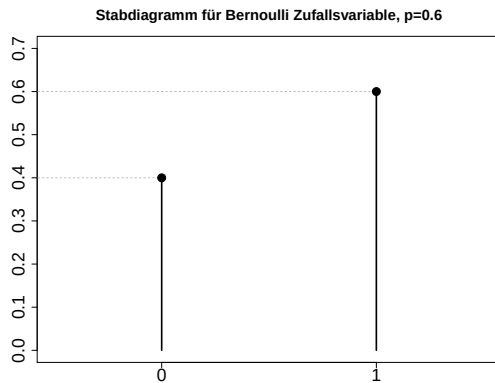
Die Vorstellung einer (un-)fairen Münze liefert eine sehr einfache Zufallsvariable:

1.3-1 Definition Eine ZV X welche genau 2 Werte annehmen kann heißt Bernoulli-verteilt.



1.3-2 Bemerkung

- Die möglichen Werte von X werden typischerweise $\{0, 1\}$ genannt.
- Schreibweise: $\mathbb{P}(X = 1) = p$, $\mathbb{P}(X = 0) = 1 - p = q$ und $X \sim \text{Ber}(p)$
- Erwartungswert: $\mathbb{E}(X) = 0 \cdot q + 1 \cdot p = p$
- Varianz: $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = 0^2 \cdot q + 1^2 \cdot p - p^2 = pq$
- Standardabweichung: $\sigma_X = \sqrt{pq}$



1.3-3 Beispiel Seien $X, Y \sim \text{Ber}(p)$. Setze $Z_1 := X + Y$, $Z_2 = X - Y$ unabhängig.

- Wie sind Z_1 und Z_2 verteilt?
- Sind Z_1 und Z_2 unkorreliert?
- Sind Z_1 und Z_2 unabhängig?

Lösung:

(a) Mögliche Werte:

X	Y	Z_1	Z_2	mit WK
0	0	0	0	$\mathbb{P}(X = 0, Y = 0) = \mathbb{P}(X = 0) \cdot \mathbb{P}(Y = 0) = q^2$
0	1	1	-1	qp
1	0	1	1	pq
1	1	2	0	p^2

$\Rightarrow$	z	0	1	2	z	-1	0	1
	$\mathbb{P}(Z_1 = z)$	q^2	$2pq$	p^2	$\mathbb{P}(Z_2 = z)$	pq	$p^2 + q^2$	pq
	z	-1	0	1				
	$\mathbb{P}(Z_1 Z_2 = z)$	pq	$p^2 + q^2$	pq				

(b) $\text{Cov}(Z_1, Z_2) = \mathbb{E}(Z_1 Z_2) - \mathbb{E}(Z_1)\mathbb{E}(Z_2) = 0$, denn:

$$\mathbb{E}(Z_1) = 0 \cdot q^2 + 1 \cdot 2pq + 2 \cdot p^2 = 2p(q + p) = 2p$$

$$\mathbb{E}(Z_2) = -pq + pq = 0 = \mathbb{E}(Z_1 Z_2)$$

$\Rightarrow Z_1$ und Z_2 sind unkorreliert

(c) Es müsste bspw. gelten $\underbrace{\mathbb{P}(Z_1 = 0, Z_2 = 1)}_0 = \underbrace{\mathbb{P}(Z_1 = 0)}_{q^2} \underbrace{\mathbb{P}(Z_2 = 1)}_{pq}$, ist aber falsch.

$\Rightarrow Z_1$ und Z_2 nicht unabhängig.



1.3.1.2 BINOMIALVERTEILUNG

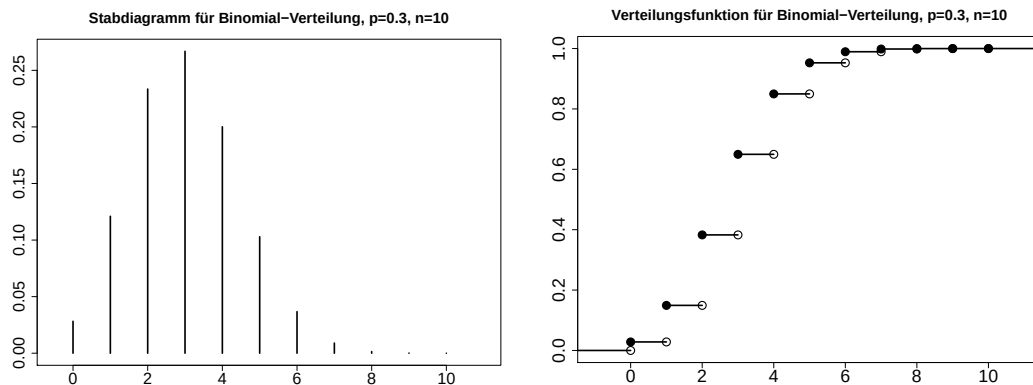
1.3-4 Definition Die ZV X heißt binomialverteilt mit den Parametern n und p (wobei $n \in \mathbb{N}, p \in [0, 1]$), wenn sie die Werte $0, \dots, n$ mit den WKen

$$p_i = \mathbb{P}(X = i) = \binom{n}{i} p^i (1-p)^{n-i}, \quad i = 0, \dots, n$$

annimmt.

1.3-5 Bemerkung

- Kurschreibweise: $X \sim \text{Bin}(n, p)$
- Erwartungswert: $\mathbb{E}(X) = np$
- Varianz: $\text{Var}(X) = np(1-p)$



1.3-6 Satz Sind $X_1, \dots, X_n$ unabhängige Bernoulliverteilte ZVen (alle mit Parameter p), dann

$$X_1 + X_2 + \dots + X_n \sim \text{Bin}(n, p)$$

1.3-7 Beispiel (Massenproduktion mit Ausschuss)

Ein Massenprodukt (Schokoriegel) mit einem Ausschussanteil von 3% wird in 20er Packungen verkauft. Wie groß ist die WK, dass eine Packung maximal 2 Ausschussstücke enthält?

Lösung:

$X_i \dots$ ZV mit:

$X_i = 1 \dots$ Schokoriegel i in der Packung ist Ausschuss,

$X_i = 0 \dots$ Schokoriegel i ist keine Ausschuss.

$$Y = \sum_{i=1}^{20} X_i$$

Wir wissen $X_i \sim \text{Ber}(0, 03)$

Annahme X_i sind unabhängig.

$\xrightarrow{\text{Satz 1.3.6}} Y \sim \text{Bin}(20, 0, 03)$

Gesucht: $\mathbb{P}(Y \leq 2)$

$$\begin{aligned} \mathbb{P}(Y \leq 2) &= \mathbb{P}(Y = 0) + \mathbb{P}(Y = 1) + \mathbb{P}(Y = 2) \\ &= \binom{20}{0} \cdot 0,03^0 \cdot 0,97^{20} + \binom{20}{1} 0,03^1 \cdot 0,97^{19} + \binom{20}{2} 0,03^2 \cdot 0,97^{18} \\ &= 0,979 \end{aligned}$$



1.3.1.3 DISKRETE GLEICHVERTEILUNG

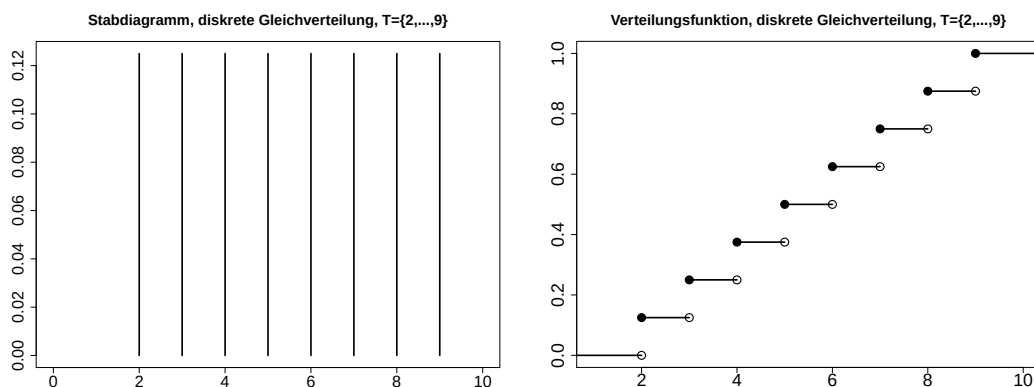
1.3-8 Definition Eine ZV X genügt der diskreten Gleichverteilung auf der Menge $T = \{x_1, \dots, x_n\}$, falls sie nur Werte aus T annehmen kann und

$$\mathbb{P}(X = x_1) = \dots = \mathbb{P}(X = x_n) = \frac{1}{n}$$

gilt.

1.3-9 Bemerkung

- Kurzschreibweise: $X \sim U(T)$
- Erwartungswert: $\mathbb{E}(X) = \frac{1}{n} \sum_{i=1}^n x_i$
- Varianz: $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2$
- Beispiel: faire Münze, fairer Würfel, ...



1.3-10 Bemerkung (Beziehung zum Laplace-Experiment)

- Wir betrachten ein Laplace-Experiment mit

$$\Omega = \{\omega_1, \dots, \omega_n\}$$

und dem WK-Maß $\mathbb{P}$, d.h.

$$\mathbb{P}(\{\omega_1\}) = \dots = \mathbb{P}(\{\omega_n\}) = \frac{1}{n}$$

- Die ZV $X : \Omega \rightarrow \mathbb{R}$, $X(\omega) = \omega$ ist damit gleichverteilt. Denn:

$$\mathbb{P}(X = \omega_i) = \mathbb{P}(\{\omega \in \Omega | X(\omega) = \omega_i\}) = \mathbb{P}(\omega_i) = \frac{1}{n}$$

1.3.1.4 HYPERGEOMETRISCHE VERTEILUNG

1.3-11 Definition Eine ZV X heißt Hypergeometrisch verteilt, mit ganzzahligen Parametern N , M und n ($0 < M \leq N, 0 < n \leq N$), wenn sie nur die Werte $T = \{\max\{0, n + M - N\}, \dots, \min\{n, M\}\}$ annehmen kann und für jedes $m \in T$ gilt:

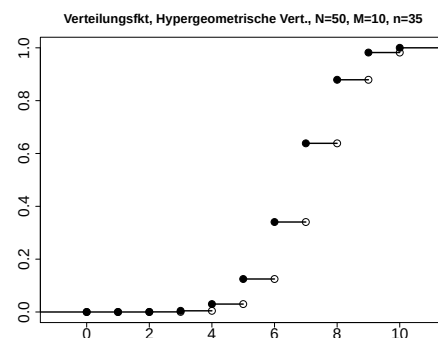
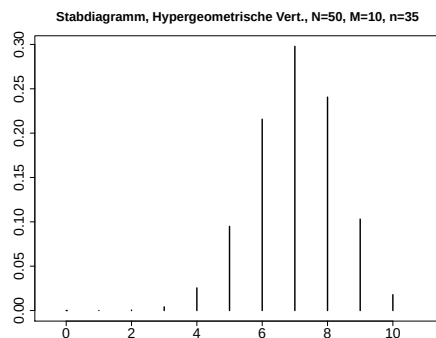
$$p_m := \mathbb{P}(X = m) = \frac{\binom{M}{n} \binom{N-M}{n-m}}{\binom{N}{n}}$$



1.3-12 Bemerkung

- Kurzschreibweise: $X \sim \text{Hyp}(N, M, n)$
- Erwartungswert und Varianz: $\mathbb{E}(X) = n \frac{M}{N}$ und $\text{Var}(X) = n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}$
- Anwendung: Stichprobe ohne Zurücklegen (bspw. Qualitätskontrolle, Lotto)
 - N Objekte, davon M mit bestimmten Merkmal (bspw. Ausschuss, Gewinnzahl)
 - n Objekte werden entnommen
 - $X \dots$ Anzahl der Objekte unter den n entnommenen, die das Merkmal besitzen

$$\Rightarrow X \sim \text{Hyp}(N, M, n)$$



1.3-13 Beispiel In einer Lostrommel befinden sich 20 Lose, davon 5 Gewinnlose. Jemand zieht 3 Lose (ohne Zurücklegen). Wie groß ist die WK, dass sich darunter genau 2 Gewinnlose befinden?

Lösung:

$X \dots$ Anzahl der Gewinnlose unter den 3 gezogenen.

Es gilt $X \sim \text{Hy}(20, 5, 3)$. Daher $\mathbb{P}(X = 2) = \frac{\binom{5}{2} \binom{15}{1}}{\binom{20}{3}} = \frac{10 \cdot 15}{1140} = 0,1316$

1.3-14 Bemerkung Wie kommt man darauf gerade diese Formel zu verwenden?

Idee: Modellierung als Laplace-Experiment.

$$\Omega = \{\{a_1, a_2, a_3\} \mid a_i \neq a_j \text{ falls } i \neq j \text{ und } a_1, a_2, a_3 \in \underbrace{\{g_1, \dots, g_5\}}_{\text{Gewinnlose}}, \underbrace{\{n_1, \dots, n_{15}\}}_{\text{Nieten}}\} \rightsquigarrow \text{Elementarer-}$$

eignisse gleich-wahrscheinlich. Es gilt: $|\Omega| = \binom{20}{3}$.

Das Ereignis, das uns interessiert ist:

$$A = \{\{a_1, a_2, a_3\} \in \Omega \mid \{a_1, a_2, a_3\} \text{ enthält genau 2 der } \{g_1, \dots, g_5\} \text{ und genau 1 der } \{n_1, \dots, n_{15}\}\}$$

Es gilt nun $A = \{X = 2\}$ und $|A| = \binom{5}{2} \cdot \binom{15}{1} \Rightarrow \mathbb{P}(X = 2) = \frac{\binom{5}{2} \binom{15}{1}}{\binom{20}{3}}$

1.3.1.5 GEOMETRISCHE VERTEILUNG

1.3-15 Definition Eine ZV X heißt geometrisch verteilt mit dem Parameter $p \in (0, 1)$, falls sie nur die Werte $1, 2, \dots$ annehmen kann und

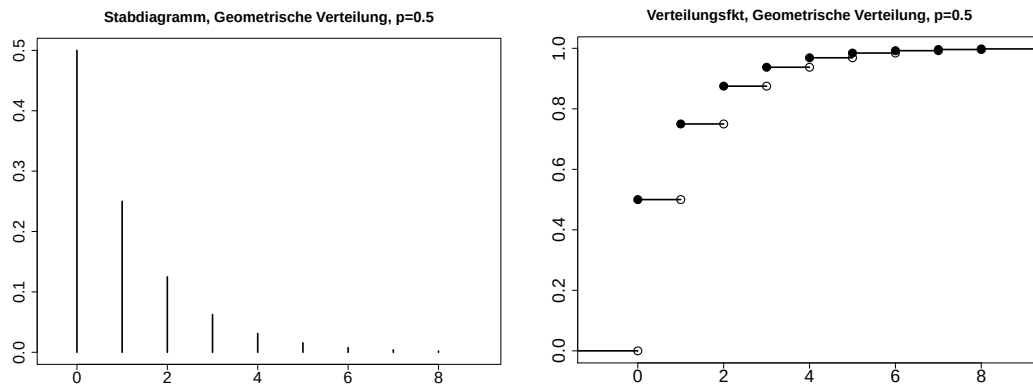
$$p_m := \mathbb{P}(X = m) = p(1 - p)^{m-1} \quad m = 1, 2, \dots$$

gilt.



1.3-16 Bemerkung

- Kurzschreibweise: $X \sim \text{Geo}(p)$
- Varianz: $\text{Var}(X) = \frac{1-p}{p^2}$
- Anwendung: Anzahl der Versuche bis der erste Erfolg eintritt, bei hintereinander ausführen von unabhängigen identischen Bernoulli Zufallsexperimenten.



1.3-17 Beispiel Der Postbote hat ein Paket für Petra bei ihrem Nachbarn abgegeben. Petra klopft jeden Nachmittag an seine Tür. Leider ist er nur mit WK 0,3 zu Hause. Annahme: Anwesenheiten des Nachbarn an verschiedenen Tagen sind unabhängig voneinander.

- Wie groß ist die WK, dass Petra ihr Paket erst beim 5. Klingeln bekommt?
- Wie groß ist die WK, dass Petra ihr Paket spätestens beim 5. Klingeln bekommt?
- Wie viele Tage muss sie im Mittel warten, bis sie ihren Nachbarn antrifft?

Lösung:

Y sei die ZV, die beschreibt, ob Petra ihren Nachbarn am i -ten Tag antrifft:

$\{Y_i = 1\}$... sie trifft ihn am i -ten Tag an

$\{Y_i = 0\}$... sie trifft ihn am i -ten Tag nicht an

laut Voraussetzung: Y_i $i = 1, 2, \dots$ sind unabhängig und $\mathbb{P}(Y_i = 1) = 0,3 = 1 - \mathbb{P}(Y_i = 0)$

Wir führen also unabhängige, identische Bernoulli-Zufallsexperimente aus und fragen uns nach dem ersten Erfolg.

(a) gesucht:

$$\begin{aligned} \mathbb{P}(Y_1 = 0, \dots, Y_4 = 0, Y_5 = 1) &= \mathbb{P}(Y_1 = 0) \cdot \dots \cdot \mathbb{P}(Y_4 = 0) \cdot \mathbb{P}(Y_5 = 1) \\ &= 0,7^4 \cdot 0,3 \\ &= 0,07203 \end{aligned}$$

Setzen wir X ... Tag an dem der erste „Erfolg“ eintritt, so gilt $\mathbb{P}(X = 5) = 0,7^4 \cdot 0,3$

Allgemein gilt $P(X = m) = 0,7^{m-1} \cdot 0,3 \Rightarrow X$ ist geometrisch verteilt mit Parameter $p = 0,3$.

(b) Gesucht:

$$\begin{aligned} \mathbb{P}(X \leq 5) &= \mathbb{P}(\{X = 1\} \cup \{X = 2\} \cup \{X = 3\} \cup \{X = 4\} \cup \{X = 5\}) \\ &= \mathbb{P}(X = 1) + \dots + \mathbb{P}(X = 5) \\ &= 0,7^0 \cdot 0,3 + 0,7^1 \cdot 0,1 + \dots + 0,7^4 \cdot 0,3 \\ &= (0,7^0 + \dots + 0,7^4) \cdot 0,3 \\ &= 0,8919 \end{aligned}$$



Diese Rechnung funktioniert für beliebiges m . Daher gilt:

$$F_X(m) = \mathbb{P}(X \leq m) = 0,3 \sum_{i=0}^{m-1} 0,7^i = 0,3 \frac{1 - 0,7^m}{1 - 0,7} = 1 - 0,7^m$$

(c) Gesucht: $\mathbb{E}(X) = \frac{1}{0,3} = \frac{10}{3}$
 $\Rightarrow$ erwartete Wartezeit ist $3, \bar{3}$ Tage.

1.3-18 Bemerkung Für eine geometrisch verteilte ZV X mit Parameter p gilt

$$F_X(m) = \mathbb{P}(X \leq m) = 1 - (1 - p)^m$$

für $n \in \mathbb{N}_0$. Dazwischen ist F_X konstant.

1.3.1.6 POISSON-VERTEILUNG

1.3-19 Definition Eine ZV X heißt Poisson-verteilt mit dem Parameter $\lambda > 0$, falls sie nur die Werte $0, 1, 2, \dots$ annehmen kann und

$$p_m := \mathbb{P}(X = m) = \frac{\lambda^m}{m!} e^{-\lambda} \quad m = 0, 1, 2, \dots$$

1.3-20 Bemerkung

- Kurzschreibweise: $X \sim \text{Poi}(\lambda)$
- Erwartungswert: $\mathbb{E}(X) = \lambda$
- Varianz: $\text{Var}(X) = \lambda$
- Anwendung: Bedientheorie, Zuverlässigkeitstheorie
 - Anzahl der Kunden pro Zeiteinheit
 - Anzahl der Störungen im Produktionsprozess eines Betriebs pro Zeiteinheit

1.3-21 Beispiel In einer Fließbandanlage tritt im Durchschnitt alle 5 Stunden eine Störung auf (d.h. im Mittel 0,2 Störungen pro Stunde). Die Zahl der Störungen in einer bestimmten Zeiteinheit kann als Poissonverteilt angenommen werden.

Wie groß ist die WK, dass in einer 8-Stunden-Schicht mehr also 2 Störungen auftreten?

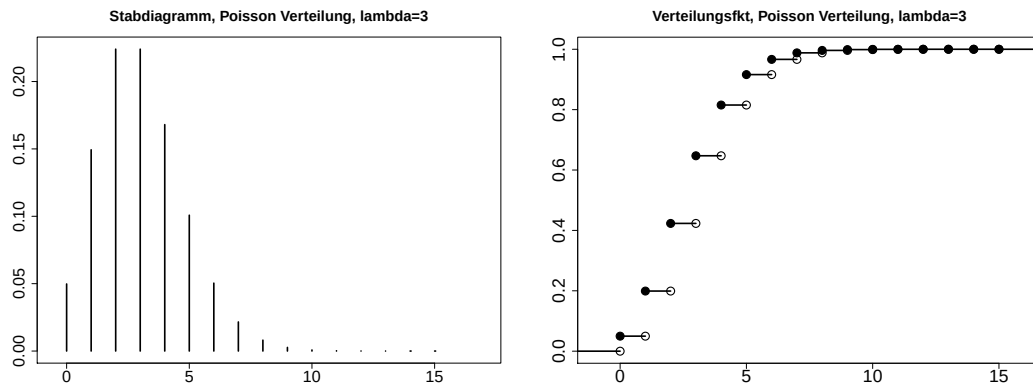
Lösung:

$X \dots$ Anzahl der Störungen in 8 Stunden

- $X \sim \text{Poi}(\lambda)$ mit $\mathbb{E}X = 8 \cdot 0,2 = 1,6 = \lambda$

$$\begin{aligned} \Rightarrow \mathbb{P}(X > 2) &= 1 - \mathbb{P}(X \leq 2) \\ &= 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) - \mathbb{P}(X = 2) \\ &= 1 - e^{-1,6} \left(\frac{1,6^0}{0!} + \frac{1,6^1}{1!} + \frac{1,6^2}{2!} \right) \\ &= 0,2166 \end{aligned}$$





1.3.2 SPEZIELLE STETIGE VERTEILUNGEN

- (D1) stetige Gleichverteilung (Bus-Beispiel)
- (D2) Normalverteilung (Zentraler Grenzwertsatz → wichtig!)
- (D3) Exponentialverteilung
- (D4) χ^2 -Verteilung (Chi-Quadrat-Verteilung)
- (D5) t -Verteilung
- (D6) F -Verteilung

S4-S5 vor allem für Statistik relevant

1.3.2.1 STETIGE GLEICHVERTEILUNG

1.3-22 Definition Eine ZV X heißt stetig gleichverteilt auf dem Intervall $I \subset \mathbb{R}$, falls für alle Intervalle $J \subset I$ gilt:

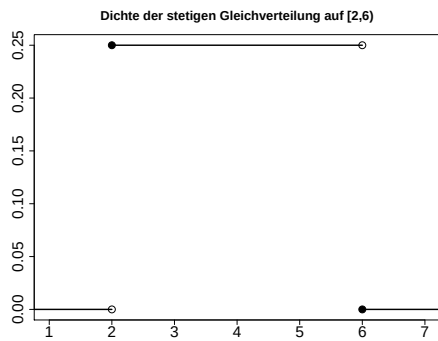
$$\mathbb{P}(X \in J) = \frac{|J|}{|I|}$$

1.3-23 Bemerkung

- Kurzschreibweise: $X \sim U(I)$
- Sei $a < b$. Ein Intervall I kann die Form (a, b) , $[a, b]$, $[a, b)$ oder $(a, b]$ haben. Dann gilt $|I| = b - a$. Ist I von dieser Form, so gilt:
- $\mathbb{E}(X) = \frac{a+b}{2}$, $\text{Var}(X) = \frac{1}{12}(b-a)^2$
- Dichte und Verteilungsfunktion:

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{sonst} \end{cases}, \quad F(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$$





1.3.2.2 NORMALVERTEILUNG

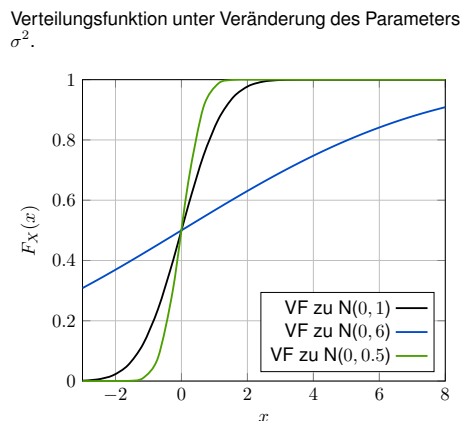
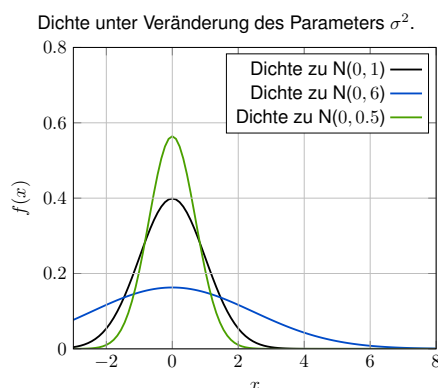
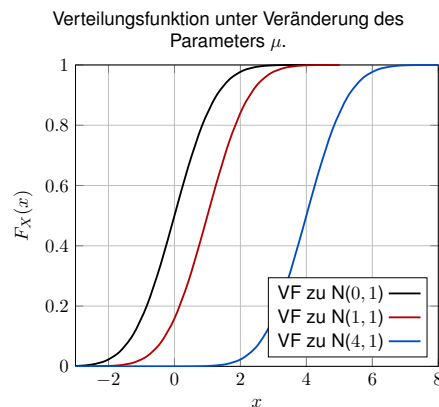
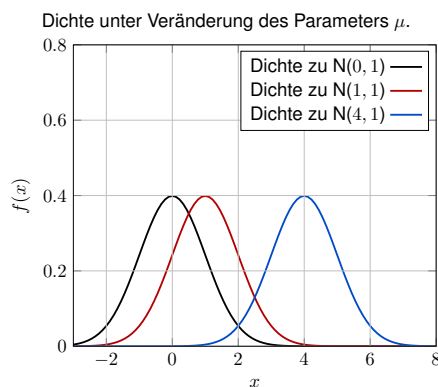
1.3-24 Definition Eine ZV X heißt normalverteilt mit den Parametern μ und σ^2 , ($\mu \in \mathbb{R}$, $\sigma > 0$), wenn sie die Dichte

$$f: \mathbb{R} \rightarrow \mathbb{R}, \quad f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

besitzt.

1.3-25 Bemerkung

- Kurzschreibweise: $X \sim \mathcal{N}(\mu, \sigma^2)$
- $\mathbb{E}(X) = \mu$
- $\text{Var}(X) = \sigma^2$
- Verteilungsfunktion nicht in geschlossener Form angebar (nur als Integraldarstellung oder unendliche Reihe)



1.3-26 Satz Gilt $X \sim \mathcal{N}(\mu, \sigma^2)$, dann gilt für die transformierte ZV

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

Wir sagen dann: Y ist standardnormalverteilt.

1.3-27 Bemerkung Für $Y \sim \mathcal{N}(0, 1)$ gilt:

- für die Verteilungsfunktion

$$\begin{aligned}\Phi(x) &:= F_Y(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt \\ &= 0,5 + \frac{1}{\sqrt{2\pi}} \int_0^x \exp\left(-\frac{t^2}{2}\right) dt\end{aligned}$$

- Werte von Φ lassen sich Tabellen oder Software entnehmen.
- Jede beliebige Normalverteilung lässt sich auf die Standardnormalverteilung Φ zurückführen.
- Symmetrie: $\Phi(-x) = 1 - \Phi(x)$

1.3-28 Bemerkung Für $X \sim \mathcal{N}(\mu, \sigma^2)$ gilt:

- $F_X(x) = \mathbb{P}(X \leq x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$
- $\mathbb{P}(X \geq a) = 1 - \Phi\left(\frac{a - \mu}{\sigma}\right)$
- $\mathbb{P}(a \leq X \leq b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$
- $\mathbb{P}(X = a) = 0$
- Anwendung:
 - Messfehler
 - geometrische und physikalische Kenngrößen von Produkten (Länge, Masse, Widerstand, ...)
 - biologische Merkmale
 - allgemein: Summe großer Anzahl von unabhängigen Größen

1.3-29 Beispiel (Drehteil) Ein Drehteil besitzt einen Soll-Durchmesser von 500 mm. Die Toleranzgrenzen sind 499,6 mm und 500,3 mm.

Die von der Maschine hergestellten Teile besitzen in Wirklichkeit (statistisch überprüft) einen normalverteilten Durchmesser mit $\mu = 500$ und $\sigma = 0,2$.

Wie groß ist die WK, dass ein solches Teil. . .

- innerhalb der Toleranzgrenzen liegt?
- einen Durchmesser kleiner als die untere Toleranzgrenze hat?



- (c) Wie genau muss die Maschine arbeiten (also wie groß darf σ sein) damit maximal 1% der produzierten Teile einen Durchmesser von maximal 499,6 mm haben?

Lösung:

X ... Durchmesser in mm

dann $X \sim \mathcal{N}(500, 0,2^2)$.

(a)

$$\begin{aligned}
 \mathbb{P}(499,6 \leq X \leq 500,3) &= \Phi\left(\frac{500,3 - 500}{0,2}\right) - \Phi\left(\frac{499,6 - 500}{0,2}\right) \\
 &= \Phi(1,5) - \Phi(-2) \\
 &= \Phi(1,5) - (1 - \Phi(2)) \\
 &= \Phi(1,5) + \Phi(2) - 1 \\
 &= 0,93319 + 0,97725 - 1 \quad | \text{ Tabelle} \\
 &= 0,91044 \\
 &\approx 91\%
 \end{aligned}$$

(b) $\mathbb{P}(X < 499,6) = \Phi\left(\frac{499,6 - 500}{0,2}\right) = 1 - \Phi(2) = 1 - 0,97725 = 0,02275$

(c) Nun ist σ unbekannt. Also:

$X \sim \mathcal{N}(500, \sigma^2)$

Wir suchen σ mit

$$\begin{aligned}
 0,01 &\geq \mathbb{P}(X < 499,6) = \Phi\left(\frac{499,6 - 500}{\sigma}\right) \\
 &= \Phi\left(\frac{-0,4}{\sigma}\right) \\
 &= 1 - \Phi\left(\frac{0,4}{\sigma}\right) \\
 \Leftrightarrow 0,01 &\stackrel{!}{=} 1 - \Phi\left(\frac{0,4}{\sigma}\right) \\
 \Leftrightarrow \Phi\left(\frac{0,4}{\sigma}\right) &= 0,99 \\
 \Leftrightarrow \frac{0,4}{\sigma} &= \Phi^{-1}(0,99) =: z_{0,99} \quad 0,992\text{-Quantil} \\
 z_{0,99} &= 2,326 \quad | \text{ Tabelle} \\
 \Rightarrow \sigma &= 0,172
 \end{aligned}$$

Bei ein Standardabweichung von max. 0,172 wird im Mittel höchstens 1% Ausschuss produziert.

1.3-30 Bemerkung Sei Φ die Verteilungsfunktion zur Standard-Normalverteilung.

- Für $\alpha \in (0, 1)$ benötigt man oft $\Phi^{-1}(\alpha)$.
Dies ist das α -Quantil q_α .
- Im Fall der Normalverteilung schreibt man oft $q_\alpha = z_\alpha \quad \left(= \Phi^{-1}(\alpha) \right)$.
- Die Werte z_α entnimmt man einer Tabelle (o.ä.).
- $z_\alpha = -z_{1-\alpha}$



1.3-31 Satz Seien $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ und $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ unabhängige ZVen. Dann:

$$X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

1.3.2.3 EXPONENTIALVERTEILUNG

1.3-32 Definition Die ZV X heißt exponentialverteilt mit dem Parameter $\lambda > 0$, wenn sie die folgende Dichte besitzt:

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{falls } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

1.3-33 Bemerkung

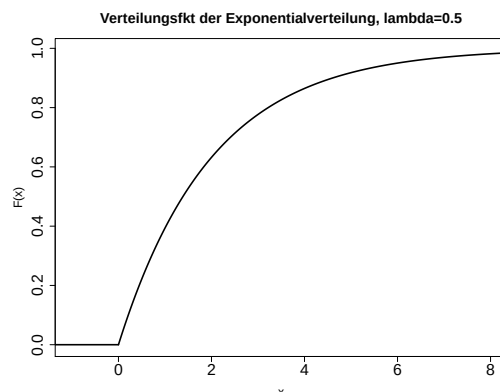
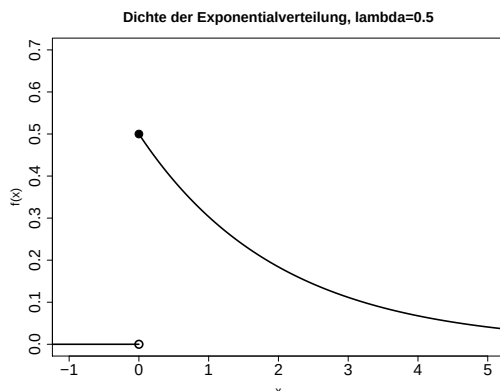
- Kurzschreibweise: $X \sim \text{Exp}(\lambda)$

- $\mathbb{E}(X) = \frac{1}{\lambda}, \quad \text{Var}(X) = \frac{1}{\lambda^2}$

- Verteilungsfunktion:

$$F(x) = \begin{cases} 1 - \exp(-\lambda x) & \text{falls } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

- Anwendung: Bedientheorie, Zuverlässigkeitstheorie, Verteilung von Zeitdauern wie Lebenszeiten, Reparaturzeiten, Wartezeiten, ...



1.3-34 Beispiel (Parallelschaltung) Ein System besteht aus 3 unabhängig voneinander arbeitenden, parallel geschalteten Elementen. Es sei bekannt, dass die Lebensdauer der einzelnen Elemente exponentialverteilt ist. Die mittlere Lebensdauer eines Elementes ist 1000 [h]

- Wie groß ist die WK, dass Element 1 höchstens 500 Std. funktioniert?
- Wie groß ist die WK, dass das System mindestens 500 Std funktioniert?
- Für welchen Zeitraum beträgt die Zuverlässigkeit des Systems 99%

Lösung:

X_i ... zufällige Lebensdauer von Element i , $i = 1, 2, 3$

$$\Rightarrow X_i \sim \exp(\lambda), \quad \mathbb{E}(X_i) = \frac{1}{\lambda} = 1000 \Rightarrow \lambda = \frac{1}{1000}$$

(a) ges.: $\underbrace{\mathbb{P}(X_1 \leq 500)}_{F_{X_1}(500)} = 1 - \exp\left(-\frac{1}{1000} \cdot 500\right) = 1 - \exp\left(-\frac{1}{2}\right) = 0,3935$



(b) X ... Lebensdauer des Systems

$$X = \max(X_1, X_2, X_3)$$

Wie ist X verteilt?

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) = \mathbb{P}(\max(X_1, X_2, X_3) \leq x) \\ &= \mathbb{P}(\{X_1 \leq x\} \cap \{X_2 \leq x\} \cap \{X_3 \leq x\}) \\ &= \mathbb{P}(X_1 \leq x) \cdot \mathbb{P}(X_2 \leq x) \cdot \mathbb{P}(X_3 \leq x) \\ &= \left(1 - \exp\left(-\frac{1}{1000}x\right)\right)^3 \quad \text{für alle } x \geq 0 \end{aligned}$$

gesucht:

$$\begin{aligned} \mathbb{P}(X \geq 500) &= 1 - \mathbb{P}(X \leq 500) \\ &= 1 - \left(1 - \exp\left(-\frac{500}{1000}\right)\right)^3 \\ &= 0,9391 \end{aligned}$$

(c) gesucht: Zeit t , so dass $X \geq t$ mit WK von mind. 0,99. Also: $0,99 \leq \mathbb{P}(X \geq t)$

$$\begin{aligned} &0,99 = \mathbb{P}(X \geq t) \\ \Leftrightarrow &0,99 = 1 - F_X(t) \\ \Leftrightarrow &F_X(t) = 0,01 \\ \Leftrightarrow &\left(1 - \exp\left(-\frac{t}{1000}\right)\right)^3 = 0,01 \\ \Leftrightarrow &t = 242,6 \text{ h} \end{aligned}$$

1.3.2.4 χ^2 -VERTEILUNG

1.3-35 Definition Eine stetige ZV heißt χ^2 -verteilt mit $n \in \mathbb{N}$ Freiheitsgraden, falls X die Dichte

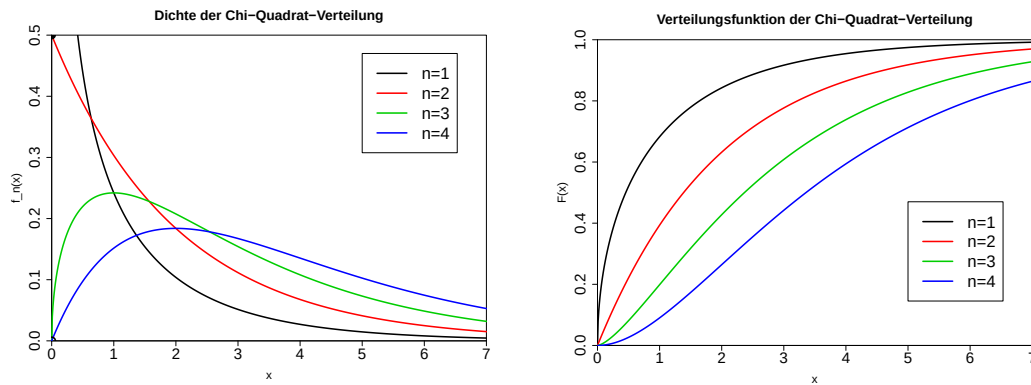
$$f_n(x) = \begin{cases} \frac{x^{\frac{n}{2}-1} \exp\left(-\frac{x}{2}\right)}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} & \text{falls } x > 0 \\ 0 & \text{sonst} \end{cases} \quad (x \in \mathbb{R})$$

besitzt. Hier ist Γ die Gammafunktion, d.h. $\Gamma(x) := \int_0^\infty t^{x-1} e^{-t} dt$ für $x > 0$.

1.3-36 Bemerkung

- Kurzschreibweise: $X \sim \chi^2(n)$
- $\mathbb{E}(X) = n$, $\text{Var}(X) = 2n$
- Verteilungsfunktion zeigen wir hier nicht (lässt sich mittels Gammafunktion darstellen)
- Quantile: Ist $\alpha \in (0, 1)$ und F_X die Verteilungsfunktion zu $X \sim \chi^2(n)$, so bezeichnen wir das α -Quantil q_α mit $\chi_{n,\alpha}^2 = q_\alpha = F_X^{-1}(\alpha)$
- Anwendung: Statistik, insbesondere Testtheorie





Ein Grund für die große Bedeutung der χ^2 -Verteilung (in der Statistik) ist:

1.3-37 Satz Seien $X_1, X_2, \dots, X_n$ unabhängige standard-normalverteilte ZVen. Dann ist

$$X := X_1^2 + X_2^2 + \dots + X_n^2$$

χ^2 -verteilt mit n Freiheitsgraden.

1.3.2.5 T-VERTEILUNG

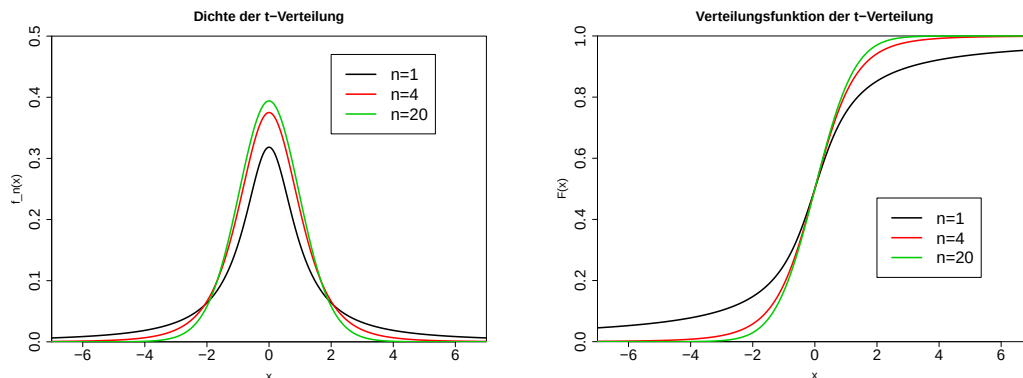
1.3-38 Definition Eine stetige ZV X heißt t -verteilt mit $n \in \mathbb{N}$ Freiheitsgraden, falls X die Dichte

$$f_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \quad (x \in \mathbb{R})$$

besitzt. Hier ist Γ wieder die Gammafunktion.

1.3-39 Bemerkung

- Kurzschreibweise: $X \sim t(n)$
- falls $n > 1$: $\mathbb{E}(X) = 0$, für $n = 1$ existiert $\mathbb{E}(X)$ nicht
- falls $n > 2$: $\text{Var}(X) = \frac{n}{n-2}$, für $n = 1, 2$ existiert $\text{Var}(X)$ nicht
- Verteilungsfunktion zeigen wir nicht (lässt sich mittels Betafunktion darstellen)
- Quantile: Ist $\alpha \in (0, 1)$ und F_X die Verteilungsfunktion zu $X \sim t(n)$, so bezeichnen wir das α -Quantil q_α mit $t_{n,\alpha} = q_\alpha = F_X^{-1}(\alpha)$
- Anwendung: Statistik, insbesondere Testtheorie



Ein Grund für die große Bedeutung der t -Verteilung (in der Statistik) ist:



1.3-40 Satz Seien Y und Z unabhängige ZVen mit $Y \sim \chi^2(n)$ und $Z \sim \mathcal{N}(0, 1)$. Dann ist

$$X = \frac{Z}{\sqrt{\frac{Y}{n}}}$$

t -verteilt mit n Freiheitsgraden.

1.3.2.6 F-VERTEILUNG

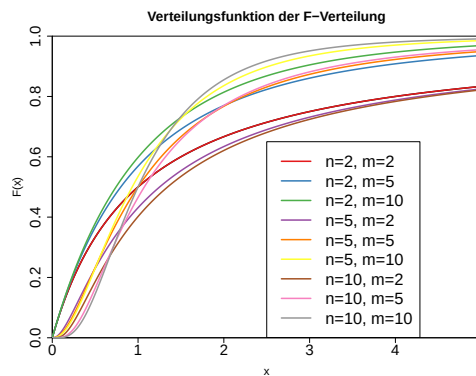
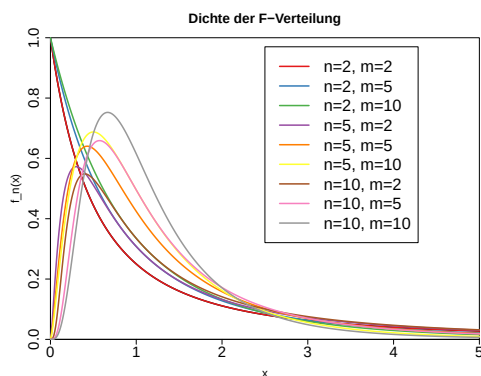
1.3-41 Definition Eine stetige ZV X heißt F -verteilt mit $m \in \mathbb{N}$ Freiheitsgraden im Zähler und $n \in \mathbb{N}$ Freiheitsgraden im Nenner, falls X die Dichte

$$f_{m,n}(x) = \begin{cases} m^{\frac{m}{2}} n^{\frac{n}{2}} \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \cdot \frac{x^{\frac{m}{2}-1}}{(mx+n)^{\frac{m+n}{2}}} & \text{falls } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

besitzt. Hier ist Γ wieder die Gammafunktion.

1.3-42 Bemerkung

- Kurzschreibweise: $X \sim F(m, n)$
- falls $n > 2$: $\mathbb{E}(X) = \frac{n}{n-2}$, für $n = 1, 2$ existiert $\mathbb{E}(X)$ nicht
- falls $n > 4$: $\text{Var}(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$, für $n = 1, 2, 3, 4$ existiert $\text{Var}(X)$ nicht
- Quantile: Ist F_X die Verteilungsfunktion zu $X \sim F(m, n)$, so bezeichnen wir das α -Quantil q_α mit $F_{m,n,\alpha} = q_\alpha = F_X^{-1}(\alpha)$
- Anwendung: Statistik, insbesondere Testtheorie



Ein Grund für die große Bedeutung der F -Verteilung (in der Statistik) ist:

1.3-43 Satz Seien Y und Z unabhängige ZVen mit $Y \sim \chi^2(m)$ und $Z \sim \chi^2(n)$. Dann ist

$$X = \frac{\frac{Y}{m}}{\frac{Z}{n}} \sim F(m, n).$$

1.4 GRENZWERTSÄTZE

Wir betrachten nun Folgen von ZVen $X_1, X_2, X_3, \dots$



1.4-1 Definition Sind $X_1, X_2, \dots$ ZVen, welche alle die gleiche Verteilungsfunktion haben, so sagen wir „die ZV sind IDENTISCH VERTEILT“.

1.4.1 GESETZ DER GROSSEN ZAHLEN

1.4-2 Beispiel Wir werfen eine faire Münze n mal und zählen die Ereignisse „Kopf“ und „Zahl“:

n	Anzahl Kopf	Anzahl Zahl	rel. Anz. Kopf
50	22	28	0,44
200	89	111	0,445
1000	493	507	0,493
100 000	50 256	49 744	0,50256

Die relative Anzahl Kopf nähert sich anscheinend der 0,5 immer weiter an. Diesen Zusammenhang formalisiert das Gesetz der großen Zahlen.

1.4-3 Satz Seien $X_1, X_2, \dots$ unabhängige und identisch verteilte ZVen mit Erwartungswert μ und Varianz σ^2 und sei

$$\overline{X_n} = \frac{1}{n} \cdot (X_1 + \dots + X_n)$$

das arithmetische Mittel der ersten n ZVen.

Dann gilt für jede (noch so kleine) Zahl $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\overline{X_n} - \mu| < \varepsilon) = 1.$$

Insbesondere gilt

$$\mathbb{P}(|\overline{X_n} - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{n \cdot \varepsilon^2}.$$

Um zu verstehen, warum das GdgZ gilt benötigen wir die folgende Ungleichung:

1.4-4 Satz (Tschebyschew-Ungleichung) Sei X eine ZV so dass $\text{Var}(X)$ existiert: Dann gilt für beliebiges $a > 0$:

$$\mathbb{P}(|X - \mathbb{E}X| \geq a) \leq \frac{\text{Var}(X)}{a^2}.$$

1.4-5 Bemerkung Die T -Ungleichung liefert eine obere Schranke an die WK, dass eine ZV um einen Mindestabstand a von ihrem Erwartungswert abweicht. Die Schranke ist klein, falls

- a groß ist
- $\text{Var}(X)$ klein ist

Beweis. Sei also X_1, X_2 unabhängig identisch verteilt mit Erwartungswert μ und Varianz σ^2 und sei $\overline{X_n} = \frac{1}{n} \sum_{i=1}^n X_i$. Dann gilt:

$$\mathbb{E}(\overline{X_n}) = \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n \underbrace{\mathbb{E}X_i}_{\mu} = \mu$$



und wegen der Unabhängigkeit gilt auch:

$$\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) \stackrel{\text{da unabh.}}{=} \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n}$$

Nun wenden wir Tschebyschew an:

$$\underbrace{\mathbb{P}(|\bar{X}_n - \mathbb{E}\bar{X}_n| \geq \varepsilon)}_{\mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon)} \leq \underbrace{\frac{\text{Var}(\bar{X}_n)}{\varepsilon^2}}_{\frac{\sigma^2}{n\varepsilon^2}}$$

$$\mathbb{P}(|\bar{X}_n - \mu| < \varepsilon) = 1 - \mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) \geq 1 - \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 1$$

□

1.4-6 Beispiel Gegeben: unabhängige Poisson-verteilte ZVen mit $X_i \sim \text{Poi}(3)$ für alle i

$\Rightarrow \mathbb{E}(X_i) = 3 = \text{Var}(X_i)$ für alle i

$\Rightarrow$ für $\varepsilon = 0,1$ und $n = 5000$ gilt $\mathbb{P}(|\bar{X}_n - 3| < 0,1) \geq 1 - \frac{3}{5000 - 0,1^2} = 0,94$

1.4.2 DER ZENTRALE GRENZWERTSATZ

1.4-7 Satz (Zentraler Grenzwertsatz)

Seien $X_1, X_2, \dots$ unabhängige und identisch verteilte Zufallsvariablen mit Erwartungswert μ und Varianz σ^2 . Für $n \in \mathbb{N}$ setzen wir

$$S_n = X_1 + \dots + X_n.$$

Für die standardisierte Zufallsvariable

$$Z_n = \frac{S_n - \mathbb{E}S_n}{\sqrt{\text{Var } S_n}} = \frac{S_n - n\mu}{\sqrt{n} \cdot \sigma}$$

gilt dann

$$\lim_{n \rightarrow \infty} \underbrace{\mathbb{P}(Z_n \leq z)}_{F_{Z_n}(z)} = \Phi(z) \quad (z \in \mathbb{R})$$

wobei Φ (wie immer) die Verteilungsfunktion der Standardverteilung ist.

1.4-8 Bemerkung

- Der Satz sagt aus, dass für großes n die ZV Z_n nahezu normalverteilt ist.
- Wesentlich: es ist KEINE Annahme über die Verteilung der X_i gemacht.

- mit $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ gilt

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$$

- Sprechweise: „ Z_n ist asymptotisch/approximativ (standard-)normalverteilt.“



- Schreibweise: Die Gleichung

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_1 \leq z) = \Phi(z) \quad (z \in \mathbb{R})$$

wird als

$$Z_n \stackrel{a}{\sim} \mathcal{N}(0, 1)$$

abgekürzt. In diesem Sinne kann man auch

$$S_n \stackrel{a}{\sim} \mathcal{N}(n\mu, n\sigma^2) \text{ und } \bar{X}_n \stackrel{a}{\sim} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

verstehen.

1.4-9 Beispiel Wir werfen einen fairen Würfel mehrmals hintereinander.

X_i ... Ergebnis des i -ten Wurfs

$S_n = \sum_{i=1}^n X_i$... Augensumme der ersten n Würfe

ZGWS: $S_n \stackrel{a}{\sim} \mathcal{N}(n\mu, n\sigma^2)$

S_n ist asymptotisch normalverteilt mit Erwartungswert $n \cdot \mu = 3,5$ und Varianz $n\sigma^2 = n \cdot 2,91\bar{6}$, denn:

$$\mu = \mathbb{E}(X_n) = \frac{1}{6}(1 + \dots + 6) = 3,5 \text{ und}$$

$$\sigma^2 = \text{Var}(X_n) = \frac{1}{6} \underbrace{(1^2 + \dots + 6^2)}_{\mathbb{E}(X_n^2)} - \underbrace{3,5^2}_{(\mathbb{E}X_n)^2} = 2,91\bar{6}$$

1.4-10 Beispiel Es werden Schrauben mit einem zufälligen Gewicht mit EW 4 g und Standardabweichung 0,3 g hergestellt. Berechnen Sie mit dem ZGWS (unter Annahme der Unabhängigkeit):

(a) WK, dass der Inhalt einer Packung mit 200 Schrauben maximal 795 g wiegt.

(b) Welches Nettogewicht wird eine 200er Packung mit WK 0,99 überschreiten?

Lösung:

X_i ... Gewicht der Schraube i in Gramm, $i = 1, \dots, 200$

$$S = \sum_{i=1}^{200} X_i$$

$$\Rightarrow \mathbb{E}S = 800 \quad \text{Var } S = \text{Var} \left(\sum_{i=1}^{200} X_i \right) = \sum_{i=1}^{200} \underbrace{\text{Var } X_i}_{0,3^2} = 200 \cdot 0,09 = 18$$



(a)

$$\begin{aligned}\mathbb{P}(S \leq 795) &= \mathbb{P}(S - \mathbb{E}S \leq 795 - \mathbb{E}S) \\&= \mathbb{P}\left(\underbrace{\frac{S - \mathbb{E}S}{\sqrt{\text{Var } S}}}_Z \leq \frac{795 - \mathbb{E}S}{\sqrt{\text{Var } S}}\right) \\&= \mathbb{P}\left(Z \leq \frac{-5}{\sqrt{18}}\right) \\&= \mathbb{P}(Z \leq -1,1785) \\&\approx \mathbb{P}(Z \leq -1,18) \quad (\text{runden}) \\&\approx \Phi(-1,18) \quad (\text{ZGWS}) \\&= 1 - \Phi(1,18) = 1 - 0,881 = 0,119\end{aligned}$$

Wahrscheinlichkeit ist etwa 20%.

(b) Gesucht: Gewicht a mit $\mathbb{P}(S > a) = 0,99$

$$\begin{aligned}0,99 &\stackrel{!}{=} 1 - \mathbb{P}(S \leq a) \\&= 1 - \mathbb{P}\left(\frac{S - \mathbb{E}S}{\sqrt{\text{Var } S}} \leq \frac{a - \mathbb{E}S}{\sqrt{\text{Var } S}}\right) \\&= 1 - \Phi\left(\frac{a - 800}{\sqrt{18}}\right) \\&\Leftrightarrow 0,01 = \Phi(\dots) \\&\Leftrightarrow \underbrace{\Phi^{-1}(0,01)}_{-2,3243} = \frac{a - 800}{\sqrt{18}} \\&\Leftrightarrow a = 790,13\end{aligned}$$

1.4-11 Bemerkung Bei der Anwendung des ZGWS spielt oft die Gleichung

$$\mathbb{P}\left(\sum_{i=1}^n X_i \leq a\right) = p$$

ein Rolle. Im Beispiel (a) war p gesucht (mit a und n gegeben), in (b) war a gesucht (mit p und n gegeben).

Es könnte auch mal n gesucht sein (siehe Hausaufgabe).
Spezialfall des ZGWS:

1.4-12 Satz (von Moivre-Laplace)

Gilt $S_n \sim \text{Bin}(n, p)$, so gilt

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq z\right) = \Phi(z)$$

bzw.

$$\frac{S_n - np}{\sqrt{np(1-p)}} \stackrel{a}{\sim} \mathcal{N}(0, 1).$$



1.4-13 Bemerkung

- Satz sagt, dass für GROSSES n nahezu $S_n \sim \mathcal{N}(np, np(1-p))$ gilt.
- FAUSTREGEL für großes n :

$$np(1-p) \geq 9$$

- Sind $X_1, X_2, \dots$ unabhängig identisch Bernoulli-verteilte ZVen mit Parameter p , dann

$$\mathbb{E}(X_1) = p, \quad \text{Var}(X_1) = p(1-p)$$

und

$$S_n = \sum_{i=1}^n X_i \sim \text{Bin}(n, p)$$

Formel aus Satz 1.4-14 folgt jetzt aus ZGWS.

1.4-14 Beispiel Betrachten den 12 000-fachen Münzwurf. Frage: Wie groß ist die WK, dass die Anzahl der Ergebnisse „Zahl“ weniger als 20 vom Erwartungswert abweichen?

Lösung:

S ... Anzahl des Eintretens von Zahl bei 12 000 Würfeln

$$\Rightarrow S \sim \text{Bin}(12\,000, 0,5)$$

$$\xRightarrow{\text{Moiivre-Laplace}} S \overset{a}{\sim} \mathcal{N}(6\,000, 3\,000)$$

$$\text{Es gilt: } \mathbb{E}(S) = 12\,000 \cdot 0,5 = 6\,000$$

$$\text{Gesucht: } \mathbb{P}(|S - 6\,000| \leq 19)$$

Wir berechnen diese WK approximativ mit Satz von Moivre-Laplace. Dazu:

$$\mathbb{P}(|S - 6\,000| \leq 19) = \mathbb{P}(|S - 6\,000| \leq 19,99) = \mathbb{P}(|S - 6\,000| \leq 19,5)$$

Diese Werte wären theoretisch gleich (weil Verteilung diskret), in der Approximation (die dann stetig Verteilt ist) dann aber nicht mehr. Daher nehmen wir für die Approximation den Wert zwischen den beiden Extremen: die 19,5 (Stetigkeits-Korrektur)!

Damit gilt:

$$\begin{aligned} \mathbb{P}(5\,980,5 \leq S \leq 6\,019,5) &\approx \Phi\left(\frac{6\,019,5 - 6\,000}{\sqrt{3\,000}}\right) - \Phi\left(\frac{5\,980,5 - 6\,000}{\sqrt{3\,000}}\right) \\ &= 2\Phi\left(\frac{19,5}{\sqrt{3\,000}}\right) - 1 \\ &= 0,27817416 \end{aligned}$$

$$\text{Bemerkung: Exakter Wert } \sum_{m=5\,901}^{6\,019} \binom{12\,000}{m} \cdot \left(\frac{1}{2}\right)^{12\,000} = 0,2781725$$



2 STATISTIK

2.1 DESKRIPTIVE STATISTIK

2.1.1 GRUNDBEGRIFFE

2.1-1 Definition (Grundgesamtheit und Merkmale)

- Grundgesamtheit Ω ... klar festgelegte Menge von gleichartigen Objekten, die hinsichtlich bestimmter Eigenschaften untersucht werden sollen
- $\omega \in \Omega$... Merkmalsträger / statistische Einheit / Untersuchungseinheit
- Merkmal ... die in der Grundgesamtheit zu untersuchende Eigenschaft
- Zustandsraum S ... Menge der möglichen Merkmalsausprägungen / unterschiedliche Eigenschaften
- $s \in S$... Merkmalsausprägung
Mathematische Darstellung:

$$X : \Omega \rightarrow S, \omega \mapsto x$$

Beachte: X ordnet jedem Element aus Ω ein Merkmal zu. X wird daher oft selbst als MERKMAL bezeichnet.

2.1-2 Beispiel (Notenspiegel)

- $\Omega = \{\text{alle Schüler einer Klasse}\}$
- $S = \{1, 2, 3, 4, 5, 6\}$
- X ... Funktion, welche jedem Schüler seine Zensur zuordnet, bspw. $X(\omega_1) = 3$; $X(\omega_2) = 5$;

2.1-3 Beispiel (medizinischer Fragebogen zur Pneumonie)

- $\Omega = \{\text{alle Patienten, die in einem bestimmten Jahr an Pneumonie erkrankt sind}\}$
- abfrage Merkmale: Alter, Geschlecht, Größe, Dauer des Krankenhausaufenthalts
- Mehrdimensionaler Zustandsraum, z.B.:
 $X(\omega_1) = (65, "m", 182 \text{ cm}, 5 \text{ d}) \in S$
 $X(\omega_2) = (34, "w", 162 \text{ cm}, 2 \text{ d}) \in S$
- X ... Funktion, welche jedem Patienten seine Merkmals-Vektoren zuordnet
- Beachte: Mehrdimensionale Merkmale behandelt man in der MULTIVARIATEN STATISTIK



2.1-4 Bemerkung Merkmale lassen sich auf verschiedene Weisen in Klassen einteilen. Wir betrachten 3 dieser möglichen Einteilungen:

(A) Ein Merkmal heißt

- **QUALITATIVES MERKMAL**, falls die Ausprägungen eine Qualität wiedergeben (und nicht ein Ausmaß). Insbesondere gibt es nur endlich viele Ausprägungen. Für qualitative Merkmale gibt es keine zwingende Ordnung/Reihenfolge.
Beispiele: Geschlecht, Religionszugehörigkeit oder Parteipräferenz
- **QUANTITATIVES MERKMAL**, falls die Ausprägung ein Ausmaß bzw. eine Intensität widerspiegeln. Die Ausprägungen sind in diesem Fall Zahlen (mit oder ohne Maßeinheit).
Beispiele: Alter, Größe oder Einkommen.

(B) Ein Merkmal heißt

- **DISKRET**, wenn es endlich viele oder abzählbar unendlich viele Ausprägungen annehmen kann.
Beispiele: Zensuren, Einwohnerzahl, Produktionszahlen einer Maschine an verschiedenen Tagen
- **STETIG**, wenn überabzählbar viele Ausprägungen angenommen werden können.
Beispiele: Gewicht, Zeitmessung in 100 m Lauf, Länge einer Schraube

(C) Ein Merkmal heißt

- **NOMINALSKALIERT**, genau dann wenn es qualitativ ist (also qualitatives Merkmal = nominalskaliertes Merkmal)
Beispiele: Geschlecht, Religionszugehörigkeit oder Parteipräferenz
- **ORDINALSKALIERT**, wenn es eine Rangordnung der Merkmalausprägung gibt, jedoch die Abstände zwischen den Merkmalsausprägungen nicht interpretiert werden können.
Beispiele: Dienstrang beim Militär, Zufriedenheit mit Produkt (gut > mittel > schlecht)
- **METRISCH SKALIERT (ODER KARINALSKALIERT)**, falls es eine Rangordnung der Merkmalsausprägung gibt und die Abstände zwischen den Ausprägungen messbar und interpretierbar sind.
Weitere Unterscheidung für metrisch skalierte Merkmale:
 - **INTERVALLSKALA**: natürlicher Nullpunkt existiert nicht
Beispiele: IQ-Skala, Temperatur in Celsius-Skala, Jahreszahlen
 - **VERHÄLTNISSKALA**: natürlicher Nullpunkt existiert
Beispiele: Zeitdauer, Masse, Preis

2.1-5 Bemerkung Statistisch Sinnvolle Auswertungen

- **NOMINALSKALA**
 - Häufigkeiten durch Zählen der einzelnen Ausprägungen
 - geeignetes Lagemaß: Modalwert
 - kein sinnvolles Streuungsmaß
- **ORDINALSKALA**
 - Häufigkeiten durch Zählen der einzelnen Ausprägungen
 - geeignetes Lagemaß: Modalwert, Median



- geeignetes Streuungsmaß: Spannweite
- METRISCHE SKALA
 - Häufigkeiten durch Zählen der einzelnen Ausprägungen
 - geeignetes Lagemaß: Modalwert, Median, arithmetisches Mittel
 - geeignetes Streuungsmaß: Spannweite, Standardabweichung, Varianz, ...

2.1-6 Bemerkung

- Bei Nominal und Ordinalskala sind keine Rechenoperationen wie Addition, Subtraktion, Multiplikation oder Division erlaubt.
- Bei Intervallskala ist Differenzenbildung erlaubt (jedoch keine Quotienten), da kein natürlicher Nullpunkt existiert.
- Bei Verhältnisskala ist Quotientenbildung erlaubt (jedoch keine Differenzen), da natürlicher Nullpunkt existiert.

Nun wollen wir Stichproben einführen.

Vorüberlegung:

- Ist $X : \Omega \rightarrow S$, $\omega \mapsto x$ ein Merkmal mit $S \subseteq \mathbb{R}$ (metrische Skala), so interessiert uns wie dieses Merkmal auf der Grundgesamtheit (GG) verteilt ist, d.h.

$$F_X(z) = \mathbb{P}(X \leq z)$$

(das WK-Maß $\mathbb{P}$ entsteht durch zufälliges (gleichverteiltes) rausgreifen eines Merkmalsträgers aus Ω)

Verteilung des Merkmals in $\Omega \rightsquigarrow$ Verteilung $\mathbb{P}$ bzw. F_X

z.B. ein Viertel der Personen in der GG hat Körpergröße $> 1,8 \text{ m} \rightsquigarrow \mathbb{P}(X > 1,8) = 0,25$

- Problem: Oft ist Ω zu groß, als dass man alle Werte $X(\omega)$, $\omega \in \Omega$ erheben kann (Gründe sind etwa: Kosten, Zeit, ...).
- Idee: Einschränkung auf möglichst „representative“ Teilmenge von Messungen der Merkmale. Berechnung der Kennzahlen, Eigenschaften, ... auf dieser Teilmenge.
- Hoffnung: Diese Berechnung geben uns Aufschluss über die Zusammensetzung der Merkmale.
- Ziehen daher Stichprobe aus den Daten.

2.1-7 Definition Sei ein Merkmal X gegeben und seien $X_1, \dots, X_n$ unabhängige, identische wie X verteilte Zufallsvariablen. Dann heißt der Vektor

$$\underline{X} = (X_1, \dots, X_n)^T$$

mathematische Stichprobe vom Umfang n . Jede Realisierung $\underline{x} = (x_1, \dots, x_n)^T$ von $\underline{X}$ heißt konkrete Stichprobe (Beobachtungsreihe).



2.1-8 Bemerkung Sei $X : \Omega \rightarrow S$, $\omega \mapsto x$ ein Merkmal

- Um die Stichprobe vom Umfang n zu modellieren wählen wir n unabhängige identisch (wie X) verteilte ZV:

$$X_1, \dots, X_n \quad (\text{große Buchstaben})$$

(Vor der Beobachtung, Mathematische Stichprobe $\rightsquigarrow$ Induktive Statistik)

- Nach der Auswertung dieser Variablen (einsetzen von ω) erhalten wir Realisierungen dieser Zufallsvariablen:

$$x_1, \dots, x_n \quad (\text{kleine Buchstaben})$$

(Nach der Beobachtung, konkrete Stichprobe $\rightsquigarrow$ Deskriptive Statistik)

2.1.2 EINDIMENSIONALES DATENMATERIAL

Erinnerung:

eindimensional

- $S = \{1, \dots, 6\}$ (Schulnoten, Würfel)
- $S = \{0, 1\}$ (Geschlecht, Münze, ...)

mehrdimensional

- $S = \mathbb{R}^2$ (Körpergröße und Gewicht)
- $S = \{1, \dots, 6\}^2$ (2 mal würfeln)

2.1.2.1 STICHPROBENFUNKTIONEN

2.1-9 Definition Sei $(X_1, \dots, X_n)$ eine mathematische Stichprobe. Sei f eine Funktion auf S^n , also $f : S^n \rightarrow \mathbb{R}$, $(x_1, \dots, x_n) \mapsto f(x_1, \dots, x_n) = y$.

Dann heißt die Zufallsvariable

$$T := f(X_1, \dots, X_n)$$

STICHPROBENFUNKTION.

Es folgen spezielle Stichprobenfunktionen.

2.1-10 Definition Sei $(X_1, \dots, X_n)$ eine mathematische Stichprobe zum Merkmal X . Wir definieren:

- (Stichproben-)Mittelwert:

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

- (Stichproben-)Streuung/Varianz

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

- (Stichproben-)Standardabweichung

$$S = \sqrt{S^2}$$

- Variationskoeffizient

$$V = \frac{S}{\bar{X}}$$



- Spannweite

$$R = \max(X_1, \dots, X_n) - \min(X_1, \dots, X_n)$$

Sei $X = (X_1, \dots, X_n)$ dann bezeichnet

$$(X_{(1)}, X_{(2)}, \dots, X_{(n)})$$

den Vektor der geordneten Stichproben, d.h.

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}.$$

Wir definieren damit

- α -Quantil $\tilde{X}_\alpha$ mittels

$$\tilde{X}_\alpha := \begin{cases} X_{(k)} & \text{falls } \alpha n \text{ keine ganze Zahl ist und} \\ & k \text{ kleinste ganze Zahl größer } \alpha n. \\ \frac{1}{2}(X_{(\alpha n)} + X_{(\alpha n+1)}) & \text{falls } \alpha n \text{ ganzzahlig.} \end{cases}$$

dann gilt:

- (1) höchstens $\alpha \cdot n$ Stichprobenwerte sind kleiner als $\tilde{X}_\alpha$.
- (2) höchstens $(1 - \alpha)n$ SP-Werte sind größer als $\tilde{X}_\alpha$

- Median := $\tilde{X}_{0,5}$
- Inter-Quartilsabstand (Inter-Quartil-Range)

$$IQR = \tilde{X}_{0,75} - \tilde{X}_{0,25}$$

2.1-11 Definition Sei $X_1, \dots, X_n$ eine mathematische Stichprobe. Dann heißt

$$S_n : \mathbb{R} \rightarrow \mathbb{R}, S_n(z) = \frac{\text{Anzahl der } X_i \text{ mit } X_i \leq z}{n}$$

empirische Verteilungsfunktion der Stichprobe.

2.1-12 Bemerkung

- Bezeichnungen für konkrete SP $(x_1, \dots, x_n)$ analog, nur mit kleinen Buchstaben:

$$\bar{x}, s^2, s, v, \tilde{x}_\alpha, s_n(z)$$

- Stichproben-Mittelwert, -Varianz, -Standardabweichung, Variationskoeffizient sind nur sinnvoll für METRISCH skalierte Merkmale!
- Quantile, Median, IQR und empirische Verteilungsfunktionen auch sinnvoll für ordinal skalierte Daten.
- Für nominal skalierte Merkmale ist keine der oben genannten Funktionen sinnvoll. Hier verwendet man z.B. den Modalwert (Häufigkeit aufgetretener Werte) zur Charakterisierung.

Die empirische Verteilungsfunktion S_n ist eine Näherung der theoretischen Verteilungsfunktion F von X :



2.1-13 Satz (Glivenko-Cantelli, Hauptsatz der Statistik)

Sei $X_1, X_2, \dots$ eine Folge von unabhängigen, identisch mit Verteilungsfunktion F verteilten ZVen und S_n die empirische VF von den ersten n Zufallsvariablen. Dann gilt für jede noch so kleine Zahl $\varepsilon > 0$ und jedes $x \in \mathbb{R}$:

$$\lim_{n \rightarrow \infty} \mathbb{P}(|S_n(x) - F(x)| < \varepsilon) = 1$$

2.1.2.2 AUFBEREITUNG STATISTISCHER DATEN AM BEISPIEL

2.1-14 Beispiel X ... Anzahl der Störungen im Maschinenpark eines Betriebes in einer Woche

- $n = 20$ Beobachtungen (20 verschiedene Wochen)
- konkrete Stichprobe:
 $(x_1, \dots, x_{20}) = (4, 2, 6, 3, 3, 1, 5, 2, 2, 1, 0, 4, 2, 5, 5, 3, 7, 2, 1, 3)$
- Beobachtete Ausprägungen a_j :
 $a_1 = 0, a_2 = 1, \dots, a_8 = 7$

- Häufigkeitstabelle:

Ausprägung a_j	Abs. Häufigkeit h_j	Summen der abs. Häufigkeiten $\sum_{i=1}^j h_i$	relative Häufigkeit w_j	Summe der rel. Häufigkeiten $s_j = \sum_{i=1}^j w_i$
0	1	1	0,05	0,05
1	3	4	0,15	0,2
2	5	9	0,25	0,45
3	4	13	0,2	0,65
4	2	15	0,1	0,75
5	3	18	0,15	0,9
6	1	19	0,05	0,95
7	1	20	0,05	1

Beachte: bei metrisch oder ordinal skalierten Merkmalen ordnet man die Ausprägungen der Größe nach.

Summen sind auch nur für metrisch oder ordinal skalierte Merkmale sinnvoll.

- Graphische Darstellung mittels Stabdiagramm:
 - absolute/relative Häufung h_j/w_j auf y -Achse
 - Ausprägung a_j auf x -Achse

2.1.3 ZWEIDIMENSIONALES DATENMATERIAL

2.1.3.1 ANSCHAUUNG AM BEISPIEL

- Zwei Merkmale X und Y werden an n Untersuchungseinheiten beobachtet.



- Wir erhalten eine 2-dimensionale mathematische Stichprobe

$$(X_1, Y_1), \dots, (X_n, Y_n)$$

und die konkrete Stichprobe

$$(x_1, y_1), \dots, (x_n, y_n)$$

- Ausprägungen von $X : a_1, \dots, a_l, Y : b_1, \dots, b_m$
- Ausprägungen von $(X, Y) : (a_j, b_k)$ mit $j \in \{1, \dots, l\}, k \in \{1, \dots, m\}$
- absolute Häufigkeit von (a_j, b_k) wird mit h_{jk} bezeichnet.

DARSTELLUNGSMÖGLICHKEITEN

(1) Häufigkeitstabelle

Die Häufigkeiten $H = (h_{jk})$ werden in einer Tabelle dargestellt. Diese heißt auch Kreuztabelle oder Kontingenztafel.

Besonders geeignet für diskrete Merkmale mit wenigen Ausprägungen.

(2) Streudiagramm

Graphische Veranschaulichung, wobei die Werte $(x_1, y_1), \dots, (x_n, y_n)$ in einem 2-dimensionalen Koordinatensystem als Punkte dargestellt werden.

Besonders geeignet für stetige Merkmale oder diskrete Merkmale mit beiden Ausprägungen.

2.1-15 Beispiel

- X ... Anzahl der Störungen im Maschinenpark eines Betriebes innerhalb einer Woche
- X ist quantitativ, metrisch skaliert, diskret
- $n = 20$ Beobachtungen (20 Arbeitsperioden von je einer Woche)
- Konkrete Stichprobe:

$$(x_1, \dots, x_{20}) = (4, 2, 6, 3, 3, 1, 5, 2, 2, 1, 0, 4, 2, 5, 5, 3, 7, 2, 1, 3)$$

- Beobachtete Ausprägungen a_j :

$$a_1 = 0, a_2 = 1, \dots, a_8 = 7$$

- Häufigkeitstabelle schafft Übersicht:

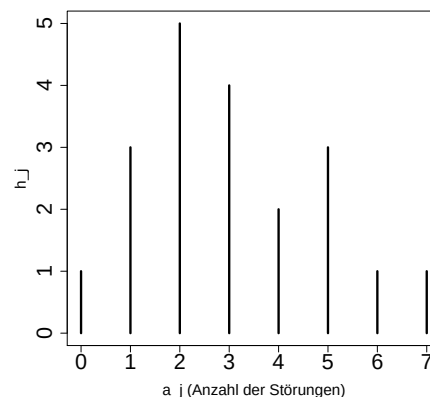


Ausprägungen a_j	absolute Häufigkeiten h_j	Summe der absoluten Häufigkeiten $\sum_{i=1}^j h_j$	relative Häufigkeit w_j	Summe der relativen Häufigkeiten $s_j = \sum_{i=1}^j w_j$
0	1	1	0,05	0,05
1	3	4	0,15	0,20
2	5	9	0,25	0,45
3	4	13	0,20	0,65
4	2	15	0,10	0,75
5	3	18	0,15	0,90
6	1	19	0,05	0,95
7	1	20	0,05	1

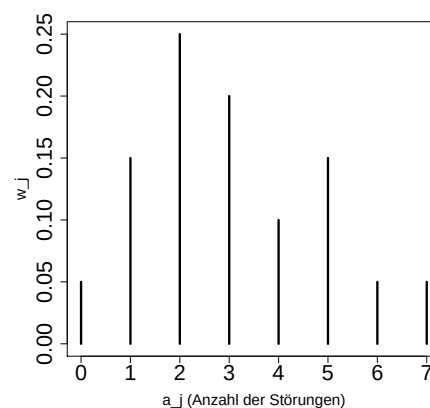
Beachte: bei metrischen oder ordinal skaliertem Merkmal ordnet man die Ausprägungen der Größe nach. Summen (Spalte 3 und 5) sind auch für metrische oder ordinal skalierte Merkmale sinnvoll.

- Graphische Darstellung mittels Stabdiagramm

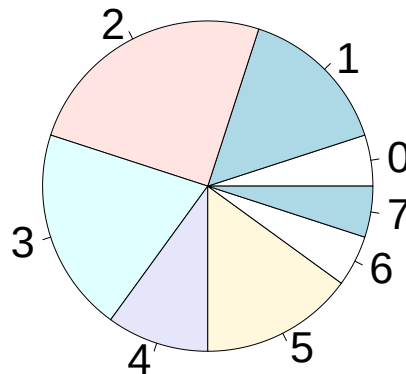
- absolute Häufigkeit h_j auf y -Achse
Ausprägung a_j auf x -Achse



- relative Häufigkeit w_j auf y -Achse
Ausprägung a_j auf x -Achse



- Graphische Darstellung mittels Kreisdiagramm



Abgetragen: Anzahl der Störungen

Winkel φ_j proportional zur relativen Häufigkeit w_j

$$\varphi_j = 360^\circ \cdot w_j$$

a_j	w_j	φ_j
0	0,05	18°
1	0,15	54°
2	0,25	90°
3	0,20	72°
4	0,10	36°
5	0,15	54°
6	0,05	18°
7	0,05	18°

- Stichprobenmittelwert

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \sum_{j=1}^k h_j a_j \\ &= \frac{1}{20} (4 + 2 + 6 + 3 + 3 + 11 + 5 + 2 + 2 + 1 + 0 + 4 + 2 + 5 + 5 + 3 + 7 + 2 + 1 + 3) \\ &= 3,05\end{aligned}$$

- Stichprobenvarianz

$$\begin{aligned}s^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \sum_{j=1}^k h_j (a_j - \bar{x})^2 \\ &= \frac{1}{19} (1 \cdot (0 - 3,05)^2 + 3 \cdot (1 - 3,05)^2 + \dots) \\ &= 3,418\end{aligned}$$

- Median

Nötig: geordnete Stichprobe

Für gerade Anzahl von Stichproben:

$$(x_{(1)} \dots x_{(20)}) = (0, 1, 1, 1, 2, 2, 2, 2, 2, \mathbf{3}, \mathbf{3}, 3, 3, 4, 4, 5, 5, 5, 6, 7)$$

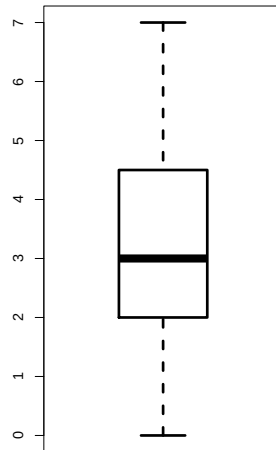


$$\tilde{x}_{0,5} = \frac{1}{2}(x_{(0,5 \cdot 20)} + x_{(0,5 \cdot 20 + 1)}) = \frac{1}{2}(3 + 3) = 3$$

Genauso werden die Quantile $\tilde{x}_q$ für beliebiges $q \in (0, 1)$ bestimmt.
Für ungerade Anzahl von Stichproben:

$$\tilde{x}_{0,5} = x_{(0,5 \cdot n \text{ [aufgerundet]})}$$

- Graphische Darstellung mittels Boxplot

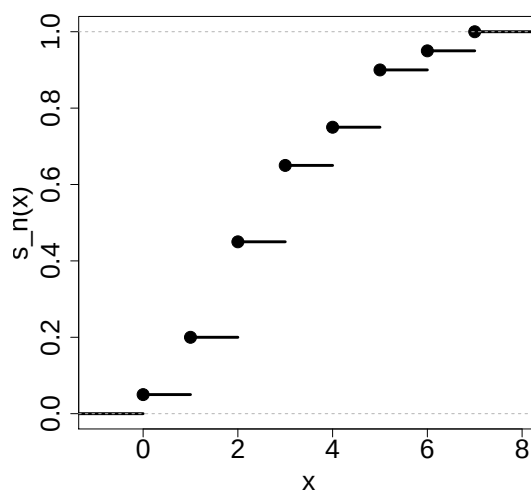


y-Achse: Anzahl der Störungen
Graphische Darstellung von

- Maximum
- 0,75-Quantil
- Median
- 0,25-Quantil
- Minimum

- empirische Verteilungsfunktion

$$s_n(x) = \begin{cases} 0 & \text{falls } x < a_1 \\ s_j & \text{falls } a_j \leq x < a_{j+1} \\ 1 & \text{falls } x \geq a_k \end{cases} \quad s_n(j_j) = s_j$$



2.1-16 Beispiel (stetige bzw. metrische Daten)

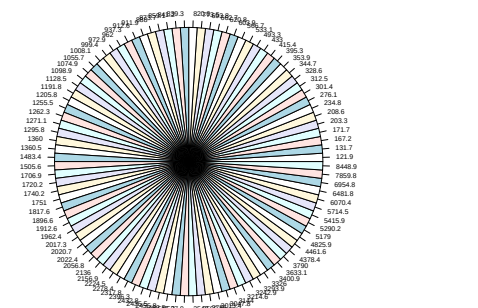
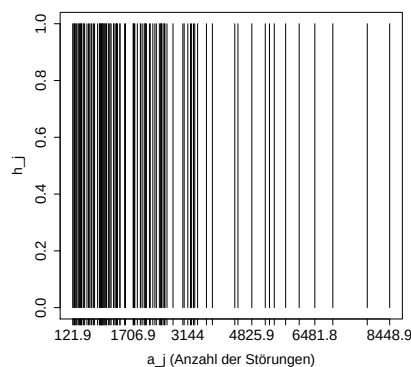
- Bei 100 elektrischen Bauteilen des gleichen Typs wurde die Lebensdauer überprüft. Die sortierten Daten lauten:

$$(x_{(1)}, \dots, x_{(100)}) = (121.9, 131.7, 167.2, 171.7, 203.3, 208.6, \dots, 6954.8, 7859.8, 8448.9)$$

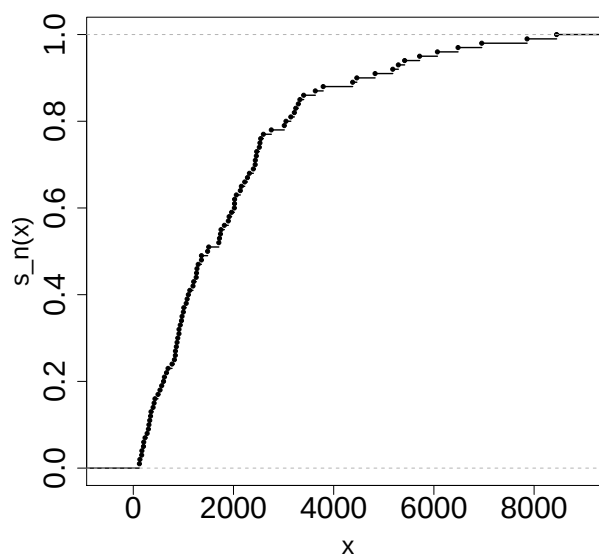
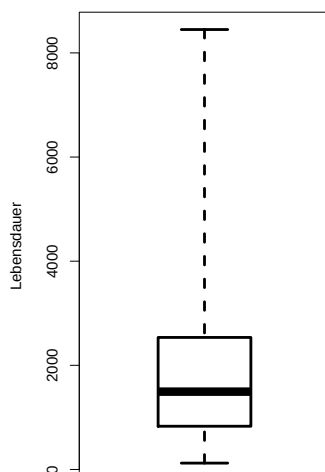
Jeder Wert kommt nur einmal vor.

- Darstellung:
 - Häufigkeitstabelle
 - Stabdiagramm
 - Kreisdiagramm

sehr ungünstig



- Besser geeignet:
 - Boxplot (übersichtlich, aber Informationsverlust)
 - Verteilungsfunktion (übersichtlich, ohne Informationsverlust)
 - Klassierung der Daten (und anschließend Histogramm, Häufigkeitstabelle, Kreisdiagramm, ...)



- KLASSIERUNG DER DATEN
Um auf für stetige Merkmale Häufigkeitstabellen und Säulendiagramme zu nutzen, fassen wir „ähnliche“ Daten in Klassen zusammen. Genauer:



Wir unterteilen das Intervall $(x_{\min}, x_{\max})$ in k Intervalle auf und betrachten dann diese gruppierten Daten in Häufigkeitstabelle und Diagrammen.

Empfehlung zur Einteilung der Klassen:

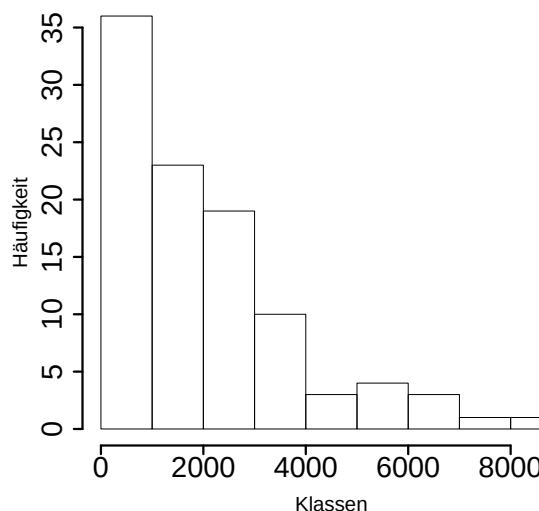
- Anzahl k der Klassen:
 - ♦ $k \approx \sqrt{n}$ falls $n \leq 400$
 - ♦ 20 falls $n > 400$
- Klassenbreite d_j muss nicht notwendigerweise für jede Klasse gleich sein. Falls doch

$$d_j = d \approx \frac{x_{\max} - x_{\min}}{k}$$

• Häufigkeitstabelle für klassierte Daten

j	Klasse K_j	absolute Klassenhäufigkeit h_j	Klassenmitte m_j	relative Klassenhäufigkeiten w_j	relative Summenhäufigkeit $s_1 + \dots + s_j$
1	[0,1000)	36	500	0,36	0,36
2	[1000,2000)	23	1500	0,23	0,59
3	[2000,3000)	19	2500	0,19	0,78
4	[3000,4000)	10	3500	0,1	0,88
5	[4000,5000)	3	4500	0,03	0,91
6	[5000,6000)	4	5500	0,04	0,95
7	[6000,7000)	3	6500	0,03	0,98
8	[7000,8000)	1	7500	0,01	0,99
9	[8000,9000)	1	8500	0,01	1

• Histogramm mit konstanter Klassenbreite

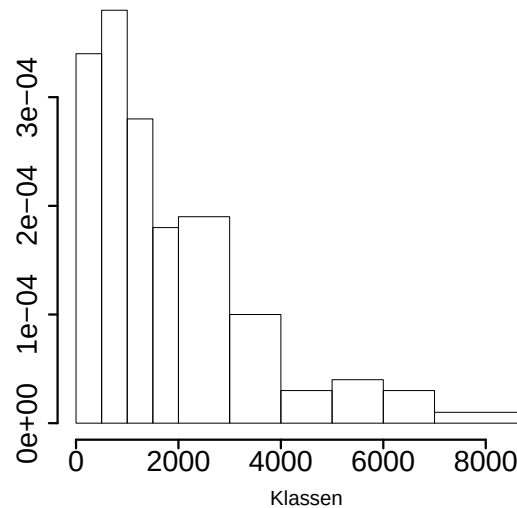


Säulenfläche proportional zur absoluten Häufigkeit h_j

Bei konstanter Klassenbreite ist auch Säulenhöhe proportional zur absoluten Häufigkeit h_j

• Histogramm mit unterschiedlicher Klassenbreite





Säulenfläche proportional zur absoluten Häufigkeit h_j
 Bei unterschiedlicher Klassenbreite ist die y -Achse so skaliert, dass gilt:

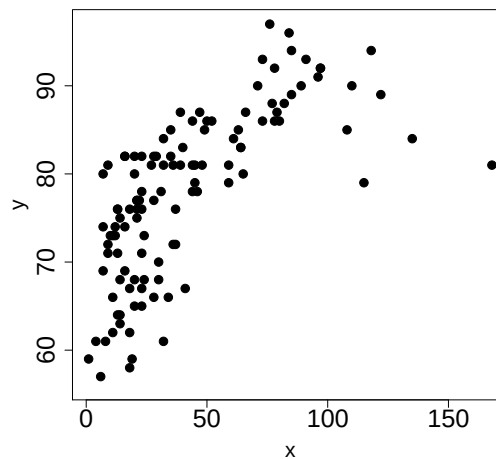
$$\sum_{j=1}^k d_j \cdot l_j = 1$$

wobei $d_j \dots$ Klassenbreite, $l_j \dots$ Säulenhöhe

2.1-17 Beispiel Uns liegt eine Stichprobe der Größe 116 von (gleichzeitigen) Messungen des Ozongehaltes X der Luft (in 1/1Mio Teilchen) und der Temperatur Y (in Grad Fahrenheit)) vor:

$$((x_1, y_1), (x_2, y_2), \dots, (x_{150}, y_{150})) = ((41, 67), (36, 72), (115, 79), \dots, (18, 62))$$

- Streudiagramm



- KLASSIERUNG UND KONTINGENZTAFEL

Wir stellen uns vor die Daten liegen uns nur in klassierter Form vor:

- Ozongehalt X nimmt Werte 1, 2, 3 an, wobei
 - 1 ... Ozongehalt < 50 ;
 - 2 ... Ozongehalt in $[50, 100)$;
 - 3 ... Ozongehalt ≥ 100
- Temperatur Y nimmt Werte 0, 1 an, wobei
 - 0 ... Temperatur < 75 ;
 - 1 ... Temperatur ≥ 75

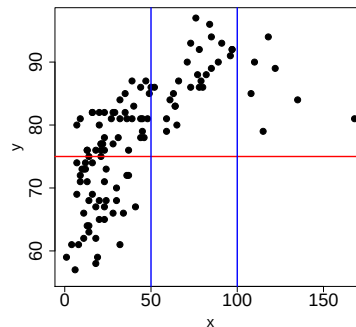


Nun sieht die konkrete Stichprobe wie folgt aus:

$$((x_1, y_1), (x_2, y_2), \dots, (x_{150}, y_{150})) = ((1, 0), (1, 0), (3, 1), \dots, (1, 0))$$

Kontingenztafel und Streudiagramm

	$Y = 0$	$Y = 1$
$X = 1$	40	41
$X = 2$	0	28
$X = 3$	0	7

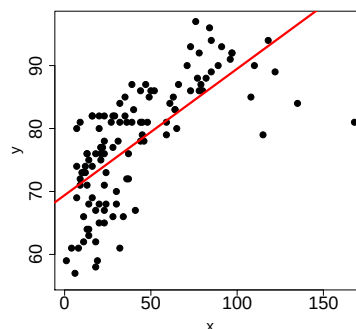


- Regressionsgerade:

$$a_1 = r_{xy} \frac{s_y}{s_x} = \frac{s_{xy}}{s_x \cdot s_y} \cdot \frac{s_y}{s_x} = \frac{s_{xy}}{s_x^2} = 0,2$$

und $a_0 = \bar{y} - a_1 \cdot \bar{x} = 77,87 - 0,2 \cdot 42,13 = 69,4$ also

$$y = 0,2x + 69,4$$



2.1.3.2 STICHPROBENFUNKTIONEN FÜR 2-DIMENSIONALE MERKMALE

2.1-18 Definition Zu den beiden metrisch skalierten Merkmalen sei die zweidimensionale Stichprobe $(X_1, Y_1), \dots, (X_n, Y_n)$ gegeben.

Wir definieren die folgenden Stichprobenfunktionen:

- Stichprobenkovarianz:

$$S_{X,Y} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

- Stichproben-Korrelationskoeffizient (nach Pearson)

$$R_{X,Y} = \frac{S_{X,Y}}{S_X \cdot S_Y} \in [-1, 1]$$



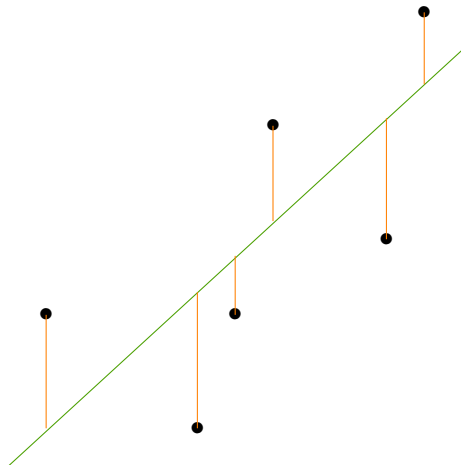
2.1-19 Bemerkung Für 2 ZVen X und Y kennen wir schon den Korrelationskoeffizienten

$$\varrho_{X,Y} = \frac{\text{Cov } X, Y}{\sigma_X \sigma_Y}.$$

Der Stichprobenkorrelationskoeffizient $R_{X,Y}$ (bzw. $r_{x,y}$ für die konkrete Stichprobe) ist ein Schätzwert für (das unbekannte) $\varrho_{X,Y}$.

Wie findet man die Gerade g , welche die Daten $(x_1, y_1), \dots, (x_n, y_n)$ „am besten“ approximiert?

- dazu sollen die Werte $(y_i - g(x_i))^2$ für alle i möglichst klein sein.



- wir minimieren dazu die Summe der quadrierten Abstände, also wir suchen g so, dass

$$\sum_{i=1}^n (y_i - g(x_i))^2$$

minimal wird. Dies führt zu folgender Definition:

2.1-20 Satz Sei eine konkrete Stichprobe $(x_1, y_1), \dots, (x_n, y_n)$ mit $s_x \neq 0$ gegeben. Die Gerade

$$g : \mathbb{R} \rightarrow \mathbb{R}, g(x) = a_0 + a_1 x$$

mit

$$a_1 = r_{xy} \frac{s_y}{s_x} \text{ und } a_0 = \bar{y} - a_1 \bar{x}$$

ist die eindeutige Lösung des Minimierungsproblems (siehe Summe aus vorhergehender Bemerkung). D.h. für eine beliebige Gerade $\tilde{g}$ gilt

$$\sum_{i=1}^n (y_i - g(x_i))^2 \leq \sum_{i=1}^n (y_i - \tilde{g}(x_i))^2.$$

g heißt REGRESSIONSGERADE.

(nur sinnvoll für metrische Merkmale)

2.1-21 Bemerkung Falls eine 2-dimensionale Stichprobe aus mindestens einem ordinalskalierten Merkmal besteht, verwendet man den Korrelationskoeffizient nach Spearman.
Vorgehen:



(1) Ordne der Größe nach:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

Z.B.:

$x_{(i)}$	-2	-2	0	1	5	6	6	6	8
Platz	1	2	3	4	5	6	7	8	9
Rang	1,5	1,5	3	4	5	7	7	7	9

(2) Rang einer Ausprägung berechnen mittels

$$R(a) = \begin{cases} k & \text{falls } a \text{ nur einmal (auf Platz } k) \text{ auftritt} \\ \frac{k_1 + k_2}{2} & \text{falls } a \text{ auf den Plätzen } k_1 \text{ bis } k_2 \text{ auftritt} \end{cases}$$

(3) $R(x_i)$ = Rang der Ausprägung x_i

Setze

$$\tilde{x} := (R(x_1), R(x_2), \dots, R(x_n))$$

(4) gehe genauso für y_i vor:

$$\tilde{y} = (R(y_1), \dots, R(y_n))$$

(5) Der Rangkorrelationskoeffizient nach Spearman ist nun

$$r_{x,y}^{(S)} := r_{\tilde{x}, \tilde{y}}.$$

2.2 SCHÄTZTHEORIE

2.2.1 EINFÜHRUNG

- Ziel: Anhand einer Stichprobe Rückschlüsse auf die Verteilung eines Merkmals X in einer Grundgesamtheit zu ziehen.
- Gegeben: Stichprobe $X_1, \dots, X_n$ zu Merkmal X .
- Bekannt: Der Verteilungstyp (Normal-/Poisson-/... verteilung). D.h. wir unterstellen, dass die tatsächliche Verteilung aus einer vorgegebenen ein- oder mehrparametrischen Schar von Verteilungsfunktionen stammt.
Raum der möglichen Parameter: Θ (groß Theta)
- Unbekannt: der wahre Parameter $\vartheta \in \Theta$
- Aufgabe: ϑ anhand der Stichprobe schätzen.
- Beispiele:
 - Exponentialverteilung: Setzen voraus, dass $X \sim \text{Exp}(\lambda)$ wobei $\lambda = \vartheta$ der unbekannte Parameter ist. $\Theta = (0, \infty)$
 - Normalverteilung: Setzen voraus, dass $X \sim \mathcal{N}(\mu, \sigma^2)$, wobei μ und σ^2 unbekannt sind. Dann:

$$\Theta = \mathbb{R} \times (0, \infty)$$

Wir unterscheiden:

- PUNKTSCHÄTZER
Stichprobenfunktion $T = T(X_1, \dots, X_n)$ liefert zu jeder konkreten Stichprobe eine konkrete Schätzung $\hat{\vartheta}$ des Parameters ϑ .



- **BEREICHSSCHÄTZER**

Gesucht ist hier ein Bereich $I = I(X_1, \dots, X_n)$ mit

$$\mathbb{P}(\vartheta \in I) \geq 1 - \lambda$$

für kleines (vorgegebenes) λ .

Meist ist I ein INTERVALL.

2.2.2 PUNKTSCHÄTZER

2.2.2.1 EIGENSCHAFTEN VON PUNKTSCHÄTZERN

2.2-1 Definition Eine Stichprobenfunktion $T(X_1, \dots, X_n)$ welche zur Schätzung eines Parameters $\vartheta \in \Theta$ verwendet wird heißt **SCHÄTZFUNKTION** (oder „Punktschätzer“ oder „Schätzer“) für ϑ .

2.2-2 Bemerkung

- Eine Stichprobenfunktion ist eine Zufallsvariable.
- Um die Abhängigkeiten von $T(X_1, \dots, X_n)$ von der Stichprobengröße n zu verdeutlichen schreibt man oft auch T_n statt $T(X_1, \dots, X_n)$.
- Oft verwendet man auch um den Schätzer zu bezeichnen den gleichen wie den zu schätzenden Parameter, jedoch versehen mit einem $\hat{\cdot}$, also etwa $\hat{\vartheta}$ (statt T).

2.2-3 Beispiel

- Gesuchter Parameter: Mittlere Abfüllmenge bei einem Getränkehersteller.

Schätzfunktion: $T = \bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$

- Gesuchter Parameter: Varianz der Abfüllmenge bei einem Hersteller

Schätzfunktion: $T = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

2.2-4 Definition Ein Punktschätzer T für ϑ heißt

- (a) **ERWARTUNGSTREU** (oder unverzerrt), wenn

$$\mathbb{E}(T) = \vartheta.$$

- (b) **ASYMPTOTISCH ERWARTUNGSTREU**, wenn

$$\lim_{n \rightarrow \infty} \mathbb{E}(T_n) = \vartheta$$

- (c) **(SCHWACH) KONSISTENT**, wenn für jedes $\varepsilon > 0$ gilt:

$$\lim_{n \rightarrow \infty} \mathbb{P}(|T_n - \vartheta| < \varepsilon) = 1$$

2.2-5 Satz Ist ein Schätzer T_n

- erwartungstreu und
- erfüllt $\lim_{n \rightarrow \infty} \text{Var}(T_n) = 0$

so ist er auch (schwach) konsistent.



2.2-6 Beispiel Sei X ein Merkmal. Setze $\mu := \mathbb{E}(X)$ und $\sigma^2 = \text{Var}(X)$.

(a) $T := \bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$ ist erwartungstreuer Schätzer für $\vartheta = \mu$, denn

$$\mathbb{E}(T) = \mathbb{E}\left(\frac{1}{n}(X_1 + \dots + X_n)\right) = \frac{1}{n} \cdot (\underbrace{\mathbb{E}(X_1)}_{\mu} + \dots + \underbrace{\mathbb{E}(X_n)}_{\mu}) = \mu$$

(b) Wegen (a) und

$$\text{Var}(T) = \text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \stackrel{(\text{unabh.})}{=} \frac{1}{n^2} \sum_{i=1}^n \underbrace{\text{Var} X_i}_{\sigma^2} = \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0$$

ist $T = \bar{X}$ sogar ein konsistenter Schätzer für μ .

(c) $T = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ ist erwartungstreuer Schätzer für $\vartheta = \sigma^2$ (siehe Übung).

2.2.3 BEREICHSSCHÄTZER

2.2.3.1 PROBLEMSTELLUNG

Wozu Bereichsschätzer?

- Punktschätzer liefert Wert, der den wahren Parameter höchstwahrscheinlich nicht genau trifft.
- Frage: Wie weit liegt die Schätzung neben dem wahren Wert?
- Bereichsschätzer liefern einen (zufälligen) Bereich, welcher mit hoher WK den wahren Wert überdeckt.
- Meist (und hier immer) ist dieser Bereich ein Intervall.
Daher auch „Intervallschätzer“ oder „Konfidenzintervall“.

2.2-7 Definition Ein Intervall

$$I(X_1, \dots, X_n) = [g_u(X_1, \dots, X_n), g_o(X_1, \dots, X_n)]$$

welches zur vorgegebenen $\alpha \in (0, 1)$ die Bedingung

$$\mathbb{P}(\vartheta \in I(X_1, \dots, X_n)) \geq 1 - \alpha$$

erfüllt heißt KONFIDENZINTERVALL (oder „Vertrauensintervall“) zum Niveau $1 - \alpha$. man nennt $1 - \alpha$ das KONFIDENZNIVEAU.

2.2-8 Bemerkung

- $I(X_1, \dots, X_n)$ ist zufällig!
- $\vartheta \in \Theta$ ist fest (aber unbekannt).
- Um ein Konfidenzintervall (KI) zu bestimmen, wählt man ZUERST $\alpha \in (0, 1)$. Typische Werte: $\alpha = 0,01$, $\alpha = 0,05$.



- Die Intervallgrenzen g_u und g_o können $-\infty$ bzw. $+\infty$ annehmen. D.h. Intervalle der Art $(-\infty, b]$ oder $[a, \infty)$ sind möglich.
- Die Realisierung $I(x_1, \dots, x_n)$ der KI für eine konkrete Stichprobe heißt konkretes Konfidenzintervall.

Warum nicht einfach $\alpha = 0$ wählen?

- d.h. wir suchen I mit $\mathbb{P}(\vartheta \in I) = 1$
- dies liefert typischer Weise $I = \Theta$ und damit keinen Erkenntnisgewinn.

2.2-9 Beispiel (und allgemeine Vorgehensweise)

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$ ein Merkmal mit BEKANNTEM σ^2 .

Gesucht: KI für $\vartheta = \mu$ zum Niveau $1 - \alpha$.

- (1) Punktschätzer bestimmen: Kennen wir schon aus Beispiel 2.2-6:

$$T = \bar{X}$$

- (2) Verteilung von T bestimmen:

T ist normalverteilt (wegen Satz 1.3-31) und $\mathbb{E}(T) = \mu$, $\text{Var}(T) = \frac{\sigma^2}{n}$ (siehe Bsp. 2.2.6)

$$\text{also } T \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

- (3) Transformation von T , so dass T in bekannte Verteilung übergeht.

- (4) s_1 und s_2 so wählen, dass

$$\mathbb{P}(S \in [s_1, s_2]) \geq 1 - \alpha$$

- Wähle zuerst α_1 und α_2 mit $\alpha = \alpha_1 + \alpha_2$.
- Wähle s_1 und s_2 so, dass $\mathbb{P}(S \leq s_1) \leq \alpha$ und $\mathbb{P}(S \geq s_2) \leq \alpha_2$



Also $s_1 = z_{\alpha_1} = -z_{1-\alpha_1}$ und $s_2 = z_{1-\alpha_2}$. Dann:

$$\mathbb{P}\left(-z_{1-\alpha_1} \leq \frac{\bar{X} - \mu}{\sigma} \sqrt{n} \leq z_{1-\alpha_2}\right) \geq 1 - \alpha$$

- (5) Umstellen liefert:

$$\mathbb{P}\left(\bar{X} - \frac{z_{1-\alpha_2}\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{z_{1-\alpha_1}\sigma}{\sqrt{n}}\right) \geq 1 - \alpha$$

d.h. das gesuchte KI ist

$$I = \left[\bar{X} - \frac{z_{1-\alpha_2}\sigma}{\sqrt{n}}, \bar{X} + \frac{z_{1-\alpha_1}\sigma}{\sqrt{n}}\right]$$

Für Prüfung: KI muss abgelesen (nicht gebildet) werden.



2.2-10 Bemerkung

- Oft wird $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$ gewählt.
- Teilweise werden aber auch einseitige KIs betrachtet. Dann gilt:

$$\alpha_1 = \alpha, \alpha_2 = 0 \text{ oder } \alpha_1 = 0, \alpha_2 = \alpha$$

2.2.3.2 KONFIDENZINTERVALLE BEI NORMALVERTEILUNG

Sei X ein normalverteiltes Merkmal mit Erwartungswert μ und Varianz σ^2 . Konfidenzintervalle für μ und σ^2 sind der Tabelle zu entnehmen.

Achtung: Es ist jeweils zu beachten, ob der andere Parameter bekannt ist oder auch aus den Daten geschätzt wurde.

2.2-11 Beispiel In einem Abfüllautomaten werde Zucker in Tüten abgefüllt. Der Sollwert beträgt $\mu_0 = 1\,000$ [g]. Die tatsächliche Abfüllmenge ist jedoch normalverteilt mit EW μ und Varianz σ^2 . Aus einer SP vom Umfang $n = 50$ ergab sich der Wert $\bar{x} = 988,7$ und $s = 31,2$.

Gesucht: einseitiges KI der Art $(-\infty, a]$ für μ zum Niveau $1 - \alpha = 0,99$.

Lösung: KI für μ , wobei σ^2 unbekannt ist:

$$\begin{aligned} I(x_1, \dots, x_{50}) &= \left[\bar{x} - \frac{t_{n-1, 1-\alpha_2}}{\sqrt{n}} \cdot s, \bar{x} - \frac{t_{n-1, 1-\alpha_1}}{\sqrt{n}} \cdot s \right] \\ &\text{mit } \alpha_2 = 0, \alpha_1 = \alpha = 0,01 \\ &= \left[988,7 - \frac{t_{49,1}}{\sqrt{50}} \cdot 31,2, 988,7 - \frac{t_{49,099}}{\sqrt{50}} \cdot 31,2 \right] \\ &\text{mit } t_{49,1} = \infty \quad t_{49,099} = 2,405 \\ &= (-\infty, 999,31] \end{aligned}$$

Beachte: $999,31 < 1\,000$

2.2-12 Bemerkung Ob ein einseitiges oder zweiseitiges KI gefragt ist, hängt von der konkreten Anwendung ab.

- Aus Sicht des Abnehmers der Ware:
 $I = (-\infty, 999,31] \Rightarrow$ Ablehnung der Ware, da ein solches KI den wahren Mittelwert mit WK 0,99 enthält. Der wahre Mittelwert liegt also mit 99% WK unter den gewünschten 1 000, nämlich wahrscheinlich höchstens bei 999,31.
- Aus Sicht des Herstellers:
Ansatz: $I_2 = [a, \infty)$ liefert $I_2 = [978,09, \infty)$. Da $1\,000 \in I_2$, hat er keine Einwände.
- Aus Sicht eines unabhängigen Kontrolleurs:
Zweiseitiges KI ($\alpha_1 = \alpha_2 = 0,005$) ergibt $I = [976,87, 1\,000,53]$. Da $1\,000 \in I_3$ hat er keine Einwände.

2.2.3.3 KONFIDENZINTERVALL FÜR EINE UNBEKANNTE WAHRSCHEINLICHKEIT

Problem: Betrachte zufälliges Ereignis A mit $\mathbb{P}(A) = p$.

Zur Schätzung von p betrachten wir n unabhängige Versuchswiederholungen. $\rightsquigarrow$ Modellierung mit n unabhängigen ZV $X_1, \dots, X_n$, welche alle Bernoulli-verteilt sind mit p . Erinnerung: $\mathbb{E}(X_i) = p$ und $T = \bar{X}$ ist konsistenter Schätzer für p .

Gesucht ist KI für $\vartheta = p$.

2 Möglichkeiten:



(1) ASYMPTOTISCHES KI (beruht auf ZGWS) liefert Intervall, welches nahezu ein KI ist.
Faustregel zur Anwendung: $n\bar{x}(1 - \bar{x}) \geq 9$

(2) EXAKTES KI liefert tatsächliches KI

(siehe Tabelle)

2.2-13 Beispiel Bei der Produktion von USB-Sticks wird eine SP der Größe $n = 200$ entnommen. 12 dieser USB-Sticks sind unbrauchbar.

Gesucht: KI zum Niveau $1 - \alpha = 0,95$ für Ausschussteil p .

Lösung: $n = 200$, $\bar{x} = \frac{12}{200} = 0,06$, $\alpha = 0,05$.

Faustregel: $n\bar{x}(1 - \bar{x}) = 11,28 > 9 \Rightarrow \text{ok.}$

$Z_{1-\frac{\alpha}{2}} = Z_{0,975} = 1,96$.

Einsetzen:

- asymptotisches KI

$$I_a = [0,0346, 0,1020]$$

- exaktes KI

$$F_1 = F_{24, 378, 0,025} = 0,5103$$

$$F_2 = F_{26, 376, 0,975} = 1,6509$$

$$I_e = [0,0314, 0,1025]$$

2.2.4 BOOTSTRAPPING

Bootstrapping liefert eine Methode um Güte von Schätzern zu bewerten (auch wenn keine klassische Formel für KI vorhanden ist).

Frage: wie bewerten wir die Güte von Schätzern, wenn

- keine Annahme über zugrunde liegende Verteilung gemacht werden kann/soll.
- wir einen Wert schätzen, für den es (in der Literatur) keinen Ansatz für ein KI gibt.

Grundidee:

- Es liegt nur EINE Stichprobe vor (Größe n)
- Satz von Glivenko-Cantelli besagt, dass für großes n die theoretische Verteilung (reale Verteilung) durch die empirische Verteilung (Verteilung der Stichprobe) approximiert werden kann.
- Entsprechend der empirischen Verteilung generieren wir nun zufällig k neue Stichproben (der Größe n), sogenannte Bootstrap-Stichproben. Dies entspricht n -fachem Ziehen mit Zurücklegen (aus original SP) $\rightsquigarrow$ liefert k unabhängige SP der Größe n .
Anhand dieser lassen sich die Varianz des Schätzers, sowie Konfidenzintervalle schätzen.

2.2-14 Definition Sei $(X_1, \dots, X_n)$ eine SP zum Merkmal X und $(x_1, \dots, x_n)$ eine konkrete SP. Sei $T(X)$ eine reellwertige SP-Funktion. Das folgende Vorgehen generiert die Bootstrap-Schätzung für die Varianz von $T(X)$, sowie ein $(1 - \alpha)$ -Bootstrap-Konfidenzintervall für $T(X)$.

- (1) Ziehe n -mal aus $x_1, \dots, x_n$ mit Zurücklegen und nenne diese Bootstrap-SP $x_1^{(b)}$
- (2) Wiederhole (1) noch $(k - 1)$ mal und erhalte so die Bootstrap-SPen $x_2^{(B)}, \dots, x_k^{(B)}$



(3) Wende die SP-Funktion T auf die Bootstrap-SPen an:

$$t_1 = T(x_1^{(B)}), \dots, t_n = T(x_k^{(B)})$$

(4) Die Bootstrap-Schätzung für die Varianz von $T(X)$ ist nun:

$$s_B^2 = \frac{1}{k-1} \sum_{i=1}^k (t_i - \bar{t})^2 \quad \bar{t} = \frac{1}{k} \sum_{i=1}^k t_i$$

(5) Die Quantile $\tilde{t}_{\frac{\alpha}{2}}, \tilde{t}_{1-\frac{\alpha}{2}}$ des Vektors $t_1, \dots, t_k$ liefern Schätzer für das 2-seitige KI.:

$$I_B = [\tilde{t}_{\frac{\alpha}{2}}, \tilde{t}_{1-\frac{\alpha}{2}}]$$

2.2-15 Beispiel Die Qualität eines Laser Entfernungsmesser soll überprüft werden. Dafür wurde 10 mal die gleiche Entfernung in mm gemessen:

$$x = (1432, 1431, 1429, 1430, 1425, 1431, 1432, 1436, 1442, 1422).$$

Es soll ein Bootstrap Konfidenzintervall zum Niveau $1 - \alpha = 0,8$ für die Standardabweichung angegeben werden. Außerdem ist eine Schätzung für die Varianz des Schätzers für die Standardabweichung zu bestimmen.

Lösung:

- Wir verwenden $k = 15$ Bootstrap-Wiederholungen (in der Praxis sollte k deutlich größer sein). 15 mal n -maliges Ziehen mit Zurücklegen ergibt:
- für alle 15 Bootstrap-Stichproben berechnen wir die Stichprobenstandardabweichung:

$$t_1 = 6,85, t_2 = 5,83, \dots, t_{15} = 4,65$$

- Für alle 15 Bootstrap-Stichproben die Stichprobenstandardabweichung, geordnet:

$$2,71, 2,96, 3,23, 4,65, 4,98, 5,03, 5,06, 5,83, 6,06, 6,46, 6,70, 6,85, 6,85$$

- Quantile: $\alpha = 0,2, \frac{\alpha}{2} = 0,1$
 $0,1 \cdot 15 = 1,5$ und $0,9 \cdot 15 = 13,5$ sind keine Zahlen, daher wird 1,5 und 13,5 aufgerundet und es gilt

$$\tilde{t}_{0,1} = 2,96 \text{ und } \tilde{t}_{0,9} = 6,85$$

- Konfidenzintervall:

$$I = [2,96, 6,85]$$

- Varianz:

$$\begin{aligned} s_B^2 &= \frac{1}{14} \sum_{i=1}^{15} 5(t_i - \tilde{t})^2 \\ &= ((2,71 - 5,14)^2 + \dots + (6,85 - 5,14)^2) / 14 = 1,87 \end{aligned}$$



2.2-16 Bemerkung

- Wie groß ist k zu wählen?
→ Möglichst groß! Mithilfe von PCs ist $k = 1\,000$ und mehr meist kein Problem.
- Vorteile:
 - keine Verteilungsannahmen nötig
 - auf beliebige Stichprobenfunktionen anwendbar
 - einfach
- Nachteil:
 - Ungenauigkeit, da 2 Approximationsschritte vorgenommen werden:
 - (1) theoretische Verteilung
 $\rightsquigarrow$ empirische Verteilung
 - (2) zufällige Stichproben aus empirischen Verteilung
 - nicht deterministisch (liefert bei jedem Durchführen neue Werte)
- Bootstrapping ist eine Methode aus dem Bereich Resampling.

2.3 TESTTHEORIE

2.3.1 GRUNDIDEE

- Gegeben: Stichprobe $X_1, \dots, X_n$ zu Merkmal X .
- Aufgabe: Annahmen (Hypothese) über die (unbekannte) Verteilung von X überprüfen.

Zwei Fälle:

- (a) Verteilung ist bis auf einen Parameter ϑ bekannt. Hypothese betrifft nur den Parameter ϑ (bspw. $\vartheta = \vartheta_0$ wobei ϑ_0 der Sollwert ist).
- (b) Verteilungstyp unbekannt (nicht parametrische Tests).

Idee:

- Definiere Nullhypothese (H_0 , der für mich unproblematische Normalfall) und Alternativhypothese (H_1 , problematischer Fall)
- Ein Test ist eine Entscheidungsregel, die anhand der Stichprobe zu „Ablehnen“ oder „Nicht-Ablehnen“ der Nullhypothese führt.
- Eine solche Entscheidung kann natürlich auch falsch sein („Nullhypothese ablehnen, obwohl sie stimmt“ oder „Nullhypothese nicht ablehnen, obwohl sie nicht stimmt“).
- Wollen WK für „ H_0 ablehnen, obwohl H_0 stimmt“ (peinlicher Irrtum) durch vorgegebenes $\alpha \in (0, 1)$ beschränken.

2.3-1 Definition Ein (statistischer) Test ist eine Abbildung φ , welche eine Stichprobe $\mathbf{X} = (X_1, \dots, X_n)$ nach $\{0, 1\}$ abbildet. Seien zusätzlich $\alpha \in (0, 1)$ und $\{P_\vartheta \mid \vartheta \in \Theta\}$ die Menge aller möglichen Verteilungen der Stichprobe und $\Theta_0 \subseteq \Theta$, sowie $\Theta_1 = \frac{\Theta}{\Theta_0}$ gegeben. Dann heißt φ Test der Nullhypothese $\vartheta \in \Theta_0$ zum Signifikanzniveau α , wenn

$$\mathbb{P}_\vartheta(\underbrace{\varphi = 1}_{H_0 \text{ ablehnen}}) \leq \alpha \text{ für alle } \underbrace{\vartheta \in \Theta_0}_{H_0 \text{ stimmt aber}}.$$



2.3-2 Bemerkung Interpretation des Testergebnis:

$\varphi(\mathbf{X}) = 0$... Anhand der Stichprobe $\mathbf{X} = (x_1, \dots, x_n)$ lässt sich der Verdacht, dass die Alternativhypothese richtig ist nicht rechtfertigen.

$\varphi(\mathbf{X}) = 1$... Die Stichprobe $\mathbf{X}$ spricht gegen die Nullhypothese. Wir verwerfen sie daher und nehmen an, dass die Alternativhypothese gilt.

2.3-3 Bemerkung (konkrete Vorgehensweise)

Gegeben:

- Merkmal X mit (unbekannter) Verteilung $\mathbb{P}_{\vartheta^*}$.
- Stichprobe $\mathbf{X} = (X_1, \dots, X_n)$,
konkrete SP $\mathbf{x} = (x_1, \dots, x_n)$.
- $\{\mathbb{P}_{\vartheta} \mid \vartheta \in \Theta\}$... Menge aller in Frage kommenden Verteilungen.

(0) Beschreibung der Zufallsvariablen, wie sie verteilt ist und was bekannt ist.

(1) Wahl des Signifikanzniveaus $\alpha \in (0, 1)$

(2) Aufstellen einer (Null-)hypothese

Wähle $\Theta_0 \subseteq \Theta$ dann

$$H_0 : \vartheta^* \in \Theta_0$$

$$H_1 : \vartheta^* \in \Theta_1 = \Theta \setminus \Theta_0$$

(bspw. X ist normalverteilt mit unbekanntem Erwartungswert μ , wir vermuten $\mu = \mu_0$, befürchten aber $\mu \neq \mu_0$. Dann $\Theta = \mathbb{R}$, $\Theta_0 = \{\mu_0\}$ und $\Theta_1 = \mathbb{R} \setminus \{\mu_0\}$. $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$)

(3) Konstruktion (und Berechnung) einer Testgröße

$$T = T(\mathbf{X})$$

mit bekannter Verteilung, falls H_0 richtig ist.

T soll Unterschiede zwischen der hypothetischen Verteilung $\mathbb{P}_{\vartheta}$, $\vartheta \in \Theta_0$ und der tatsächlichen Verteilung widerspiegeln.

Für eine konkrete SP $\mathbf{x}$ wird der Testwert $t = T(\mathbf{x})$ berechnet.

(4) Konstruktion eines kritischen Bereichs K mit der Eigenschaft

$$\mathbb{P}_{\vartheta}(T \in K) \leq \alpha \text{ für alle } \vartheta \in \Theta_0$$

(nach Möglichkeit mit $=$ statt $\leq$).

(5) Entscheidungsregel:

- Fall $t \in K$: Ablehnen der Nullhypothese H_0 (Test ist signifikant; $\varphi(\mathbf{x}) = 1$)
- Fall $t \notin K$: Auf Basis des durchgeführten Test ist nichts gegen die Nullhypothese einzuwenden (Test ist nicht signifikant; $\varphi(\mathbf{x}) = 0$).

2.3-4 Bemerkung Konstellation für Hypothesen (bei Parametertests)

- $H_0 : \vartheta = \vartheta_0, \quad H_1 : \vartheta \neq \vartheta_0$
- $H_0 : \vartheta \leq \vartheta_0, \quad H_1 : \vartheta > \vartheta_0$
- $H_0 : \vartheta \geq \vartheta_0, \quad H_1 : \vartheta < \vartheta_0$



2.3-5 Bemerkung (Fehler) Bei Tests kann es zu verschiedenen Fehlern kommen:

- (1) Fehler 1. Art: Entscheidung für H_1 , obwohl H_0 wahr ist (α -Fehler, type-I-error)

$$\varphi(\mathbf{x}) = 1 \text{ obwohl } H_0 \text{ gilt.}$$

- (2) Fehler 2. Art: H_0 wird nicht abgelehnt, obwohl H_1 wahr ist (β -Fehler, type-II-error)

$$\varphi(\mathbf{x}) = 0 \text{ obwohl } H_1 \text{ gilt.}$$

	H_0 gilt	H_1 gilt
Test verwirft H_0 , $\varphi(\mathbf{x}) = 1$	Fehler 1. Art	korrekt
Test verwirft H_0 nicht, $\varphi(\mathbf{x}) = 0$	korrekt	Fehler 2. Art

2.3-6 Bemerkung Sei φ ein Test zur Nullhypothese $\vartheta \in \Theta_0$ mit $\alpha \in (0, 1)$.

- Laut Konstruktion ist WK für Fehler 1. Art durch α beschränkt:

$$\mathbb{P}_{\vartheta}(\varphi = 1) \leq \alpha \text{ für } \vartheta \in \Theta_0$$

- β ist eine Schranke für den Fehler 2. Art, falls

$$\mathbb{P}_{\vartheta}(\varphi = 0) \leq \beta \text{ für alle } \vartheta \in \Theta_1$$

(Test verwirft H_0 nicht, obwohl H_1 richtig ist)

Wählt man die kleinstmögliche Schranke β , so heißt $1 - \beta$ die Macht (oder Power) des Tests.

2.3.2 PARAMETERTESTS

Geben Übersicht zu wichtigsten Test.

Zum Durchführen wichtig:

- Welche Nullhypothese wird getestet?
- Testvoraussetzungen?
- Testgröße T ?
- Verteilung von T unter H_0 ?
- Kritischer Bereich?

2.3.2.1 TESTS FÜR μ UND σ UNTER NORMALVERTEILUNG

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$ ein normalverteiltes Merkmal und $X_1, \dots, X_n$ zugehörige Stichprobe. Tests werden entsprechend der Tabelle 1 durchgeführt.

2.3-7 Beispiel Bei der Herstellung von Zylindern kann der Durchmesser als normalverteilt angenommen werden. Die Standardabweichung σ des Durchmessers kann als Maß für die Qualität der Produkte angesehen werden. Der Hersteller gibt an, dass die Standardabweichung σ höchstens $0,03 \text{ [mm]}$ beträgt. Der Käufer zweifelt dies an, entnimmt eine Stichprobe vom Umfang $n = 40$ um die Aussage des Herstellers zu widerlegen. Aus der Stichprobe ergeben sich die Werte $\bar{x} = 50,03 \text{ [mm]}$ und $s = 0,037 \text{ [mm]}$.

Führen Sie einen Test aus Sicht des Käufers zum Signifikanzniveau $\sigma = 0,05$ durch.

Lösung: Verwende χ^2 -Steuertest, wobei $\vartheta = \sigma^2$ und μ unbekannt ist.

Setze $\sigma_0 = 0,03$.



(1) Signifikanzniveau: $\alpha = 0,05$

(2) Hypothesen: $H_0 : \sigma^2 \leq 0,03^2$ $H_1 : \sigma^2 > 0,03^2$
(MERKE: Was man zeigen will, wird als Alternativhypothese formuliert. Hier: der Käufer will zeigen, dass die Abweichung höher als angegeben ist.)

(3) Testgröße:

$$t = \frac{(n-1)S^2}{\sigma_0^2 \sim \chi^2(n-1)}$$

$$\text{Konkreter Testwert: } t = T(x) = \frac{(n-1)s^2}{\sigma_0^2} = \frac{39 \cdot 0,037^2}{0,03^2} = 59,32$$

(4) Kritischer Bereich: $K = (\chi_{n-1,1-\alpha}^2, \infty) = (\chi_{39,0,95}^2, \infty) = (54,57, \infty)$
(aus Quantiltabelle für die χ^2 -Verteilung ablesen)

(5) Entscheidung: $t \in K \Rightarrow H_0$ wird abgelehnt ($\varphi(xi) = 1$) (6) Interpretation: Die Stichprobe bestätigt die Vermutung des Käufers. Die Nullhypothese, dass „die Standardabweichung maximal 0.03 [mm] beträgt“, wird verworfen.

2.3-8 Bemerkung

- Die Entscheidung „Ablehnen von H_0 “ oder nicht hängt von α ab. Hätten wir im Beispiel $\alpha = 0,01$ gewählt, so erhielten wir $t \notin K = (62,43, \infty)$, d.h. gegen H_0 wäre nichts einzuwenden gewesen.
- α ist immer VOR der Durchführung des Tests zu wählen.
Derjenige α -Wert, für den die Grenze des kritischen Bereichs genau mit dem konkreten Wert t der Testgröße übereinstimmt (also die Grenzstelle zwischen Ablehnung und Nicht-Ablehnung) heißt P-WERT.
- Es gilt:
 $p < \alpha \Rightarrow$ Ablehnen von H_0
 $p \geq \alpha \Rightarrow$ gegen H_0 ist nichts einzuwenden.
Beachte: p-Wert ist die typische Ausgabe von Statistik-Software beim Durchführen von Tests.
- p-Wert ist die WK, dass die Testgröße T den konkreten Wert t oder einen extremeren annimmt.
Liegt das t in dem kritischen Bereich K (dessen Flächeninhalt α ist), wird H_0 abgelehnt, der p-Wert ist die Fläche ab t (und ist kleiner α , wenn H_0 abgelehnt wird).
Liegt das t nicht im kritischen Bereich K , dann ist der p-Wert die Fläche ab t (und $p \geq \alpha \Leftrightarrow t \notin K$).
Bei der beidseitigen Betrachtung wird für p-Wert die Fläche von $\pm t$ bis $\pm \infty$ betrachtet. Dabei gilt wieder: $p < \alpha \Leftrightarrow H_0$ wird abgelehnt und anders herum.

2.3.2.2 TESTS FÜR 2 UNABHÄNGIGE STICHPROBEN UNTER NORMALVERTEILUNG

- Betrachte 2 normalverteilte Merkmale mit je einer Stichprobe.
- Die SPen können von unterschiedlicher Größe sein.
- Es sollen Hypothesen zum Vergleich der Erwartungswerte bzw. der Varianzen untersucht bzw. getestet werden.
- Beispiele:



- Sind Männer im Mittel größer als Frauen?
- Ist das Einkommen in den neuen und den alten Bundesländern im Mittel gleich?
- Ist die Varianz der Körpergrößen bei Männern und bei Frauen die gleiche?
- Vorgehen: Entsprechend der Schritte (1)-(5) aus Bemerkung 2.3-3, wobei Testgröße, kritischer Bereich, ... der Tabelle 2 zu nehmen sind.

2.3-9 Beispiel Klaus behauptet sein Ruhepuls ist niedriger als der von Peter. Nachdem Klaus 10 Tage und Peter 15 Tage mit einem Messgerät den Ruhepuls gemessen haben, versucht Klaus mit einem Test (zum Sig.niveau $\alpha = 0,05$) seine Behauptung zu bestätigen. Welches Ergebnis zeigt sich, wenn folgendes gemessen/berechnet wurde:

$$\bar{x}_K = 55, \quad \bar{x}_P = 57$$

$$s_K = 3, \quad s_P = 4$$

(Annahme: Merkmale normalverteilt und $\sigma_P = \sigma_K$)

(1) $\alpha = 0,05$

(2) $H_0 : \mu_K \geq \mu_P \quad H_1 : \mu_K < \mu_P$ (weil Klaus zeigen will, dass sein Ruhepuls niedriger ist)

$$(3) \quad T = \frac{(\bar{X} - \bar{Y}) \sqrt{\frac{n_1 n_2}{n_1 + n_2}}}{\sqrt{\dots}}$$

$$t = \frac{(55 - 57) \cdot \sqrt{\frac{10 \cdot 15}{10 + 15}}}{\sqrt{\frac{9 \cdot 3^2 + 14 \cdot 4^2}{23}}} = -1,345$$

(4) $K = (-\infty, -t_{n_1+n_2-2, 1-\alpha}) = (-\infty, -1,7139)$

(5) Entscheidung: $t \notin K \Rightarrow H_0$ wird nicht verworfen.

(6) Interpretation: Auf Grundlage der vorliegenden Daten kann (bei einem Signifikanzniveau von $\alpha = 0,05$) nicht gezeigt werden, dass „der Ruhepuls von Klaus im Mittel niedriger ist als der von Peter“.

Beachte: Hätte Klaus weniger Angst vor einem falsch-positiven Ergebnis gehabt, hätte er mit $\alpha = 0,1$ gerechnet. Dann wäre $t \in K$ heraus gekommen.

2.3.2.3 TESTS FÜR UNBEKANNTE WAHRSCHEINLICHKEITEN

(A) eine Stichprobe

Gegeben:

- A zufälliges Ereignis mit $\mathbb{P}(A) = p \in [0, 1]$
- X mit Parameter p Bernoulli-verteiltes Merkmal mit SP $X_1, \dots, X_n$
- $\sum_{i=1}^n X_i = n\bar{X} \dots$ absolute Häufigkeit von A bei n unabhängigen Versuchen
- $\bar{X} \dots$ relative Häufigkeit

Aufgabe: Nullhypothese $p = p_0$, $p \leq p_0$ bzw $p \geq p_0$ für vorgegebenes $p_0 \in [0, 1]$ überprüfen.

Vorgehen: entsprechend Bemerkung 2.3-3 und Tabelle 3

(B) zwei SPen

Gegeben:

- A, B zufällige Ereignisse mit $\mathbb{P}(A) = p_1, \mathbb{P}(B) = p_2$



- $X \sim \text{Bin}(p_1)$, $Y \sim \text{Ber}(p_2)$ Merkmale mit SPen $X_1, \dots, X_n$ und $Y_1, \dots, Y_n$

Aufgabe: Nullhypothese $p_1 = p_2$, $p_1 \leq p_2$ bzw. $p_1 \geq p_2$ überprüfen

Vorgehen: entsprechend Bemerkung 2.3-3 und Tabelle 4

2.3-10 Beispiel Bei der Herstellung von Zahnprothesen wird der Ausschussanteil untersucht. Es ist zu prüfen ob sich zeigen lässt, dass das neue Herstellungsverfahren (B) gegenüber dem alten Verfahren (A) eine Verbesserung ist.

Verfahren	SP-Umfang	Ausschussanzahl
A	400	29
B	500	25

Es ist ein Test zum Signifikanzniveau $\alpha = 0,05$ durchzuführen.

(1) $\alpha = 0,05$

(2) $H_0 = p_A \leq p_B$ $H_1 = p_A > p_B$
 $p_A \dots$ Ausschussverfahren von Verteilung A
 $p_B \dots$ Ausschussverfahren von Verteilung B

(3) $T = \frac{\bar{X} - \bar{Y}}{\sqrt{\hat{p}(1-\hat{p})\frac{n_1+n_2}{n_1 \cdot n_2}}} \sim \mathcal{N}(0,1)$ mit $\hat{p} = \frac{n_1\bar{X} + n_2\bar{Y}}{n_1 + n_2}$

konkreter Testwert: $t = \frac{\bar{x} - \bar{y}}{\sqrt{\hat{p}(1-\hat{p})\frac{n_1+n_2}{n_1 \cdot n_2}}} = \frac{\frac{29}{400} - \frac{25}{500}}{\sqrt{0,06(1-0,06)\frac{900}{400 \cdot 500}}} = 1,412$

$\hat{p} = \frac{29 + 25}{900} = 0,06$

Prüfen der Faustregel: $n\bar{x}(1-\bar{x}) = 26,9 > 9$ $n\bar{y}(1-\bar{y}) = 23,8 > 9 \Rightarrow \text{ok}$

(4) Kritischer Bereich

$K = (z_{1-\alpha}, \infty) = (z_{0,95}, \infty) = (1,645, \infty)$

(5) $t \notin K \Rightarrow H_0$ wird nicht verworfen.

(6) Interpretation: Auf Grundlage der vorliegenden SP lässt sich die (Alternativ-)hypothese, dass „das neue Herstellungsverfahren gegenüber dem alten Verfahren einen geringeren Ausschussanteil hat“ nicht nachweisen.

2.3.3 PARAMETERFREIE TESTS

Vorgehensweise ist fast wie bei parametrischen Tests, mit dem Unterschied, dass hier nicht nur die Parameterwerte in die Hypothesen eingehen. Wir lernen hier „nur“ den χ^2 -Unabhängigkeitstest kennen.

2.3.3.1 χ^2 -UNABHÄNGIGKEITSTEST

Aufgabenstellung: Zu überprüfen ist, ob zwei diskrete Merkmale X und Y unabhängig sind. Dafür liegt eine 2-dimensionale SP vor.

$$(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$$

Kontingenztafel:

X habe Ausprägungen $1, \dots, l$



Y habe Ausprägungen $1, \dots, m$

$X \setminus Y$	1	2	...	m	$\sum$
1	H_{11}	H_{12}	...	H_{1m}	H_{1*}
2	H_{21}	H_{22}	...	H_{2m}	H_{2*}
$\vdots$	$\vdots$			$\vdots$	$\vdots$
l	H_{l1}	H_{l2}	...	H_{lm}	H_{l*}
$\sum$	H_{*1}	H_{*2}	...	H_{*m}	n

- H_{jk} ... Häufigkeit des gleichzeitigen Auftreten von $X = j$ und $Y = k$

- $H_{j*} = \sum_{k=1}^m H_{jk}$... Randhäufigkeit (Zeilensumme)

- $H_{*k} = \sum_{j=1}^l H_{jk}$... Randhäufigkeit (Spaltensumme)

Gehe nun in die üblichen Schritte (1)-(5) bzw. (6) und nach Tabelle 6 vor.

2.3-11 Bemerkung

- Falls die Merkmal X bzw. Y stetig sind, so lässt sich ein χ^2 -Unabhängigkeitstest rechnen, wenn man die Daten in Klassen einteilt.
- Faustregel für Anwendbarkeit:

$$\frac{H_{j*} \cdot H_{*k}}{n} \geq 5 \text{ für alle } j, k$$

- Die Teststatistik $T = \sum \sum \dots$ wird groß, falls die beobachteten Häufigkeiten H_{jk} stark von den erwarteten Häufigkeiten $\frac{1}{n} \cdot H_{j*} \cdot H_{*k}$ abweichen.

Die Größe $\frac{H_{j*} H_{*k}}{n}$ beschreibt die erwartete Häufigkeit unter der Annahme von Unabhängigkeit bei gegebenen Randhäufigkeiten H_{j*} und H_{*k} .

- Im Spezialfall $m = l = 2$ gibt es eine vereinfachte Formel für T , siehe Tabelle.

2.3-12 Beispiel (mit $m = l = 2$) Von 90 Schülern (47 Mädchen, 43 Jungs) der 4. Klasse einer Grundschule gehen 34 auf das Gymnasium. Von diesen 34 Schülern sind 20 Mädchen. Lässt sich mit einem Test zum Signifikanzniveau 0,05 zeigen, dass die Wahl der Bildungseinrichtung vom Geschlecht abhängig ist?

(1) $\alpha = 0,05$

- (2) H_0 : X und Y unabhängig H_1 = X und Y nicht unabhängig wobei
 X ... Wahl der Bildungseinrichtung (1 =Gymnasium, 2 =kein Gymnasium)
 Y ... Geschlecht (1 ... männlich, 2 ... weiblich)

(3) $T = n \cdot \frac{(H_{11}H_{22} - H_{12}H_{21})^2}{H_{1*}H_{2*}H_{*1}H_{*2}}$
 Kontingenztafel:



$X \setminus Y$	1	2	Σ
1	14	20	34
2	29	27	56
Σ	43	47	90

$$t = 90 \frac{(14 \cdot 27 - 20 \cdot 29)^2}{34 \cdot 56 \cdot 43 \cdot 47} = 0,945$$

(4) Kritischer Bereich

$$K = (\chi_{1,1-0,05}, \infty) = (3,84, \infty)$$

(5) $t \notin K \Rightarrow H_0$ wird nicht verworfen

(6) Interpretation: Auf Grundlage der SP lässt sich die Unabhängigkeit von Bildungseinrichtungswahl und Geschlecht nicht widerlegen.

2.3-13 Beispiel 550 Personen wurden nach Rauchgewohnheiten und sportlichen Betätigungen befragt. Von den 381 Nichtrauchern treiben 56 regelmäßig, 143 gelegentlich und 182 nie Sport. Unter den Rauchern betragen entsprechende Häufigkeiten 15, 46 und 108.

Kann man (mit Signifikanzniveau 0,01) zeigen, dass das Rauchverhalten und sportliche Betätigung als abhängig angesehen werden können?

Lösung:

$X \dots$ Rauchgewohnheit (1 = Nichtraucher, 2 = Raucher)

$Y \dots$ Sport (1 = regelmäßig, 2 = gelegentlich, 3 = nie)

$X \setminus Y$	1	2	3	Σ
1	56	143	182	381
2	15	46	108	169
Σ	71	189	290	550

(1) $\alpha = 0,01$

(2) $H_0 : X$ und Y unabhängig $H_1 = X$ und Y nicht unabhängig

$$(3) T = \sum_{j=1}^2 \sum_{k=1}^3 \left(H_{jk} - \frac{H_{j*} H_{*k}}{n} \right)^2 \cdot \frac{n}{H_{j*} H_{*k}}$$

Trage $\frac{h_{j*} h_{*k}}{n}$ in Tabelle ab:

$X \setminus Y$	1	2	3	Σ
1	49,2	130,9	200,9	381
2	21,8	58,1	108	89,1
Σ	71	189	290	550

$$\text{z.B. } \frac{381 \cdot 71}{550} = 49,2$$

$$t = \frac{(56 - 49,2)^2}{49,2} + \frac{(143 - 130,9)^2}{130,9} + \dots + \frac{(108 - 89,1)^2}{89,1} = 12,45$$

(4) $K = (\chi_{1,2,0,99}, \infty) = (9,81, \infty)$

(5) $t \in K \Rightarrow H_0$ ablehnen

(6) Interpretation: Auf Grundlage der SP lässt sich die Unabhängigkeit von Rauchgewohnheiten und sportlicher Betätigung widerlegen (sie sind also abhängig).

