



Vorlesung - Wahrscheinlichkeitsrechnung und Statistik 2024/2025

Wahrscheinlichkeitsrechnung

Wahrscheinlichkeit im Alltag ...

Laut den meteorologischen **Vorhersagen** wird es morgen regnen.

Ob ich **riskiere** und die **Wette** verlieren werde? Ich werde mit **Sicherheit** gewinnen!

Ist das wirklich **unmöglich**? Ist dies tatsächlich **möglich**???

Welch ein **Zufall**!

Wahrscheinlichkeitsrechnung ist ein Teilgebiet der Mathematik, das sich mit dem Studium der zufälligen Phänomene beschäftigt.

- *zufällig* = unvorhersehbar, unbeabsichtigt
- *aleatorius* (lat.) = zufällig
- *alea* (lat.) = Spielwürfel; Würfelspiel



→ man misst die **Chancen für Erfolg** oder das **Risiko für Misserfolg** von Ereignissen

Anwendungsgebiete:

- Glückspiele, Wetten, Loto
- meteorologische Vorhersagen
- Meinungsumfragen
- Versicherungsmathematik (Risikomessungen im Versicherungswesen / im Bankensystem)
- Kryptographie (Verschlüsselungsverfahren)
- Verarbeitung von Informationen, Komprimierbarkeit von Daten
- Generieren von Zufallszahlen (Pseudo-Zufallszahlen mithilfe von Software), zufällige Algorithmen, Simulationen, Computerspiele, maschinelles Lernen, Data-Mining, Formen- und Spracherkennung

▷ <https://www.random.org/randomness/>

man kann auch "echte Zufallszahlen" (*true random numbers*) erhalten; z.B. durch den radioaktiven Zerfall eines Atoms (nach den Gesetzen der Quantentheorie ist es nicht möglich, genau zu wissen, wann und wie dieser radioaktive Zerfall voranschreiten wird), durch elektromagnetisches Rauschen aus der Atmosphäre, durch Umgebungsgeräusche usw.

Beispiel: Generieren zufälliger Werte (in Python)

```
# Beispiel 1
import random
r = random.random()
print("Zufällige Zahl aus dem Intervall (0,1):",r)
N = random.randint(-1,5)
print("Zufällige ganze Zahl aus dem Intervall [-1,5]:",N)
Liste = ["AB", "XY", "EF", "MN", "FG"]
print("Liste:", Liste)
print("Zufällige Wahl aus Liste: ", random.choice(Liste))
k=6
print(k, "-mal zufällige Wahl aus Liste: ", random.choices(Liste, k=6))
#6-mal Ziehen mit Zurucklegen
print(4, "-mal zufällige Wahl aus Liste: ", random.sample(Liste,4))
#4-mal Ziehen ohne Zurucklegen
```

```
# Beispiel 2
import numpy as np
N=30
R = np.random.randint(1,7,size=N)
# Vektor mit N zufälligen Werten aus der Menge {1,2,...,6}
print("Werte beim",N, "-maligen Würfeln:\n",R)
total= sum(R==6)
print("Wie oft wurde 6 beim",N, "-maligen Würfeln erhalten:",total)
```

Zufällige (randomisierte) Algorithmen

Ein Algorithmus, der im Laufe seiner Ausführung gewisse Entscheidungen *zufällig* trifft, heisst zufälliger (randomisierter) Algorithmus.

- ▷ Laufzeit und/oder Korrektheit können nur noch mit gewissen Wahrscheinlichkeiten garantiert werden;
- ▷ Laufzeit, Speicherplatz, berechnetes Ergebnis (für festgehaltene Eingabe) sind Zufallsvariablen
- ▷ es gibt Beispiele wo randomisierte Algorithmen effizienter in Bezug auf Speicherplatz und Laufzeit sind

Beispiel: Bei der randomisierten Variante von Quick Sort (*Random QuickSort*) wird das Pivotelement zufällig gewählt.

Las Vegas Algorithmen: sind randomisierte Algorithmen, die immer ein korrektes Ergebnis liefern; in Abhängigkeit von den zufälligen Entscheidungen variiert die Laufzeit dieser Algorithmen. Man analysiert dann die Verteilung der Anzahl der durchgeführten Rechnungsschritte.

Beispiel: *Random QuickSort*

Monte-Carlo-Algorithmen: sind randomisierte Algorithmen, die manchmal auch ein falsches Ergebnis liefern. Man untersucht die Wahrscheinlichkeit, mit der das Ergebnis falsch ist. Durch Wiederholen des Algorithmus mit unabhängigen Zufallsbits kann jedoch die Fehlerwahrscheinlichkeit gesenkt werden. Die Rechenzeit ist konstant.

Beispiel: *Miller-Rabin-Primzahltest*: man bestimmt ob eine natürliche Zahl prim ist oder nicht. Die Ausgabe des Tests lautet entweder “sicher zusammengesetzt” oder “wahrscheinlich prim”. Der Miller-Rabin-Test liefert keine Aussage über die Faktoren einer zusammengesetzten Zahl.

Beispiel: Von welchem Typ ist folgender zufälliger Algorithmus?

Input: Sei S ein Vektor von 30 Werten aus der Menge $\{0, 1, 2\}$ (ihre Anordnung ist unbekannt; man nimmt an, dass der Vektor mindestens eine 0 enthält). Von welchem Typ ist folgender Algorithmus? (in Python geschrieben)

```
import numpy as np

N=30
S = np.random.randint(0,3,size=N) #generieren von Daten
# N= Länge des Vektors S
print(S)
#Vektor mit N zufälligen Elementen aus der Menge {0,1,2}
k=1
i= np.random.randint(N)
while S[i] != 0:
    print(k, "-te Iteration")
    print("S[" , i, "]= ", S[i])
    i= np.random.randint(low=0, high=N)
    k=k+1
# k = Anzahl Iterationen bis zufällig eine 0 gefunden wird
if S[i]==0:
    print(k, "-te Iteration")
    print("S[" , i, "]= ", S[i])
    print("In der", k, "-ten Iteration wurde zufällig eine 0 gefunden.")
```

▷ dieser Algorithmus ist 100% erfolgreich \implies es ist ein Las-Vegas Algorithmus

► Die Monte Carlo Version der vorigen Aufgabe: sei M die maximale Anzahl von Iterationen:

```
import numpy as np
print("Zweite Version")
N=50
S = np.random.randint(3,size=N)
print(S)
#Vektor mit N zufälligen Elementen aus der Menge {0,1,2}
M=3 # maximale Anzahl Iterationen, M>1
a=True
for k in range(M) :
    print(k+1, "-te Iteration")
    i= np.random.randint(low=0, high=N)
    print("S[" , i, "]= ", S[i])
    if S[i] == 0:
        print("In der", k+1, "-ten Iteration wurde zufällig eine 0 gefunden.")
        a=False
        break
if a:
    print("In", k+1, "Iterationen wurde keine 0 gefunden.")
```

▷ falls eine 0 gefunden wird, ended der Algorithmus mit dem richtigen Resultat, sonst endet der Algorithmus ohne eine 0 zu finden.

Zufällige Experimente (Versuche) und Ereignisse

Das **zufällige Experiment** ist ein Experiment dessen Ergebnis nicht vorhesehbar ist.

↔ Das **zufällige Ereignis** ist das **Ergebnis eines Experiments**.

- Würfeln mit 2 Spielwürfeln
→ beide Spielwürfel zeigen 1 an
- Wurf einer Münze
→ die Münze zeigt Zahl an
- Ziehen einer Spielkarte
→ eine 3 wurde gezogen
- Lottoziehung (6 aus 49)
→ die erste gezogene Zahl ist 23



Grundbegriffe

- das **unmögliche Ereignis**, mit \emptyset bezeichnet, taucht in der Durchführung des Experiments nie auf
 - das **sichere Ereignis** ist das Ereignis das bei jeder Durchführung des Experiments auftaucht
 - **Grundraum (Ergebnismenge)**, mit Ω bezeichnet, ist die Menge aller möglichen Ergebnisse eines Experiments
◊ der Grundraum kann endlich oder unendlich sein
 - wenn A eine Teilmenge von Ω ist, $A \subseteq \Omega$, dann wird A **zufälliges Ereignis** genannt, die Elemente von Ω heissen **elementare Ereignisse**
- *Analogie zwischen Ereignissen (Wahrscheinlichkeitsrechnung) und Mengen (Algebra)!*

Experiment: Man wirft einen Spielwürfel

Grundraum: $\Omega = \{e_1, e_2, e_3, e_4, e_5, e_6\}$

e_i : die Zahl i ($i = 1, \dots, 6$) wurde erhalten

$e_1, e_2, e_3, e_4, e_5, e_6$ elementare Ereignisse

A : eine gerade Zahl wurde erhalten $\Rightarrow A = \{e_2, e_4, e_6\}$

\bar{A} : eine ungerade Zahl wurde erhalten $\Rightarrow \bar{A} = \{e_1, e_3, e_5\}$

Operationen mit Ereignissen

- Die **Vereinigung** der Ereignisse A und B : $A \cup B = \{e \in \Omega : e \in A \text{ oder } e \in B\}$.
- Der **Durchschnitt** der Ereignisse A und B : $A \cap B = \{e \in \Omega : e \in A \text{ und } e \in B\}$.
- \bar{A} ist das **Komplement** des Ereignisses A , d.h. \bar{A} enthält alle Elementarereignisse, die nicht in A sind.
- $A, B \subseteq \Omega$ sind **disjunkte Ereignisse**, wenn $A \cap B = \emptyset$.
- Die **Differenz** der zufälligen Ereignisse A und B : $A \setminus B = A \cap \bar{B}$.
- Es gilt $A \cup \bar{A} = \Omega$, $A \cap \bar{A} = \emptyset$, $\bar{\bar{A}} = A$.

Beziehungen zwischen Ereignissen

- Aus dem Ereignis A **folgt** das Ereignis B , wenn jedes Element aus A auch Element aus B ist, d.h. $A \subseteq B$.
- Wenn aus A B folgt und aus B folgt A , dann sind die beiden Ereignisse A und B **gleich**: $A = B$.

Eigenschaften: Seien $A, B, C \subseteq \Omega$.

Die Vereinigung und Durchschnitt sind **kommutativ**

$$A \cup B = B \cup A, \quad A \cap B = B \cap A,$$

assoziativ

$$(A \cup B) \cup C = A \cup (B \cup C), \quad (A \cap B) \cap C = A \cap (B \cap C),$$

und **distributiv**

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C), \quad (A \cap B) \cup C = (A \cup C) \cap (B \cup C);$$

sie erfüllen die **Gesetze von De Morgan**

$$\overline{A \cup B} = \bar{A} \cap \bar{B}, \quad \overline{A \cap B} = \bar{A} \cup \bar{B}.$$

Relative und absolute Häufigkeit

Def. 1. Sei A ein zufälliges Ereignis das in einem Experiment auftaucht; man wiederholt das Experiment n mal (unter denselben gegebenen Bedingungen) und man bezeichnet mit k_n wie oft das Ereignis A auftaucht; die **relative Häufigkeit** des Ereignisses A ist die Zahl

$$h_n(A) = \frac{k_n}{n}.$$

k_n ist die **absolute Häufigkeit** des Ereignisses A .

Klassische Definition der Wahrscheinlichkeit

Def. 2. Wir betrachten ein Experiment welches endlich viele, gleichwahrscheinliche Ergebnisse hat. Die Wahrscheinlichkeit, dass das Ereignis A eintreitet, ist

$$P(A) = \frac{\text{Anzahl der günstigen Fälle für das Eintreten von } A}{\text{Anzahl aller möglichen Fälle innerhalb des Experiments}}.$$

➡ Nach wiederholtem Durchführen des Experiments (n hinreichend gross), unter denselben Bedingungen, ist die relative Häufigkeit $h_n(A)$ des Ereignisses A ungefähr gleich mit der Wahrscheinlichkeit $P(A)$, d.h.

$$h_n(A) \approx P(A), \text{ wenn } n \rightarrow \infty.$$

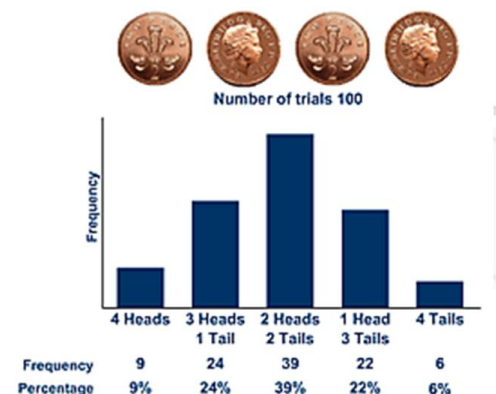
Beispiel: Experiment: Man wirft 4 Münzen.

Das Experiment wird $n = 100$ mal wiederholt. Die Ergebnisse des Experiments sind in dem Bild eingetragen.

Ereignis A : "die 4 Münzen zeigen 3 mal Zahl an."

$$f_{100}(A) = ?, \quad P(A) = ?$$

$$f_{100}(A) = \frac{22}{100} = 0.22 \text{ (aus dem Bild)}$$



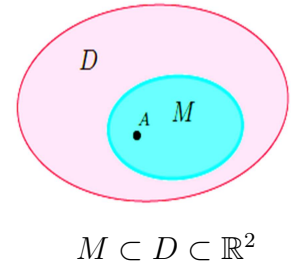
$$\Omega = \{(K, K, K, K), (K, Z, Z, Z), \dots, (Z, Z, Z, K), (Z, Z, Z, Z)\}$$

$$A = \{(K, Z, Z, Z), (Z, K, Z, Z), (Z, Z, K, Z), (Z, Z, Z, K)\} \Rightarrow P(A) = \frac{4}{2^4} = 0.25$$

Geometrische Wahrscheinlichkeit: Das Maß einer Menge entspricht der **Länge** in \mathbb{R} , der **Fläche** in \mathbb{R}^2 , dem **Volumen** in \mathbb{R}^3 . Seien $M \subset D \subset \mathbb{R}^n$, $n \in \{1, 2, 3\}$, Mengen mit endlichem Maß.

Man wählt zufällig einen Punkt $A \in D$ (in diesem Fall ist der Grundraum D). Die geometrische Wahrscheinlichkeit des Ereignisses " $A \in M$ " ist

$$P(A \in M) := \frac{\text{Maß}(M)}{\text{Maß}(D)}.$$



Axiomatische Definition der Wahrscheinlichkeit

1933 hat der russische Mathematiker **Andrei Nikolaevici Kolmogorov** im Buch "Grundbegriffe der Wahrscheinlichkeitsrechnung" die axiomatische Definition der Wahrscheinlichkeit eingeführt.

Die Wahrscheinlichkeit ist eine Funktion $P : \mathcal{K} \rightarrow \mathbb{R}$ so dass: jedem zufälligen Ereignis $A \in \mathcal{K}$ der Wert $P(A)$ zugeordnet wird

- $\hookrightarrow \mathcal{K}$ hat die Struktur einer σ -Algebra (siehe Def. 3)
- $\hookrightarrow P$ erfüllt bestimmte Axiome (siehe Def. 4)

Def. 3. Eine Familie \mathcal{K} von Ereignissen aus der Grundmenge Ω wird **σ -Algebra** genannt, wenn:

- (1) \mathcal{K} ist nicht leer;
- (2) wenn $A \in \mathcal{K}$, dann $\bar{A} \in \mathcal{K}$;
- (3) wenn $A_n \in \mathcal{K}$, $n \in \mathbb{N}$, dann $\bigcup_{n=1}^{\infty} A_n \in \mathcal{K}$.

Beispiele:

- 1) $\emptyset \neq A \subset \Omega \Rightarrow \mathcal{K} = \{\emptyset, A, \bar{A}, \Omega\}$ ist eine σ -Algebra.
- 2) **Die Menge aller Teilmengen aus Ω ist eine σ -Algebra** und wird mit $\mathcal{P}(\Omega)$ bezeichnet. Wenn Ω endlich ist, wie viele Elemente hat $\mathcal{P}(\Omega)$?
- 3) Wenn \mathcal{K} eine σ -Algebra auf Ω ist und $\emptyset \neq B \subseteq \Omega$ (ein Ereignis), dann ist $B \cap \mathcal{K} = \{B \cap A : A \in \mathcal{K}\}$ eine σ -Algebra auf B .

Satz 1. Wenn \mathcal{K} eine σ -Algebra auf Ω ist, so gelten folgende Eigenschaften:

- (1) $\emptyset, \Omega \in \mathcal{K}$; (2) $A, B \in \mathcal{K} \Rightarrow A \cap B, A \setminus B \in \mathcal{K}$; (3) $A_n \in \mathcal{K}, n \in \mathbb{N} \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{K}$.

Def. 4. Sei \mathcal{K} eine σ -Algebra in Ω . Eine Funktion $P : \mathcal{K} \rightarrow \mathbb{R}$ wird **Wahrscheinlichkeitsmaß** genannt, wenn folgende Axiome gelten:

- (1) $P(\Omega) = 1$
- (2) $P(A) \geq 0$ für alle $A \in \mathcal{K}$;
- (3) jede Folge $(A_n)_{n \in \mathbb{N}}$ von paarweise disjunkten Ereignissen (d.h. $A_i \cap A_j = \emptyset$ für alle $i \neq j$) aus \mathcal{K} gilt

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Das Tripel (Ω, \mathcal{K}, P) heisst **Wahrscheinlichkeitsraum**.

Beispiele: 1) Das einfachste Beispiel eines Wahrscheinlichkeitsmaßes erhält man im Fall eines *endlichen Grundraumes* Ω : Sei $\mathcal{K} = \mathcal{P}(\Omega)$ (die Menge aller Teilmengen von Ω) und $P : \mathcal{K} \rightarrow \mathbb{R}$ definiert durch

$$P(A) = \frac{\#A}{\#\Omega}, \text{ wobei } \#A \text{ die Anzahl der Elemente in } A \in \mathcal{P}(\Omega) \text{ ist.}$$

P , auch Zählmaß genannt, erfüllt Def. 4 und entspricht der *klassischen Definition der Wahrscheinlichkeit eines Ereignisses* (siehe Def. 2).

2) Seien $\Omega = \mathbb{N} = \{0, 1, 2, \dots\}$, $\mathcal{K} = \mathcal{P}(\mathbb{N})$ und $P : \mathcal{K} \rightarrow \mathbb{R}$ definiert durch $P(\{n\}) = \frac{1}{2^{n+1}}$, $n \in \mathbb{N}$. Es gilt $P(\mathbb{N}) = \sum_{n=0}^{\infty} \frac{1}{2^{n+1}} = 1$ und die Axiome aus Def. 4 sind erfüllt. $(\mathbb{N}, \mathcal{P}(\mathbb{N}), P)$ ist ein Wahrscheinlichkeitsraum; insbesondere Def. 4-(3) ist erfüllt, anhand des Umordnungssatzes (aus der Analysis) für Reihen mit nichtnegativen Gliedern, in diesem Fall, angewendet für eine geometrische Reihe. ♣

Satz 2. Sei (Ω, \mathcal{K}, P) ein Wahrscheinlichkeitsraum. Es gilt:

- (1) $P(\bar{A}) = 1 - P(A)$ und $0 \leq P(A) \leq 1$
- (2) $P(\emptyset) = 0$
- (3) $P(A \setminus B) = P(A) - P(A \cap B)$
- (4) $A \subseteq B \implies P(A) \leq P(B)$, d.h. P ist monoton.
- (5) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Beispiel: Aus einer Packung mit 52 Spielkarten entnimmt man zufällig eine Spielkarte. Man berechne die Wahrscheinlichkeit p , dass a) ein Ass *oder* eine Herzdame; b) eine Karte mit Herz *oder* ein Ass entnommen wurde.

L.: a) A : ein Ass wurde entnommen; D : eine Herzdame wurde entnommen; A und D sind *disjunkte* Ereignisse

$$p = P(A \cup D) = P(A) + P(D) = \frac{4+1}{52};$$

b) H : eine Karte mit Herz wurde entnommen; man bemerke, dass H und A nicht disjunkt sind

$$p = P(H \cup A) = P(H) + P(A) - P(H \cap A) = \frac{13+4-1}{52} = \frac{4}{13}.$$



Übung: Man beweise für $A, B, C \in \mathcal{K}$ die Gleichheit:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

Unabhängige Ereignisse

Def. 5. Sei (Ω, \mathcal{K}, P) ein Wahrscheinlichkeitsraum. Die Ereignisse $A, B \in \mathcal{K}$ sind **unabhängig**, wenn

$$P(A \cap B) = P(A)P(B).$$

Bemerkung: Die Ereignisse $A, B \in \mathcal{K}$ sind **unabhängig**, wenn das Auftauchen des Ereignisses A , das Auftauchen des Ereignisses B nicht beeinflusst und umgekehrt.

Übung: Man würfelt zweimal. A: die erste Zahl ist 6; B: die zweite Zahl ist 5; C: die erste Zahl ist 1.

- Sind folgende Ereignisse unabhängig oder abhängig? a) A und B; b) A und C; c) B und C.
- Sind folgende Ereignisse disjunkt? a) A und B; b) A und C; c) B und C.



Historisches Beispiel - Würfelspiel (XVII. Jh.): Der Provinzadelige Chevalier de Méré war ein leidenschaftlicher Spieler. Gerne verführte er am Pariser Hof seine Mitspieler zu folgendem Würfelspiel:

(A): *Wir werfen einen Würfel viermal. Wenn eine oder mehrere Sechsen dabei sind, gewinne ich. Wenn keine Sechsen dabei ist, gewinnen Sie.*

Tatsächlich gewann der Chevalier mit diesem Spiel regelmäßig Geld. Er dachte sich eine neue Variante aus, die ebenso lukrativ sein sollte:

(B): *Wir werfen ein Paar von Würfeln 24 mal. Wenn dabei eine Doppel-Sechsen oder mehrere sind, gewinne ich. Wenn keine Doppel-Sechsen dabei ist, gewinnen Sie.*

Würden Sie die Wette annehmen, d.h. sind Ihre Gewinnchancen größer als 50%?

(C): Wenn man statt 24 Würfeln 25 mal würfelt, ändert sich die Schlussfolgerung von (B)?

Man schätze anhand Simulationen die Wahrscheinlichkeiten folgender Ereignisse:

A: man erhält mindestens eine 6 in 4 Würfeln mit einem Würfel;

B: man erhält mindestens ein Paar (6,6) in 24 Würfeln mit zwei Würfeln;

C: man erhält mindestens ein Paar (6,6) in 25 Würfeln mit zwei Würfeln.

Berechnung der theoretischen Wahrscheinlichkeiten: \bar{A} bezeichnet das Ereignis, dass keine 6 eingetreten ist in 4 Würfeln von einem Würfel.

$$\Rightarrow P(\bar{A}) = \left(\frac{5}{6}\right)^4 \Rightarrow P(A) = 1 - \left(\frac{5}{6}\right)^4 \approx 0.5177.$$

\bar{B} bezeichnet das Ereignis, dass kein Paar (6,6) in 24 Würfeln von zwei Würfeln erhalten wurde

$$\Rightarrow P(\bar{B}) = \left(\frac{35}{36}\right)^{24} \Rightarrow P(B) = 1 - \left(\frac{35}{36}\right)^{24} \approx 0.4914$$

und analog $P(C) = 1 - \left(\frac{35}{36}\right)^{25} \approx 0.5055$. Wir vergleichen die erhaltenen Wahrscheinlichkeiten

$$P(B) < \frac{1}{2} < P(C) < P(A).$$

Schlußfolgerung: A hat die größten Gewinnchancen!



Def. 6. Die Ereignisse A_1, \dots, A_n aus \mathcal{K} sind **unabhängig (in der Gesamtheit)**, wenn für jede endliche Menge $\{i_1, \dots, i_m\} \subseteq \{1, \dots, n\}$ ($m \geq 2$) gilt

$$P(A_{i_1} \cap \dots \cap A_{i_m}) = P(A_{i_1}) \dots P(A_{i_m}).$$

A_1, \dots, A_n aus \mathcal{K} sind **paarweise unabhängig**, wenn

$$P(A_i \cap A_j) = P(A_i)P(A_j) \quad \forall i, j \in \{1, \dots, n\}, i \neq j.$$

Fragen: (1) Was bedeutet:

1a) dass 3 Ereignisse A, B, C paarweise unabhängig sind?

1b) dass die 3 Ereignisse A, B, C unabhängig (in der Gesamtheit) sind?

(2) Damit n Ereignisse B_1, \dots, B_n unabhängig (in der Gesamtheit) sind, wie viele Bedingungen müssen verifiziert werden?

Beispiel: Von den vier Flächen eines fairen Tetraeders sei eine rot, eine blau, eine grün gefärbt. Die vierte Fläche ist mit allen drei Farben bunt bemalt. Das Tetraeder wird geworfen, und es werden die folgenden Ereignisse betrachtet:

R : der Teraeder fällt auf eine Fläche, welche rote Farbe enthält

B : der Teraeder fällt auf eine Fläche, welche blaue Farbe enthält

G : der Teraeder fällt auf eine Fläche, welche grüne Farbe enthält.

Sind die 3 Ereignisse R, B, G unabhängig in der Gesamtheit?

Satz 3. Sei (Ω, \mathcal{K}, P) ein Wahrscheinlichkeitsraum und seien $A, B \in \mathcal{K}$. Folgende Aussagen sind äquivalent:

(1) A und B sind unabhängig.

(2) \bar{A} und B sind unabhängig.

(3) A und \bar{B} sind unabhängig.

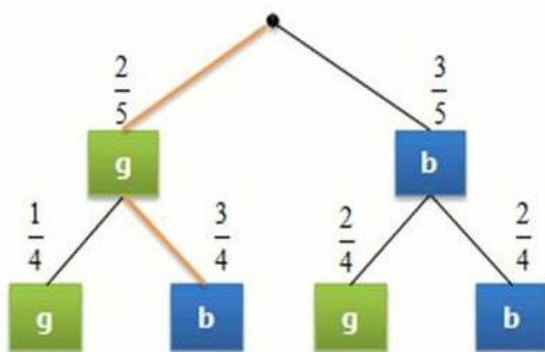
(4) \bar{A} und \bar{B} sind unabhängig.

Bedingte Wahrscheinlichkeit

Beispiel: In einer Urne sind 2 grüne und 3 blaue Kugeln. 2 Kugeln werden ohne Zurücklegen gezogen. Welches ist die Wahrscheinlichkeit, dass :

a) man eine grüne und danach eine blaue Kugel zieht?

b) beide Kugeln die gleiche Farbe haben?



1. Pfadregel: 1. Zug grün - 2. Zug blau

$$P(\text{grün, blau}) = \frac{2}{5} \cdot \frac{3}{4} = \frac{6}{20}$$

2. Pfadregel: (grün, grün) oder (blau, blau)

$$\begin{aligned} P(gg, bb) &= P(gg) + P(bb) = \frac{2}{5} \cdot \frac{1}{4} + \frac{3}{5} \cdot \frac{2}{4} \\ &= \frac{2}{20} + \frac{6}{20} = \frac{8}{20} = \frac{2}{5} \end{aligned}$$

Def. 7. Sei (Ω, \mathcal{K}, P) ein Wahrscheinlichkeitsraum und seien $A, B \in \mathcal{K}$. Die **bedingte Wahrscheinlichkeit** von A durch B ist $P(\cdot|B) : \mathcal{K} \rightarrow \mathbb{R}$ definiert durch

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

wenn $P(B) > 0$. $P(A|B)$ ist die Wahrscheinlichkeit des Eintretens des Ereignisses A unter der Bedingung, dass das Eintreten des Ereignisses B bereits bekannt ist.

Bemerkung: Die Ereignisse $A, B \in \mathcal{K}$ sind **unabhängig**, wenn das Auftauchen des Ereignisses A , das Auftauchen des Ereignisses B nicht beeinflusst und umgekehrt, d.h.

$$P(A|B) = P(A) \text{ und } P(B|A) = P(B).$$

Beispiel:

Peter hat eine Alarmanlage in sein Auto installiert, es werden folgende Ereignisse betrachtet:

A: "Alarmanlage springt an"

J: "Jemand versucht, Peters Auto aufzubrechen"

$P(\bar{A}|J)$: die (bedingte) Wahrscheinlichkeit, dass die Alarmanlage nicht anspringt unter der Bedingung, jemand versucht das Auto aufzubrechen

(diese Wahrscheinlichkeit sollte niedrig sein!)

$P(A|\bar{J})$: die (bedingte) Wahrscheinlichkeit, dass die Alarmanlage anspringt unter der Bedingung, niemand versucht das Auto aufzubrechen

(diese Wahrscheinlichkeit sollte ebenfalls niedrig sein!)

$P(A|J)$: die (bedingte) Wahrscheinlichkeit, dass die Alarmanlage anspringt unter der Bedingung, jemand versucht das Auto aufzubrechen

(diese Wahrscheinlichkeit sollte hoch sein!)

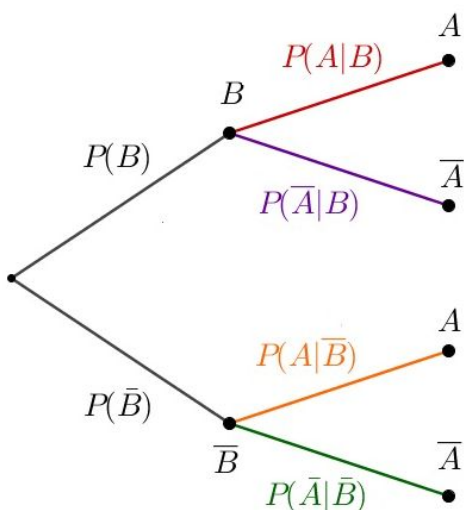
$P(J|A)$: die (bedingte) Wahrscheinlichkeit, dass jemand versucht das Auto aufzubrechen unter der Bedingung, die Alarmanlage springt an (ist nicht klug, wenn ein potentieller Dieb so handelt!) ♣

Satz 4. Für $A, B \in \mathcal{K}, P(A) > 0, P(B) > 0$ gilt:

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$$

$$P(A|B) = 1 - P(\bar{A}|B).$$

Baumdiagramm der bedingten Wahrscheinlichkeiten



$$P(A \cap B) = P(A|B)P(B)$$

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

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

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

Satz 5. (Multiplikationsregel)

Sei (Ω, \mathcal{K}, P) ein Wahrscheinlichkeitsraum und seien $A_1, \dots, A_n \in \mathcal{K}$ so dass

$$P(A_1 \cap \dots \cap A_{n-1}) > 0.$$

Es gilt

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

Zerlegung einer Grundmenge

Def. 8. Eine Menge von Ereignissen H_1, H_2, \dots, H_n aus Ω heißt **Zerlegung** (Partition, Unterteilung) von Ω , wenn

$$\bigcup_{i=1}^n H_i = \Omega$$

und für alle $i, j \in \{1, \dots, n\}$, $i \neq j$, die Ereignisse H_i und H_j disjunkt sind, d.h. $H_i \cap H_j = \emptyset$.

Beispiel: $B \subset \Omega \Rightarrow \{B, \bar{B}\}$ ist eine Zerlegung von Ω .

Satz 6. (Formel der totalen Wahrscheinlichkeit)

In einem Wahrscheinlichkeitsraum (Ω, \mathcal{K}, P) sei $\{H_1, \dots, H_n\}$ eine Zerlegung von Ω mit $H_i \in \mathcal{K}$ und $P(H_i) > 0$ für alle $i \in \{1, \dots, n\}$. Für $A \in \mathcal{K}$ gilt

$$P(A) = P(A|H_1)P(H_1) + \dots + P(A|H_n)P(H_n).$$

Beispiel: Eine Gemeinde wird zur Bürgermeisterwahl in zwei Wahlbezirke \mathcal{B}_1 und \mathcal{B}_2 eingeteilt. 60% der Wähler kommen aus \mathcal{B}_1 , 40% aus \mathcal{B}_2 . In \mathcal{B}_1 wird der Kandidat \mathcal{C} mit Wahrscheinlichkeit 0.4 gewählt, in \mathcal{B}_2 dagegen mit Wahrscheinlichkeit 0.8. Wie viel Prozent der Stimmen hat der Kandidat \mathcal{C} insgesamt bekommen?

Antwort:

B_i : Wähler ist aus Bezirk \mathcal{B}_i , $i = \overline{1, 2}$;

A : Kandidat \mathcal{C} gewinnt die Bürgermeisterwahl;

Anhand der Formel der totalen Wahrscheinlichkeit (Satz 6)

$$\Rightarrow P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) = 0.4 \cdot 0.6 + 0.8 \cdot 0.4 = 0.56.$$

Der Kandidat \mathcal{C} hat insgesamt 56% Stimmen bekommen.



Formel von Bayes

Die Bayes-Formel ist eine Methode zur Korrektur (Überarbeitung, Verbesserung) anhand neuer Daten (relevanter Informationen) einer im Voraus bestimmten Wahrscheinlichkeit (a priori; Lat. *prior* = vor). Sie beginnt mit einer Schätzung der Wahrscheinlichkeit einer bestimmten Hypothese H . Wenn wir neue Daten haben (Trainingsdaten, Beweise, Informationen, Evidenzen usw. - Engl. *evidence*) E , bezüglich der Hypothese H , kann eine korrigierte (verbesserte) Wahrscheinlichkeit für die Hypothese H berechnet werden, A-posteriori-Wahrscheinlichkeit (Lat. *posterior* = der/die/das Nachfolgende) genannt.

→ $P(H)$ die Wahrscheinlichkeit, dass die Hypothese H wahr ist, auch A-priori-Wahrscheinlichkeit genannt;

→ die bedingte Wahrscheinlichkeit $P(H|E)$ ist die, anhand der neue Daten/Informationen/Beweise, verbesserte (korrigierte) Wahrscheinlichkeit;

→ $P(E|H)$ die Wahrscheinlichkeit, dass die Daten (Informationen) auftreten, wenn die Hypothese H wahr ist;

→ $P(E|\bar{H})$ die Wahrscheinlichkeit, dass die Daten (Informationen) auftreten, wenn man weiß, dass die Hypothese H falsch ist (die Hypothese \bar{H} ist wahr).

Laut Satz 6 (mit der Partition $\{H, \bar{H}\}$) folgt

$$P(E) = P(E|H) \cdot P(H) + P(E|\bar{H}) \cdot P(\bar{H}) = P(E|H) \cdot P(H) + P(E|\bar{H}) \cdot (1 - P(H)).$$

Die Formel von Bayes ist in diesem Beispiel

$$P(H|E) = \frac{P(H \cap E)}{P(E)} = \frac{P(E|H) \cdot P(H)}{P(E)} = \frac{P(E|H) \cdot P(H)}{P(E|H) \cdot P(H) + P(E|\bar{H}) \cdot P(\bar{H})}.$$

Satz 7. (Formel von Bayes) In einem Wahrscheinlichkeitsraum (Ω, \mathcal{K}, P) sei $\{H_1, \dots, H_n\}$ eine Zerlegung von Ω mit $H_i \in \mathcal{K}$ und $P(H_i) > 0$ für alle $i \in \{1, \dots, n\}$ und sei $E \in \mathcal{K}$, so dass $P(E) > 0$. Es gilt

$$P(H_j|E) = \frac{P(E|H_j)P(H_j)}{P(E)} = \frac{P(E|H_j)P(H_j)}{P(E|H_1)P(H_1) + \dots + P(E|H_n)P(H_n)} \quad \forall j \in \{1, \dots, n\}.$$

▷ für $i \in \{1, 2, \dots, n\}$ heißt $P(H_i)$ **A-priori-Wahrscheinlichkeit** für H_i (Hypothesen)

▷ E heißt das beobachtete Ereignis (Evidenz, Engl. *evidence*);

▷ mit der Formel von Bayes berechnet man die Wahrscheinlichkeit einer Hypothese H_i unter Berücksichtigung des Ereignisses E : $P(H_i|E)$; diese werden **A-posteriori-Wahrscheinlichkeiten** genannt;

▷ $P(E|H_i)$ ist die Wahrscheinlichkeit mit der die Hypothese H_i das Auftreten des Ereignisses E voraussagt (Engl. *likelihood*)

Anwendungen der Formel von Bayes

▷ Bayes-Klassifikatoren (weisen Objekte mit bestimmten Eigenschaften der Klasse zu, zu denen sie mit der grössten Wahrscheinlichkeit gehören oder bei der durch die Einordnung die wenigsten Kosten entstehen)

↔ naiver Bayes-Lern-Algorithmus (Echtzeit-Vorhersage, Textklassifizierung, Spam-Filterung, Gefühls-Analyse in sozialen Medien)

↔ Bayes-Filter: in der Spam-Erkennung werden Bayes-Filter verwendet; von charakteristischen Wörtern in einer E-Mail wird auf die Eigenschaft, Spam zu sein, geschlossen; mit Machine Learning lässt sich der Filter verfeinern

▷ Bayessche-Netzwerke

Beispiel (bedingte Wahrscheinlichkeiten): Wie gut ist der Spam-Filter?

Statistische Daten: 2120 E-Mails wurden in eine Stichprobe einbezogen.

H : eine (zufällig gewählte) E-Mail ist *kein Spam*;

\bar{H} : eine (zufällig gewählte) E-Mail ist *Spam*.

Ein Spam-Filter soll E-Mails in *kein Spam* (Ereignis E) oder *Spam* (Ereignis \bar{E}) klassifizieren (eine Vorhersage oder Prognose angeben).

Folgende Daten wurden erhalten:

▷ RP=400 (richtig positiv) Anzahl der E-Mails die tatsächlich *kein Spam* sind und als *kein Spam* vom Spam-Filter eingeordnet wurden; $\#(H \cap E)$ ¹

▷ FP=210 (falsch positiv) Anzahl der E-Mails die tatsächlich *Spam* sind und als *kein Spam* vom Spam-Filter eingeordnet wurden; $\#(\bar{H} \cap E)$

▷ FN=310 (falsch negativ) Anzahl der E-Mails die tatsächlich *kein Spam* sind und als *Spam* vom Spam-Filter eingeordnet wurden; $\#(H \cap \bar{E})$

¹Anzahl der Elemente aus $H \cap E$

▷ $RN=1200$ (richtig negativ) Anzahl der E-Mails die tatsächlich *Spam* sind und als *Spam* vom Spam-Filter eingeordnet wurden; $\#(\bar{H} \cap \bar{E})$.

Die Konfusionsmatrix (Wahrheitsmatrix) wird verwendet, um die Leistung eines Klassifikators (z.B. des Spam-Filters) zu visualisieren.

		Realität		
		<i>kein Spam (+)</i>	<i>Spam (-)</i>	gesamt
Prognose	<i>kein Spam (+)</i>	RP	FP	RP+FP
	<i>Spam (-)</i>	FN	RN	RN+AN
	gesamt	RP+FN	FP+RN	RP+FP+FN+RN

Struktur der Konfusionsmatrix (Engl. *confusion matrix*)

		Realität		
		H : E-Mail ist <i>kein Spam (+)</i>	\bar{H} : E-Mail ist <i>Spam (-)</i>	gesamt
(Filter) Prognose	E : E-Mail ist <i>kein Spam (+)</i>	400 (richtig positiv RP)	210 (falsch positiv FP)	610
	\bar{E} : E-Mail ist <i>Spam (-)</i>	310 (falsch negativ FN)	1200 (richtig negativ RN)	1510
	gesamt	710	1410	2120

Konfusionsmatrix mit Daten aus der Stichprobe

Anhand der statistischen Daten aus der Stichprobe: a) die Wahrscheinlichkeit, dass eine E-Mail, die als *kein Spam* vom Filter klassifiziert wurde, tatsächlich *kein Spam* ist

$$P(H|E) = \frac{400}{610} \approx 0.65;$$

b) die Wahrscheinlichkeit, dass eine E-Mail, die als *Spam* vom Filter klassifiziert wurde, tatsächlich *Spam* ist

$$P(\bar{H}|\bar{E}) = \frac{1200}{1510} \approx 0.79.$$



Wann ist ein Klassifikator (ein Modell) *gut genug*?

Klassifikationsmodelle	<i>classification models in machine learning</i>
Gütekriterien der binären Klassifikation	<i>measuring the performance of a binary classification model</i>
Positiver-Vorhersagewert= $\frac{RP}{RP+FP}$	<i>positive predictive value; precision</i>
Negativer-Vorhersagewert= $\frac{RN}{RN+FN}$	<i>negative predictive value</i>
Sensitivität= $\frac{RP}{RP+FN}$ Trefferquote; Richtig-Positiv-Rate	<i>recall; probability of detection; true positive rate</i>
Spezifität= $\frac{RN}{RN+FP}$ Richtig-Negativ-Rate	<i>true negative rate</i>
Treffergenauigkeit= $\frac{RP+RN}{RP+FP+RN+FN}$ Anteil der korrekt klassifizierten Objekte	<i>accuracy</i>

Wiederholungsübung: Man zieht sukzessiv und ohne Zurücklegen zwei Karten aus einer Packung mit 52 Spielkarten;

A: die erste Karte ist ein Ass; B: die zweite Karte ist eine Dame; C: die zweite Karte ist eine 2.

1a) Sind A und B unabhängig?

1b) Man berechne $P(A \cap B)$, $P(A \cap C)$, $P(B \cap C)$.

Zufallsgrößen - zufällige Variablen

Beispiele: 1) Man wirft zwei Würfel, die Summe der erhaltenen Zahlen ist eine Zufallsgröße

$S : \Omega \rightarrow \{2, 3, \dots, 12\}$, wobei Ω alle elementaren Ereignisse enthält, dh. $\Omega = \{(\omega_i^1, \omega_j^2) : i, j = \overline{1, 6}\}$, (ω_i^1, ω_j^2) ist das elementare Ereignis der erste Würfel zeigt die Zahl i an und der zweite Würfel zeigt die Zahl j an $i, j = \overline{1, 6}$; man erhält z.B. $P(S = 5) = \frac{4}{36}$, $P(S = 6) = \frac{5}{36}$, usw.

2) Ein Spieler wirft zwei unterscheidbare Münzen

$\Rightarrow \Omega = \{(K, Z), (K, K), (Z, K), (Z, Z)\}$

Die Zufallsgröße X zeigt an wie oft Zahl (Z) aufgetaucht ist: $\Rightarrow X : \Omega \rightarrow \{0, 1, 2\}$

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

▷ Abkürzung: **Zufallsgröße** \rightarrow **ZG**

Eine Zufallsgröße ist

diskret, wenn sie endlich viele (x_1, \dots, x_n) oder abzählbar unendlich viele (x_1, \dots, x_n, \dots) Werte

stetig, wenn sie (überabzählbar viele) Werte in einem Intervall oder aus \mathbb{R}

annehmen kann.

Diskrete ZG: numerische Zufallsgrößen: Summe der Zahlen beim Würfeln, Anzahl Sechsen bei 1000 Würfelwürfen, Anzahl defekter Teile während der Produktion, Anzahl Kunden eines Geschäfts an einem Tag, Anzahl Druckfehler auf einer Zeitungsseite; kategoriale Zufallsgrößen/Merkmale (z.B. Wetterprognose: regnerisch, bewölkt, neblig, heiter \rightarrow Klassifizierung in Kategorien) ...

Stetige ZG: numerische Zufallsgrößen: Brenndauer einer Glühlampe, Abfüllmenge in Konserven, Länge von hergestellten Schrauben, Temperatur in einer Stadt innerhalb eines Jahres, Geschwindigkeit eines an einer Radarkontrolle vorbeifahrenden Autos, Laufzeit einer Maschine bis zur ersten Betriebsstörung ...

Numerische Zufallsgröße - formale Definition

Sei (Ω, \mathcal{K}, P) ein Wahrscheinlichkeitsraum.

Def. 9. $X : \Omega \rightarrow \mathbb{R}$ ist eine numerische Zufallsgröße, wenn

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{K} \text{ für alle } x \in \mathbb{R}.$$

Diskrete Zufallsgrößen: $X : \Omega \rightarrow \{x_1, x_2, \dots, x_i, \dots\}$

Die **Wahrscheinlichkeitsverteilung** von X ist:

$$X \sim \begin{pmatrix} x_1 & x_2 & \dots & x_i & \dots \\ p_1 & p_2 & \dots & p_i & \dots \end{pmatrix} = \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$$

$I \subseteq \mathbb{N}$ (Indexmenge); $p_i = P(X = x_i) > 0, i \in I$, wobei $\sum_{i \in I} p_i = 1$

Klassische diskrete Verteilungen

➡ **Diskrete Gleichverteilung:** $X \sim \text{Unid}(n), n \in \mathbb{N}^*$

$$X \sim \begin{pmatrix} 1 & 2 & \dots & n \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}$$

► Python: `scipy.stats.randint`

Beispiel: Ein Würfel wird geworfen. Sei X die ZG die anzeigt, welche Zahl aufgetaucht ist

$$\Rightarrow X \sim \begin{pmatrix} 1 & 2 & \dots & 6 \\ \frac{1}{6} & \frac{1}{6} & \dots & \frac{1}{6} \end{pmatrix}$$

```
from scipy.stats import randint #pip install scipy
import numpy
import matplotlib.pyplot as plt
from matplotlib.pyplot import bar, show, hist, grid, legend, xticks
N=1000
a=1; b=7
R = randint.rvs(a, b, size = N)
#print ("Zufällige Werte: \n", R)
x, count = numpy.unique(R, return_counts=True)
print("Die Werte",x,"haben die abs. Häufigkeiten:",count)
bar(x,count/N, width=0.8,color="cyan", edgecolor="black")
plt.grid()
plt.xlabel("Werte der Unid(6) Verteilung")
plt.ylabel("relative Häufigkeiten")
plt.title("Unid(6) ")
xticks(range(0,b))
show()
```

➡ **Bernoulli Verteilung:** $X \sim \text{Bernoulli}(p), p \in (0, 1)$

$$X \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

Beispiel: Bernoullisches Versuchsschema: Innerhalb eines Experiments taucht A (Erfolg) oder \bar{A} (Misserfolg) auf; $\mathbb{I}_A = 0 \Leftrightarrow \bar{A}$ taucht auf; $\mathbb{I}_A = 1 \Leftrightarrow A$ taucht auf
 $\Rightarrow \mathbb{I}_A \sim \text{Bernoulli}(p)$ mit $p := P(A)$

$$\mathbb{I}_A \sim \begin{pmatrix} 0 & 1 \\ 1 - P(A) & P(A) \end{pmatrix}$$

Zufällige Werte mit `scipy.stats.bernoulli` generiert:

```
from scipy.stats import bernoulli #pip install scipy
import numpy
import matplotlib.pyplot as plt
from matplotlib.pyplot import bar, show, hist, grid, legend, xticks
N=1000
Bern = bernoulli.rvs(size=N,p=0.6)
```



```

z, count = numpy.unique(Bern, return_counts=True)
print("Die Werte", z, "haben die Haufigkeiten:", count)
bar(z, count/N, width=0.9, color="red", edgecolor="black")
plt.grid()
plt.xlabel("Werte der Bernoulli Verteilung")
plt.ylabel("relative Haufigkeiten")
plt.title("Bernoulli(0.6)")
xticks(range(0, 2))
plt.show()

```

➡ **Binomialverteilung:** $X \sim \text{Bino}(n, p), n \in \mathbb{N}^*, p \in (0, 1)$

- Innerhalb eines Experiments taucht A oder \bar{A} auf (Bernoullisches Versuchsschema)
 - A Erfolg mit $P(A) = p$, \bar{A} Misserfolg mit $P(\bar{A}) = 1 - p$
 - man wiederholt das Experiment n -mal (z.B. Münzwurf, Ziehen mit Zurücklegen im Urnenmodell)
 - Zufallsgröße X ist die Anzahl der Erfolge in n Wiederholungen des Experiments
- ⇒ die möglichen Werte von X sind $0, 1, \dots, n$.

Eine Zufallsgröße X mit dem Wertebereich $\{0, \dots, n\}$ heißt **binomialverteilt** mit den Parametern $n \geq 1$ und $p \in (0, 1)$, falls gilt

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}, \quad k \in \{0, \dots, n\}.$$

▷ Zur Erinnerung: **binomische Formel**

$$(a + b)^n = \sum_{k=0}^n C_n^k a^k b^{n-k}$$

für $a = p$ und $b = 1 - p$ erhält man

$$1 = \sum_{k=0}^n C_n^k p^k (1 - p)^{n-k}$$

Beispiele:

1) Ein Würfel wird 10-mal geworfen. Sei X die ZG die anzeigt wie oft die Zahl 6 auftaucht $\Rightarrow X \sim \text{Bino}(10, \frac{1}{6})$.

2) In einer Urne mit $n_1 + n_2$ Kugeln sind n_1 weiß und n_2 schwarz.

Experiment: Man entnimmt eine Kugel, notiert ihre Farbe und legt sie in die Urne zurück.

Man wiederholt das Experiment n mal und betrachtet die Zufallsgrößen:

X_1 = Anzahl der entnommenen weißen Kugeln; X_2 = Anzahl der entnommenen schwarzen Kugeln.

$\Rightarrow X_1 \sim \text{Bino}(n, p_1)$ mit $p_1 = \frac{n_1}{n_1 + n_2}$; $X_2 \sim \text{Bino}(n, p_2)$ mit $p_2 = \frac{n_2}{n_1 + n_2}$.

➤ Zufällige Werte mit `scipy.stats.binom` generiert:

```

# Im Urnenmodell n_1=70 weisse Kugeln, n_2=30 schwarze Kugeln
# n=10 Ziehungen mit Zurücklegen
import numpy
import matplotlib.pyplot as plt
from matplotlib.pyplot import bar, show, hist, grid, legend, xticks
from scipy.stats import binom
N=2000
n=10; p= 0.7
data = binom.rvs(n, p, size= N)
#zufällige Werte der ZG die anzeigt wie viele weisse Kugeln
#in 10 Ziehungen erhalten wurden
z, count = numpy.unique(data, return_counts=True)

```



```
print("Die Werte", z, "haben die absoluten Haufigkeiten:", count)
bar(z, count/N, width=0.8, color="red", edgecolor="black")
plt.grid()
plt.xlabel("Werte der Bino(10,0.7) Verteilung")
plt.ylabel("relative Haufigkeiten")
plt.title("Bino(10,0.7) ")
xticks(range(0, n+1))
show()
```

➡ **Geometrische Verteilung:** $X \sim Geo(p)$, $p \in (0, 1)$

Innerhalb eines Experiments taucht A (Erfolg) mit $P(A) = p$ oder \bar{A} (Misserfolg) mit $P(\bar{A}) = 1 - p$ auf

X = Anzahl der Versuche **vor dem ersten Erfolg**

Mögliche Werte: $X \in \{0, 1, 2, \dots\} = \mathbb{N}$

Beispiel: X zeigt an wie oft man würfelt *bevor* die erste 6 auftaucht $\Rightarrow X \sim Geo(\frac{1}{6})$

Eine Zufallsgröße X mit dem Wertebereich $\{0, 1, 2, \dots\}$ heißt **geometrisch verteilt** mit dem Parameter $p \in (0, 1)$, falls gilt

$$P(X = k) = p(1 - p)^k, \quad k \in \{0, 1, 2, \dots\} = \mathbb{N}.$$

➤ Python: `scipy.stats.geom`

Hinweis: die generierten Werte starten von 1 nicht von 0, d.h. $P(Y = k) = p(1 - p)^{k-1}$ für $k \in \{1, 2, \dots\}$ und $Y - 1 \sim Geo(p)$.

Beispiel: Eine Nachricht wird Bit für Bit gesendet. Die Wahrscheinlichkeit, dass das ein Bit das letzte ist beträgt 0.2. Welches ist die Wahrscheinlichkeit, dass die Nachricht die Länge n hat?

Y : Anzahl der Versuche **bis** (einschließlich) zum ersten Erfolg $\Rightarrow Y$ ist auch die Länge der Nachricht $Y \in \{1, 2, 3, \dots\}$

$$\Rightarrow P(Y = k) = 0.2 \cdot 0.8^{k-1}, \quad k \in \{1, 2, \dots\} = \mathbb{N}^*$$

$\Rightarrow Y - 1 \sim Geo(0.2)$

➡ **Hypergeometrische Verteilung:** $X \sim Hyge(n_1, n_2, n)$, $n, n_1, n_2 \in \mathbb{N}^*$

In einer Urne sind n_1 rote und n_2 blaue Kugeln. Man entnimmt **ohne Zurücklegen** n ($n \leq n_1 + n_2$) Kugeln, davon sind k rot $\Rightarrow n - k$ sind blau. Wir betrachten die Zufallsgröße X = Anzahl der roten entnommenen Kugeln.

Mögliche Werte für X sind $\{0, 1, \dots, n^*\}$ mit

$$n^* = \min(n_1, n) = \begin{cases} n_1 & \text{weniger rote Kugeln als entnommene Kugeln} \\ n & \text{mehr rote Kugeln als entnommene Kugeln.} \end{cases}$$

➡ Bei der Entnahme **ohne Zurücklegen** ändern sich die Entnahme-Wahrscheinlichkeiten in jedem Schritt!

Seien $n_1, n_2, n \in \mathbb{N}$ mit $n \leq n_1 + n_2$ und $n^* = \min(n_1, n)$. Eine Zufallsgröße X mit dem Wertebereich $\{0, \dots, n^*\}$ heißt **hypergeometrisch verteilt** mit den Parametern n_1, n_2, n , falls gilt

$$P(X = k) = \frac{C_{n_1}^k C_{n_2}^{n-k}}{C_{n_1+n_2}^n}, \quad k \in \{0, \dots, n^*\}.$$

➤ Python: `scipy.stats.hypergeom`

Beispiele hypergeometrischer Verteilungen:

Team bilden		Lotto 6 aus 49	
von	$n_1 + n_2 = 12$ Schüler	von	49 Zahlen
sind	$n_1 = 5$ Jungen	sind	6 Glückszahlen
	$n_2 = 7$ Mädchen		43 keine Glückszahlen
wählen	$n = 4$ Schüler		wählen $n = 6$ Zahlen (hier $n = n_1$)
davon sind	$k = 3$ Jungen	davon sind	$k = 2$ richtig getippt
	$n - k = 1$ Mädchen		$n - k = 4$ falsch getippt
	X : Anzahl Jungen im Team		X : Anzahl richtig getippter Zahlen
	z.B. $X = 3$ Jungen im Team		z.B. $X = 2$ Zahlen waren richtig
	$P(X = 3) = \frac{C_5^3 C_7^1}{C_{12}^4}$		$P(X = 2) = \frac{C_6^2 C_{43}^4}{C_{49}^6}$

Beispiel: Von 100 Displays sind $M = 6$ defekt. Wie groß ist die Wahrscheinlichkeit dafür, dass in einer Stichprobe von $n = 5$ zufällig ausgewählten Displays höchstens eines defekt ist?

Unabhängige Zufallsgrößen

Def. 10. Seien die diskreten ZG

$$X \sim \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I} \quad \text{und} \quad Y \sim \begin{pmatrix} y_j \\ q_j \end{pmatrix}_{j \in J}$$

X und Y sind **unabhängige** ZG, falls

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) \quad \forall i \in I, j \in J.$$

Bezeichnung: $P(X = x_i, Y = y_j) = P(\{X = x_i\} \cap \{Y = y_j\}), \forall i \in I, j \in J.$

Bemerkung: Seien die Ereignisse $A_i = \{X = x_i\}, i \in I$, und $B_j = \{Y = y_j\}, j \in J$. Die diskreten ZG X und Y sind unabhängig $\iff \forall (i, j) \in I \times J$ die Ereignisse A_i und B_j sind unabhängig (siehe Def. 5).

Beispiel: Man wirft eine Münze 10 mal. Sei X die ZG, die anzeigt wie oft Kopf erhalten wurde in den ersten 5 Würfeln der Münze; sei Y die ZG, die anzeigt wie oft Kopf erhalten wurde in den letzten 5 Würfeln der Münze. X und Y sind unabhängige ZG. Was für eine Wahrscheinlichkeitsverteilung hat X , bzw. Y ?

Diskrete zufällige Vektoren

$\triangleright (X_1, \dots, X_n)$ ist ein **diskreter zufälliger Vektor**, wenn jede der Komponenten X_1, \dots, X_n eine diskrete ZG ist.

\triangleright Diskrete zufällige Vektoren werden durch ihre Wahrscheinlichkeitsverteilung charakterisiert; z.B.

$$(X, Y) \sim \begin{pmatrix} (x_i, y_j) \\ p_{ij} \end{pmatrix}_{(i,j) \in I \times J}$$

wobei $I, J \subseteq \mathbb{N}$ sind Indexmengen, $p_{ij} = P((X, Y) = (x_i, y_j)) = P(\{X = x_i\} \cap \{Y = y_j\})$,
 $p_{ij} > 0 \forall i \in I, j \in J$, mit $\sum_{(i,j) \in I \times J} p_{ij} = 1$.

Wahrscheinlichkeitsverteilung von (X, Y) in tabellarischer Form

$\begin{array}{c} Y \\ \backslash X \end{array}$	\dots	y_j	\dots
\vdots	\vdots	\vdots	\vdots
x_i	\dots	p_{ij}	\dots
\vdots	\vdots	\vdots	\vdots

mit $p_{ij} = P((X, Y) = (x_i, y_j))$.

▷ Falls X und Y unabhängige ZG sind

$$(1) \quad \Rightarrow p_{ij} = P(\{X = x_i\} \cap \{Y = y_j\}) = P(X = x_i)P(Y = y_j) \quad \forall i \in I, j \in J.$$

▷ Falls X und Y unabhängige ZG sind und ihre Wahrscheinlichkeitsverteilungen sind bekannt, so bestimmt man p_{ij} für den Vektor (X, Y) anhand der Formel (1).

➡ Wenn die Wahrscheinlichkeitsverteilung

$$(X, Y) \sim \left(\begin{array}{c} (x_i, y_j) \\ p_{ij} \end{array} \right)_{(i,j) \in I \times J}$$

bekannt ist, wie bestimmt man die Verteilungen von X , bzw. Y ?

Antwort:

$$P(X = x_i) = \sum_{j \in J} p_{ij} \quad \forall i \in I, \quad P(Y = y_j) = \sum_{i \in I} p_{ij} \quad \forall j \in J.$$

Beispiel: Gegeben ist (X_1, X_2) mit der Wahrscheinlichkeitsverteilung:

$\begin{array}{c} X_2 \\ \backslash X_1 \end{array}$	0	1	2
1	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{2}{16}$
2	$\frac{1}{16}$	$\frac{5}{16}$	$\frac{5}{16}$

a) Welches sind die Wahrscheinlichkeitsverteilungen der ZG: X_1 und X_2 ?

Operationen mit numerischen ZG

Fortsetzung des letzten Beispiels:

b) Welches sind die Wahrscheinlichkeitsverteilungen der ZG: $X_1 + X_2$, $X_1 \cdot X_2$, $X_1^2 - 1$?

c) Sind die ZG X_1 und X_2 unabhängig?

➡ Wenn die Wahrscheinlichkeitsverteilungen der unabhängigen diskreten ZG X und Y bekannt sind, wie bestimmt man die Wahrscheinlichkeitsverteilung für $X + Y$, $X \cdot Y$?

Beispiel: Seien X, Y unabhängige ZG mit

$$X \sim \left(\begin{array}{cc} 0 & 1 \\ \frac{1}{3} & \frac{2}{3} \end{array} \right), \quad Y \sim \left(\begin{array}{ccc} -1 & 0 & 1 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{array} \right)$$

Man bestimme die Wahrscheinlichkeitsverteilungen für a) $2X + 1$, Y^2 , (X, Y) !

b) $X + Y$, $X \cdot Y$, $\max(X, Y)$, $\min(X, Y^2)$?

Der Erwartungswert einer diskreten ZG

Def. 11. Sei die diskrete numerische ZG $X \sim \left(\begin{array}{c} x_i \\ p_i \end{array} \right)_{i \in I}$. Der **Erwartungswert (Mittelwert)** von X ist

$$E(X) = \sum_{i \in I} x_i P(X = x_i),$$

falls $\sum_{i \in I} |x_i| P(X = x_i) < \infty$.

➡ Der Erwartungswert (Mittelwert) einer ZG charakterisiert die zentrale Tendenz seiner Werte.

Satz 8. Seien X, Y diskrete ZG. Es gilt:

→ $E(aX + b) = aE(X) + b$ für alle $a, b \in \mathbb{R}$;

→ $E(X + Y) = E(X) + E(Y)$;

→ Wenn X und Y **unabhängige** ZG sind $\implies E(X \cdot Y) = E(X)E(Y)$.

→ sei $g: \mathbb{R} \rightarrow \mathbb{R}$ eine Funktion, so dass $g(X)$ eine diskrete ZG ist, und

$$E(g(X)) = \sum_{i \in I} g(x_i) P(X = x_i),$$

falls $\sum_{i \in I} |g(x_i)| P(X = x_i) < \infty$.

► Python: `numpy.mean(x)` berechnet für $x = [x_0, \dots, x_{n-1}]$ das arithmetische Mittel $\frac{1}{n}(x_0 + \dots + x_{n-1})$

Sind $x = [x_0, \dots, x_{n-1}]$ zufällig generierte Werte für die diskrete ZG X dann gilt

$$E(X) \approx \text{numpy.mean}(x) = \frac{1}{n}(x_0 + \dots + x_{n-1}).$$

```
import numpy
x = [[1, 3], [5, 9]]
print("Arithmetisches Mittel (Matrix):", numpy.mean(x))
print("Arithmetisches Mittel (Spalten):", numpy.mean(x, axis = 0)) #axis = 0
print("Arithmetisches Mittel (Zeilen):", numpy.mean(x, axis = 1)) #axis = 1
```

Beispiel: Man würfelt; wenn 6 auftaucht, gewinnt man 3\$, wenn 1 auftaucht gewinnt man 2\$, wenn 2,3,4,5 auftauchen, verliert man 1\$. Wie viel gewinnt oder verliert *im Mittel* ein Spieler in 30 solchen Spielen?

Lsg.: Sei X der Gewinn oder Verlust in einem Spiel

$$X \sim \begin{pmatrix} -1 & 2 & 3 \\ \frac{4}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix} \implies E(X) = (-1) \cdot \frac{4}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} = \frac{1}{6}.$$

Für $i \in \{1, \dots, 30\}$ sei X_i der Gewinn oder Verlust im i -ten Spiel; X_i hat dieselbe Verteilung wie X . Der Gewinn nach 30 Spielen ist

$$E(X_1 + \dots + X_{30}) = E(X_1) + \dots + E(X_{30}) = 30 \cdot E(X) = 5.$$

Der Spieler gewinnt *im Mittel* nach 30 Spielen 5\$.

```
import numpy
import random
s=[]
N=1000
for _ in range(N):
    spiele = random.choices([-1,-1,-1,-1,2,3], k=30)
    s.append(sum(spiele)) # Gewinn nach 30 Spielen
print("Mittlerer Gewinn (nach 30 Spielen):", numpy.mean(s))
```

Übung: Man berechne die Erwartungswerte für: $X \sim Unid(n)$, $X \sim Bernoulli(p)$, $X \sim Bino(n, p)$.

Def. 12. Seien X_1, \dots, X_n cu $n \in \mathbb{N}$, $n \geq 2$, ZG welche diskrete Werte in den Mengen $\mathcal{X}_1, \dots, \mathcal{X}_n$ annehmen. X_1, \dots, X_n sind **unabhängige Zufallsgrößen**, wenn

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \cdot \dots \cdot P(X_n = x_n)$$

für alle $x_1 \in \mathcal{X}_1, \dots, x_n \in \mathcal{X}_n$

Beispiel: Man würfelt vier Mal. Sei X_i die Zahl die im i -ten Wurf erhalten wurde.

- a) X_1, X_2, X_3, X_4 sind unabhängige ZG;
- b) $X_1 + X_2$ und $X_3 + X_4$ sind unabhängige ZG;
- c) $X_1 + X_2 + X_3$ und X_4 sind unabhängige ZG.

Satz 9. Seien X_1, \dots, X_n unabhängige ZG. Für beliebige Indizes $i_1, \dots, i_k \subset \{1, \dots, n\}$ sind die ZG X_{i_1}, \dots, X_{i_k} unabhängig.

Def. 13. Die **Verteilungsfunktion** einer diskreten Zufallsgröße X , welche die Werte $\{x_i : i \in I\}$ annimmt, ist $F : \mathbb{R} \rightarrow [0, 1]$ definiert als

$$F(x) := P(X \leq x) = \sum_{i \in I: x_i \leq x} P(X = x_i) \text{ für jedes } x \in \mathbb{R}.$$

Beispiel: Die diskrete ZG X hat folgende Wahrscheinlichkeitsverteilung:

$$P(X = -1) = 0.5, \quad P(X = 1) = 0.3, \quad P(X = 4) = 0.2.$$

$\Rightarrow X$ hat folgende Verteilungsfunktion $F_X : \mathbb{R} \rightarrow [0, 1]$

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & \text{wenn } x < -1 \\ 0.5, & \text{wenn } -1 \leq x < 1 \\ 0.5 + 0.3 = 0.8, & \text{wenn } 1 \leq x < 4 \\ 0.5 + 0.3 + 0.2 = 1, & \text{wenn } 4 \leq x. \end{cases}$$

Übung: Sei $F : \mathbb{R} \rightarrow [0, 1]$ die Verteilungsfunktion einer diskreten ZG X

$$F(x) = P(X \leq x) = \begin{cases} 0, & \text{dacă } x < -2 \\ 0.5, & \text{dacă } -2 \leq x < 1 \\ 0.7, & \text{dacă } 1 \leq x < 2 \\ 1, & \text{dacă } 2 \leq x. \end{cases}$$

Man bestimme den Erwartungswert von X .



Satz 10. Die Eigenschaften der Verteilungsfunktion F einer diskreten ZG X sind:

1. F ist monoton wachsend, d.h. aus $x_1 < x_2$ folgt $F(x_1) \leq F(x_2)$.
2. $\lim_{x \rightarrow \infty} F(x) = 1$ und $\lim_{x \rightarrow -\infty} F(x) = 0$
3. F ist rechtsseitig stetig, d.h. $\lim_{x \searrow x_0} F(x) = F(x_0)$ für $\forall x_0 \in \mathbb{R}$.
4. $P(a < X \leq b) = F(b) - F(a)$

► Jede Funktion $F : \mathbb{R} \rightarrow \mathbb{R}$, welche die Eigenschaften 1., 2. und 3. des obigen Satzes erfüllt, ist eine Verteilungsfunktion.

► Python: `scipy.stats.binom.cdf(x, n, p)`, `scipy.stats.hypergeom.cdf(x, M, n, N)`, berechnet die Verteilungsfunktion $F(x) = P(X \leq x)$ für $X \sim \text{Bino}(n, p)$, $X \sim \text{Hyge}(M, n, N)$.

```

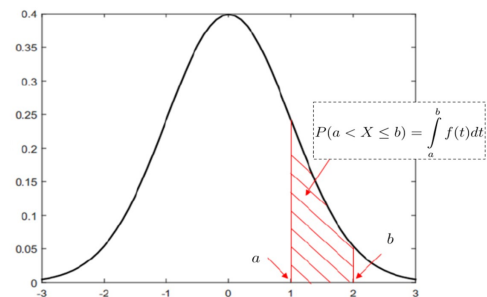
#In einer Urne sind 10 Kugeln, davon sind genau 5 rot.
# Man zieht 5-mal, mit Zurucklegen;
# sei X= ZG Anzahl gezogene rote Kugeln => X hat Bino(5,0.5) Verteilung
# Grafische Darstellung der Verteilungsfunktion fur Bino(5,0.5)
import scipy.stats
import matplotlib.pyplot as plt
import numpy as np
n=5
p=0.5
x = np.linspace(-2, n+2, 100)
y=scipy.stats.binom.cdf(x,n,p)
plt.plot(x, y, "r.", label="bino.cdf")
plt.xlabel("x")
plt.ylabel("F(x)=P(X <= x)")
plt.title("Verteilungsfkt. Bino(5,0.5)")
plt.legend()
plt.xticks(range(-2,8))
plt.grid()
plt.show()

```

Dichtefunktion einer stetigen ZG

Zur Erinnerung: Eine Zufallsgröße heißt stetig, wenn sie (überabzählbar viele) Werte in einem Intervall (oder Vereinigung von Intervallen) oder \mathbb{R} annehmen kann!

► **Stetige ZG:** sind numerische Zufallsgrößen; z.B.: die Brenndauer einer Glühlampe, die Abfüllmenge in Konserven, die Länge von hergestellten Schrauben, die Temperatur in einer Stadt innerhalb eines Jahres, die Geschwindigkeit eines an einer Radarkontrolle vorbeifahrenden Autos, die Laufzeit einer Maschine bis zur ersten Betriebsstörung, die Verspätung eines bestimmten Zuges usw.



Def. 14. Die Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ heißt **Dichtefunktion** der stetigen Zufallsgröße X , falls für alle $x \in \mathbb{R}$ gilt

$$P(X \leq x) = \int_{-\infty}^x f(t) dt.$$

Die **Verteilungsfunktion** einer stetigen Zufallsgröße X , welche die Dichtefunktion f hat, ist $F : \mathbb{R} \rightarrow [0, 1]$ definiert als

$$F(x) := P(X \leq x) = \int_{-\infty}^x f(t) dt \text{ für jedes } x \in \mathbb{R}.$$

Satz 11. Seien f Dichtefunktion und F die Verteilungsfunktion einer stetigen Zufallsgröße. Dann gilt:

(1) $f(x) \geq 0$ für alle $x \in \mathbb{R}$;

(2) $\int_{-\infty}^{\infty} f(t) dt = 1$;

$$(3) F(b) - F(a) = P(a < X \leq b) = \int_a^b f(t)dt \text{ für alle } a, b \in \mathbb{R}, a < b;$$

$$(4) P(X = a) = 0 \quad \forall a \in \mathbb{R};$$

$$(5) \text{ für alle } a, b \in \mathbb{R}, a < b, \text{ gilt}$$

$$F(b) - F(a) = P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) = \int_a^b f(t)dt ;$$

$$(6) P(X \in M) = \int_M f(t)dt, \quad M \subseteq \mathbb{R};$$

$$(7) F \text{ ist eine monoton wachsende und stetige Funktion auf } \mathbb{R};$$

$$(8) \lim_{x \rightarrow \infty} F(x) = 1 \quad \text{und} \quad \lim_{x \rightarrow -\infty} F(x) = 0;$$

$$(9) \text{ ist } F \text{ ableitbar im Punkt } x, \text{ dann gilt } F'(x) = f(x).$$

➡ Jede Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$, welche die Eigenschaften (1) und (2) des Satzes 11 erfüllt, ist eine **Dichtefunktion**.

➡ Jede Funktion $F : \mathbb{R} \rightarrow \mathbb{R}$, welche die Eigenschaften (7) und (8) des Satzes 11 erfüllt, ist die **Verteilungsfunktion** einer stetigen ZG.

➡ Die Dichtefunktion einer stetigen ZG kann in abzählbar vielen Punkten geändert werden, trotzdem bleibt sie eine Dichtefunktion für dieselbe ZG; die Dichtefunktion einer stetigen ZG ist *nicht eindeutig* im Sinne der Gleichheit in allen Punkten aus \mathbb{R} . Sei f_1 die Dichtefunktion einer stetigen ZG X und sei $f_2 : \mathbb{R} \rightarrow [0, \infty]$, so dass $f_1(x) = f_2(x)$ für alle $x \in \mathbb{R} \setminus \mathcal{N}$, wobei \mathcal{N} abzählbar ist. Dann ist f_2 auch eine Dichtefunktion für X . Aus den Eigenschaften der Integrale folgt

$$\int_{-\infty}^x f_1(t)dt = \int_{-\infty}^x f_2(t)dt \quad \forall x \in \mathbb{R},$$

also f_1 und auch f_2 entsprechen derselben Verteilungsfunktion (für X , siehe Def.14).

Klassische stetige Verteilungen

➡ **Gleichmäßige Verteilung** $X \sim \text{Unif}[a, b]$; Parameter: $a, b \in \mathbb{R}$ mit $a < b$

Eine stetige Zufallsgröße X heißt auf dem Intervall $[a, b]$ **gleichverteilt**, falls ihre Dichtefunktion gegeben ist durch

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } x \in [a, b] \\ 0 & \text{sonst} \end{cases}$$

Beispiel: Die Verkehrsnachrichten melden einen Unfall auf der Autobahn zwischen KM 60 und 100. Da keine weiteren Informationen verfügbar sind, betrachten wir die genaue Position des Unfalls als gleichverteilt auf dem genannten Streckenabschnitt.

Wie groß ist die Wahrscheinlichkeit dafür, dass sich der Unfall zwischen KM 80 und 85 ereignet hat? Man vergleiche die Simulations-Ergebnisse mit den theoretischen Ergebnissen.

$$\text{Theoretische Wkt.: } P(80 \leq X \leq 85) = \int_{80}^{85} \frac{1}{100-60} dt = \frac{85-80}{40} = 0.125.$$



➡ **Normalverteilung (Gauss-Verteilung)** $X \sim N(\mu, \sigma^2)$; Parameter: $\mu, \sigma \in \mathbb{R}$ mit $\sigma > 0$

Wegen der Gestalt ihres Graphen wird die Dichtefunktion f auch *Gaußsche Glockenkurve* genannt. Sie findet sich auf dem letzten Zehnmarkschein vor der Euro-Einführung neben dem Bildnis von Carl Friedrich Gauß (1777-1855).

Anwendungen

- Messfehler / Fehlerrechnung (z.B. in der Regressionsanalyse)



C. F. Gauss - Mathematiker, Astronom, Physiker, Fachmann der Geodesie; im Bild im Hintergrund - wichtige Gebäude aus Göttingen (Universität, Sternwarte);

- Brownsche Molekularbewegung
- Mathematische Statistik → die Normalverteilung wird oft verwendet, wenn die eigentliche, den Daten zugrunde liegende Verteilungsfunktion, unbekannt ist (der zentrale Grenzwertsatz)
- Meteorologie (Regenmenge, Sonnenstunden); Produktion (Kapazitäten, Qualitätssicherung, Füllmenge von Packungen); Gesellschaftswissenschaften (Einkommen, Intelligenz, Sozialkompetenz) Finanzmärkte (Preisänderung von Aktien) .

- die **Normalverteilung** $N(\mu, \sigma^2)$ hat folgende Dichtefunktion

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

- für $\mu = 0, \sigma = 1$: $N(0, 1)$ heißt *standard Normalverteilung*

➡ **Exponentialverteilung** $X \sim \text{Exp}(\lambda)$; Parameter $\lambda > 0$

- die Exponentialverteilung mit (Intensitäts-) Parameter $\lambda > 0$, hat folgende Dichtefunktion

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{für } x > 0 \\ 0, & \text{für } x \leq 0 \end{cases}$$

Beispiel: Sei $X \sim \text{Exp}(0.5)$ die ZG welche die Lebensdauer einer Batterie anzeigt (wie viele Monate die Batterie funktioniert). Man berechne: a) $P(2 \leq X \leq 4)$; b) $P(X > 3)$.

Theoretische Wkt.: $P(2 \leq X \leq 4) = \int_2^4 0.5e^{-0.5t} dt = -e^{-0.5t} \Big|_2^4 = e^{-1} - e^{-2} \approx 0.23254$

Theoretische Wkt.: $P(X > 3) = \int_3^\infty 0.5e^{-0.5t} dt = -e^{-0.5t} \Big|_3^\infty = e^{-1.5} \approx 0.22313.$



➡ **Student Verteilung:** $X \sim T(n)$; Parameter $n \in \mathbb{N}^*$

- die Student Verteilung mit $n \in \mathbb{N}^*$ Freiheitsgraden hat folgende Dichtefunktion

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

wobei

$$\Gamma(a) = \int_0^\infty v^{a-1} \exp(-v) dv, a > 0$$

ist die Eulersche Gamma-Funktion.

➡ **Chi-Quadrat Verteilung:** $X \sim \chi^2(n)$; Parameter $n \in \mathbb{N}^*$

• die χ^2 -Verteilung mit $n \in \mathbb{N}^*$ Freiheitsgraden hat folgende Dichtefunktion

$$f(x) = \begin{cases} 0, & \text{wenn } x \leq 0 \\ \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} x^{\frac{n}{2}-1} \exp\left(-\frac{x}{2}\right), & \text{wenn } x > 0, \end{cases}$$

Beispiel: Sei $Z \sim \text{Unif}[a, b]$. Man berechne die Verteilungsfunktion F_Z und $P(Z > \frac{2a+b}{3})$.

Lsg.: $Z \sim \text{Unif}[a, b]$

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

$$\Rightarrow F_Z(x) = P(Z \leq x) = \int_{-\infty}^x f_Z(t) dt = \begin{cases} \int_{-\infty}^x 0 dt = 0, & \text{wenn } x < a \\ \int_{-\infty}^a 0 dt + \int_a^x f_Z(t) dt = \frac{x-a}{b-a}, & \text{wenn } a \leq x < b \\ \int_{-\infty}^a 0 dt + \int_a^b \frac{1}{b-a} dt + \int_b^x 0 dt = 1, & \text{wenn } b \leq x. \end{cases}$$

Es gilt $a < \frac{2a+b}{3} < b$

$$\Rightarrow P\left(Z > \frac{2a+b}{3}\right) = 1 - P\left(Z \leq \frac{2a+b}{3}\right) = 1 - F_Z\left(\frac{2a+b}{3}\right) = 1 - \frac{\frac{2a+b}{3} - a}{b-a} = \frac{2}{3}.$$

Übung: Sei X die ZG, die Lebensdauer in Stunden eines Gerätes (bis zum ersten Ausfall) angibt. Es ist bekannt $P(X > x) = 2^{-x}, x > 0$ und $P(X > x) = 1, x \leq 0$. Man bestimme die Dichtefunktion f und $P(2 < X < 3)$. ♠

Verteilungsfunktion eines Zufallsvektors

➡ (X, Y) ist ein **diskreter Zufallsvektor**, wenn X und Y diskrete ZG sind.

➡ (X, Y) ist ein **stetiger Zufallsvektor**, wenn X und Y stetige ZG sind.

Def. 15. Sei der Zufallsvektor (X, Y) (diskret oder stetig). Die Funktion $F_{(X,Y)} : \mathbb{R} \times \mathbb{R} \rightarrow [0, 1]$

$$F_{(X,Y)}(x, y) = P(X \leq x, Y \leq y) = P(\{X \leq x\} \cap \{Y \leq y\}), \quad x, y \in \mathbb{R}$$

heißt **Verteilungsfunktion** des Zufallsvektors (X, Y) .

Wenn $F_{(X,Y)}$ bekannt ist, so berechnet man die Verteilungsfunktionen F_X und F_Y auf folgende Weise:

$$F_X(x) = P(X \leq x) = \lim_{y \rightarrow \infty} F_{(X,Y)}(x, y); \quad F_Y(y) = P(Y \leq y) = \lim_{x \rightarrow \infty} F_{(X,Y)}(x, y).$$

Def. 16. Zwei Zufallsgrößen X, Y (diskret oder stetig) sind **unabhängig**, genau dann, wenn für alle $x, y \in \mathbb{R}$ gilt

$$P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y), \text{ d.h. } F_{(X,Y)}(x, y) = F_X(x) \cdot F_Y(y).$$

Beispiel: Die Verteilungsfunktion des Zufallsvektors (X, Y) ist $F_{(X,Y)} : \mathbb{R} \times \mathbb{R} \rightarrow [0, 1]$

$$F_{(X,Y)}(x, y) = \begin{cases} 0, & \text{wenn } x < 0 \text{ oder } y < 1 \\ x(y-1), & \text{wenn } 0 \leq x < 1 \text{ und } 1 \leq y < 2 \\ x, & \text{wenn } 0 \leq x < 1 \text{ und } 2 \leq y \\ y-1, & \text{wenn } 1 \leq x \text{ und } 1 \leq y < 2 \\ 1, & \text{wenn } 1 \leq x \text{ und } 2 \leq y. \end{cases}$$

Sind X und Y unabhängige ZG? Man berechne die Dichtefunktionen f_X , bzw. f_Y ?

Lsg.: Man berechnet F_X, F_Y und danach $f_X = F'_X, f_Y = F'_Y \implies X \sim Unif[0, 1], Y \sim Unif[1, 2]$. Man beweist, dass $F_{(X,Y)} = F_X \cdot F_Y$ und es folgt, dass X und Y unabhängige ZG sind. ♣

Der Erwartungswert (Mittelwert) einer stetigen ZG

Def. 17. Der Erwartungswert oder Mittelwert einer stetigen Zufallsgröße X ist gegeben durch

$$E(X) = \int_{-\infty}^{\infty} t f_X(t) dt, \text{ vorausgesetzt, dass } \int_{-\infty}^{\infty} |t| f_X(t) dt < \infty \text{ gilt.}$$

➡ Der Erwartungswert (Mittelwert) einer ZG charakterisiert die zentrale Tendenz seiner Werte.

Satz 12. (Eigenschaften des Erwartungswertes) Es seien X und Y stetige ZG. Dann gilt

1. $E(aX + b) = aE(X) + b$ für $a, b \in \mathbb{R}$.
2. $E(X + Y) = E(X) + E(Y)$.
3. Sind X und Y **unabhängige ZG**, dann gilt $E(X \cdot Y) = E(X) \cdot E(Y)$.
4. Sei $g : \mathbb{R} \rightarrow \mathbb{R}$, so dass $g(X)$ eine stetige ZG ist $\implies E(g(X)) = \int_{-\infty}^{\infty} g(t) f_X(t) dt$, falls $\int_{-\infty}^{\infty} |g(t)| f_X(t) dt < \infty$, wobei f_X die Dichtefunktion der stetigen ZG X ist.

Beispiel: Sei X die ZG, welche anzeigt, wie lange der Weg in Minuten von Peter morgens zur Schule dauert. Man weiss $X \sim Unif[20, 26]$. Morgens hat Peter Stunden von 8 Uhr. Er geht um 7:35 von zu Hause los. Mit welcher Wahrscheinlichkeit ist er rechtzeitig in der Schule? Wie lange dauert *im Mittel* der Weg von Peter zur Schule?

Lsg.: $X \sim Unif[20, 26]$

$$\implies f_X(t) = \begin{cases} \frac{1}{26-20} = \frac{1}{6}, & \text{wenn } 20 \leq t \leq 26 \\ 0, & \text{sonst.} \end{cases}$$

$$P(\text{"Peter ist rechtzeitig in der Schule"}) = P(X \leq 25) = \int_{-\infty}^{25} f_X(t) dt = \int_{20}^{25} \frac{1}{6} dt = \frac{25 - 20}{6} = \frac{5}{6}.$$

$$E(X) = \int_{-\infty}^{\infty} t f_X(t) dt = \int_{20}^{26} \frac{t}{6} dt = \frac{1}{6} \cdot \frac{t^2}{2} \Big|_{20}^{26} = 23 \text{ (Minuten)}.$$



→ Spezialfall im Satz 12: $g(z) = (z - E(X))^2 \rightarrow E[g(X)] = E[(X - E(X))^2]$ (die Varianz von X)

Def. 18. Sei X eine diskrete oder stetige Zufallsgröße.

$V(X) = E[(X - E(X))^2]$ ist **Varianz** bzw. **Streuung** von X ,
 $St(X) = \sqrt{V(X)}$ ist die **Standardabweichung** der Zufallsgröße X .

➡ Die Varianz, bzw. Standardabweichung, kann als Kennzahl für die Größe der Fluktuationen (Streuung) der Werte der ZG X um den Erwartungswert $E(X)$ interpretiert werden.

Satz 13. Eigenschaften der Varianz (für **diskrete** bzw. **stetige ZG**):

1. $V(X) = E(X^2) - (E(X))^2$.
2. $V(aX + b) = a^2 V(X) \forall a, b \in \mathbb{R}$.
3. Sind X und Y **unabhängige ZG** $\Rightarrow V(X + Y) = V(X) + V(Y)$.

➤ Erwartungswert $E(X)$, Varianz $V(X)$, Standardabweichung $St(X)$ in Python: seien $x = [x_0, \dots, x_{n-1}]$ zufällig generierte Werte der ZG X

$$(2) \quad E(X) \approx \text{numpy.mean}(x) = \frac{1}{n}(x_0 + \dots + x_{n-1}), \text{ wenn } n \text{ groß ist}$$

$$(3) \quad V(X) \approx \text{numpy.var}(x) = \frac{1}{n} \sum_{i=0}^{n-1} (x_i - \text{numpy.mean}(x))^2, \text{ wenn } n \text{ groß ist}$$

$$St(X) \approx \text{numpy.std}(x) = \left(\frac{1}{n} \sum_{i=0}^{n-1} (x_i - \text{numpy.mean}(x))^2 \right)^{\frac{1}{2}}, \text{ wenn } n \text{ groß ist.}$$

```
#Anhand Simulationen: Berechnen des Erwartungswertes
# der Varianz, der Standardabweichung für N(0,1) Zufallswerte
import numpy
from scipy.stats import norm
N=1000
res=norm.rvs(0,1,N)
m = numpy.mean(res)
print(f"Erwartungswert mit numpy.mean: {m:.6f}")
v1=numpy.var(res) # oder mit numpy.mean([(x - m)**2 for x in res])
print(f"Varianz mit numpy.var: {v1:.6f}")
s=numpy.std(res)
print(f"Standardabweichung mit numpy.std: {s:.6f}") # oder mit v1**(1/2)
```

Diskrete ZG	Stetige ZG
<ul style="list-style-type: none"> • charakterisiert durch die Wahrscheinlichkeitsverteilung $X \sim \left(\begin{matrix} x_i \\ P(X = x_i) \end{matrix} \right)_{i \in I}$ <ul style="list-style-type: none"> • $\sum_{i \in I} P(X = x_i) = 1$ • $P(X \in A) = \sum_{i \in I: x_i \in A} P(X = x_i)$ • Verteilungsfunktion $F(x) = P(X \leq x) \forall x \in \mathbb{R}$ • $F(x) = \sum_{i \in I: x_i \leq x} P(X = x_i) \forall x \in \mathbb{R}$ • F ist rechtsstetig in jedem Punkt $x \in \mathbb{R}$ • F unstetig in den Punkten $x_i, \forall i \in I$ • $\forall a < b, a, b \in \mathbb{R}$ $P(a \leq X \leq b) = \sum_{\substack{i \in I \\ a \leq x_i \leq b}} P(X = x_i)$ $P(a < X < b) = \sum_{\substack{i \in I \\ a < x_i < b}} P(X = x_i)$ $P(a < X \leq b) = F(b) - F(a)$ <ul style="list-style-type: none"> • $P(X = a) = 0$ falls $a \notin \{x_i : i \in I\}$ • Erwartungswert $E(X) = \sum_{i \in I} x_i P(X = x_i)$ 	<ul style="list-style-type: none"> • charakterisiert durch die Dichtefunktion f $P(X \leq x) = \int_{-\infty}^x f(t) dt$ <ul style="list-style-type: none"> • $\int_{-\infty}^{\infty} f(t) dt = 1$ • $P(X \in A) = \int_A f(t) dt$ • Verteilungsfunktion $F(x) = P(X \leq x) \forall x \in \mathbb{R}$ • $F(x) = \int_{-\infty}^x f(t) dt \quad \forall x \in \mathbb{R}$ • F ist stetig in jedem Punkt $x \in \mathbb{R}$ • $\forall a < b, a, b \in \mathbb{R}$ $P(a \leq X \leq b) = P(a < X \leq b)$ $= P(a \leq X < b) = P(a < X < b) = \int_a^b f(t) dt$ $= F(b) - F(a)$ <ul style="list-style-type: none"> • $P(X = a) = \int_a^a f(t) dt = 0 \quad \forall a \in \mathbb{R}$ • ist F ableitbar im Punkt $x \Rightarrow F'(x) = f(x)$ • Erwartungswert $E(X) = \int_{-\infty}^{\infty} t f(t) dt.$

Beispiele: 1) Wenn $X \sim \text{Bino}(n, p)$, dann gilt $E(X) = np$ und $V(X) = np(1 - p)$.

Lsg.: Für $i \in \{1, \dots, n\}$ sei $X_i \sim \text{Bernoulli}(p)$ (d.h. $P(X_i = 1) = p, P(X_i = 0) = 1 - p$), so dass X_1, \dots, X_n unabhängige ZG sind. Dann gilt $X_1 + \dots + X_n \sim \text{Bino}(n, p)$, also $X_1 + \dots + X_n$ und X haben dieselbe Verteilung \Rightarrow sie haben den gleichen Erwartungswert und die gleiche Varianz

$$E(X) = E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n) = p + \dots + p = np.$$

Die ZG X_1, \dots, X_n sind unabhängig, mit Satz 13, erhält man

$$V(X) = V(X_1 + \dots + X_n) = V(X_1) + \dots + V(X_n) = np(1 - p) = np(1 - p).$$

2) Wenn $X \sim N(\mu, \sigma^2)$, dann gilt $E(X) = \mu, V(X) = \sigma^2$.

Lsg.: Die Dichtefunktion von X ist

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}, x \in \mathbb{R}.$$

Wenn $\mu = 0$ und $\sigma = 1$ erhält man die Dichtefunktion der standard Normalverteilung

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{x^2}{2} \right\}, x \in \mathbb{R}.$$

$$\text{Aus Satz 11-(2)} \implies \int_{-\infty}^{\infty} \varphi(t) dt = 1.$$

In den folgenden Rechnungen benutzt man die Variablenänderung $t = \frac{x - \mu}{\sigma}$

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x) dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} dx \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t \exp \left\{ -\frac{t^2}{2} \right\} dt + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{t^2}{2} \right\} dt \\ &= 0 + \mu \int_{-\infty}^{\infty} \varphi(t) dt = \mu. \end{aligned}$$

Man benutzt dieselbe Variablenänderung wie vorher und danach partielle Integration

$$\begin{aligned} \implies V(X) &= E[(X - \mu)^2] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x - \mu)^2 \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} dx \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 \exp \left\{ -\frac{t^2}{2} \right\} dt = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t \left(-\exp \left\{ -\frac{t^2}{2} \right\} \right)' dt \\ &= t \left(-\exp \left\{ -\frac{t^2}{2} \right\} \right) \Big|_{-\infty}^{\infty} - \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(-\exp \left\{ -\frac{t^2}{2} \right\} \right) dt \\ &= 0 - 0 + \sigma^2 \int_{-\infty}^{\infty} \varphi(t) dt = \sigma^2. \end{aligned}$$



Def. 19. Wenn (X, Y) *stetiger Zufallsvektor* ist, dann heißt $f_{(X,Y)} : \mathbb{R} \times \mathbb{R} \rightarrow [0, \infty)$ *gemeinsame Dichtefunktion* von X und Y , wenn

$$P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f_{(X,Y)}(u, v) du dv \text{ für alle } x, y \in \mathbb{R}.$$

Satz 14. Für einen *stetigen Zufallsvektor* (X, Y) gelten die Eigenschaften:

1. $f_{(X,Y)}(x, y) \geq 0$ für alle $(x, y) \in \mathbb{R}^2$;
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{(X,Y)}(u, v) du dv = 1$;
3. $F_{(X,Y)}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{(X,Y)}(u, v) du dv$, $x, y \in \mathbb{R}$, wobei $F_{(X,Y)}$ die Verteilungsfunktion des Vektors (X, Y) ist.
4. $F_{(X,Y)}$ ist eine stetige Funktion.

5. Falls $F_{(X,Y)}$ partiell ableitbar in (x, y) ist, dann gilt:

$$\frac{\partial^2 F_{(X,Y)}(x, y)}{\partial x \partial y} = f_{(X,Y)}(x, y).$$

$$6. P((X, Y) \in M) = \underbrace{\int \int_M}_{M} f_{(X,Y)}(u, v) du dv, \quad M \subset \mathbb{R}^2.$$

Wenn $f_{(X,Y)}$ bekannt ist, so bestimmt man f_X und f_Y mit folgenden Formeln

$$f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, v) dv, \quad x \in \mathbb{R}, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(u, y) du, \quad y \in \mathbb{R}.$$

Satz 15. X (mit Dichtefunktion f_X) und Y (mit Dichtefunktion f_Y) sind **unabhängige stetige Zufallsgrößen** genau dann, wenn für ihre Dichtefunktionen für alle $x, y \in \mathbb{R}$ gilt

$$f_{(X,Y)}(x, y) = f_X(x) \cdot f_Y(y).$$

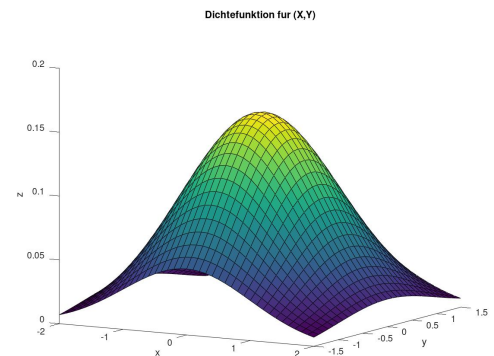
Beispiel einer zweidimensionalen normalen Dichtefunktion: (X, Y) hat folgende Dichtefunktion:

$$f_{(X,Y)}(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}, \quad x, y \in \mathbb{R}.$$

$$\Rightarrow f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dy = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad \forall x \in \mathbb{R},$$

$$\Rightarrow f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dx = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}, \quad \forall y \in \mathbb{R}.$$

$$\Rightarrow X, Y \sim N(0, 1) \text{ und sind unabhängige ZG, weil } f_{(X,Y)} = f_X \cdot f_Y.$$



Übungen: 1) Sei (X, Y) stetiger zufälliger Vektor mit gemeinsamer Verteilungsfunktion

$$F_{(X,Y)}(x, y) = \begin{cases} (1 - e^{-x})(1 - e^{-2y}) & \text{für } x > 0 \text{ und } y > 0 \\ 0 & \text{sonst.} \end{cases}$$

Sind X und Y unabhängig? Man berechne $P(1 \leq X \leq 2 \leq Y \leq 3)$.

2) Sei (X, Y) stetiger zufälliger Vektor mit gemeinsamer Dichtefunktion $f_{(X,Y)} : \mathbb{R}^2 \rightarrow \mathbb{R}$

$$f_{(X,Y)}(x, y) = \begin{cases} ce^{-2x}e^{-3y} & \text{für } x > 0 \text{ und } y > 0 \\ 0 & \text{sonst.} \end{cases}$$

Man bestimme die Konstante $c \in \mathbb{R}$. Sind X und Y unabhängig?

3) Der zufällige Vektor (X, Y) hat die Dichtefunktion

$$f_{(X,Y)} : \mathbb{R}^2 \rightarrow \mathbb{R} \quad f_{(X,Y)}(x, y) = \begin{cases} x - y, & \text{wenn } 0 \leq x \leq 1 \text{ si } -1 \leq y \leq 0 \\ 0, & \text{sonst.} \end{cases}$$

Man berechne $E(X)$ und $E(X^2)$.

Wiederholungsübung: Wenn $X \sim \text{Exp}(\lambda)$ dann zeige man, dass $E(X) = \frac{1}{\lambda}$, $V(X) = \frac{1}{\lambda^2}$.



Das Gesetz der großen Zahlen

Def. 20. $(X_n)_n$ ist eine **Folge von unabhängigen ZG**, wenn $\forall \{i_1, \dots, i_k\} \subset \mathbb{N}$ die ZG X_{i_1}, \dots, X_{i_k} unabhängig sind, d.h.

$$P(X_{i_1} \leq x_{i_1}, \dots, X_{i_k} \leq x_{i_k}) = P(X_{i_1} \leq x_{i_1}) \cdot \dots \cdot P(X_{i_k} \leq x_{i_k})$$

$$\forall x_{i_1}, \dots, x_{i_k} \in \mathbb{R}.$$

Beispiel: 1) die Zufallsgröße X_n = die angezeigte Zahl im n -ten Wurf eines Würfels $\Rightarrow (X_n)_n$ ist eine Folge von unabhängigen ZG

2) Sei für $n \in \mathbb{N}^*$

$$X_n = \begin{cases} 1, & \text{wenn im } n\text{-ten Wurf einer Münze Zahl auftaucht} \\ 0, & \text{wenn im } n\text{-ten Wurf einer Münze Kopf auftaucht.} \end{cases}$$

$(X_n)_n$ ist eine Folge von unabhängigen Ereignissen.

Def. 21. Die Folge $(X_n)_n$ von ZG **konvergiert fast sicher** zur ZG X , wenn

$$P\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\}\right) = 1.$$

Bezeichnung: $X_n \xrightarrow{f.s.} X$

Beispiel: $\Omega := [0, 1]$ Grundraum, sei P das Wahrscheinlichkeitsmaß auf $[0, 1]$, d.h. für alle $\alpha < \beta$ aus $[0, 1]$ berechnet man

$$P([\alpha, \beta]) = P([\alpha, \beta)) = P((\alpha, \beta]) = P((\alpha, \beta)) := \beta - \alpha$$

$$X_n(\omega) = \omega + \omega^n + (1 - \omega)^n, \omega \in [0, 1], n \geq 1 \Rightarrow P(\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = ???) = 1.$$

$$\lim_{n \rightarrow \infty} X_n(\omega) = \begin{cases} \omega & \text{für } \omega \in (0, 1) \\ 1 & \text{für } \omega = 0 \\ 2 & \text{für } \omega = 1. \end{cases}$$

Sei $X(\omega) = \omega$ für alle $\omega \in \Omega$, $\Rightarrow \{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = \omega\} = (0, 1)$

$$\Rightarrow P\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\}\right) = P((0, 1)) = 1 \Rightarrow X_n \xrightarrow{f.s.} X.$$

Satz 16. Sei $(X_n)_n$ Folge von unabhängigen ZG mit der gleichen Verteilung und endlichem Erwartungswert, d.h. $E(X_n) = m \in \mathbb{R}$ für alle $n \in \mathbb{N}^* \Rightarrow (X_n)_n$ erfüllt das **starke Gesetz der großen Zahlen (SGGZ)**, d.h.

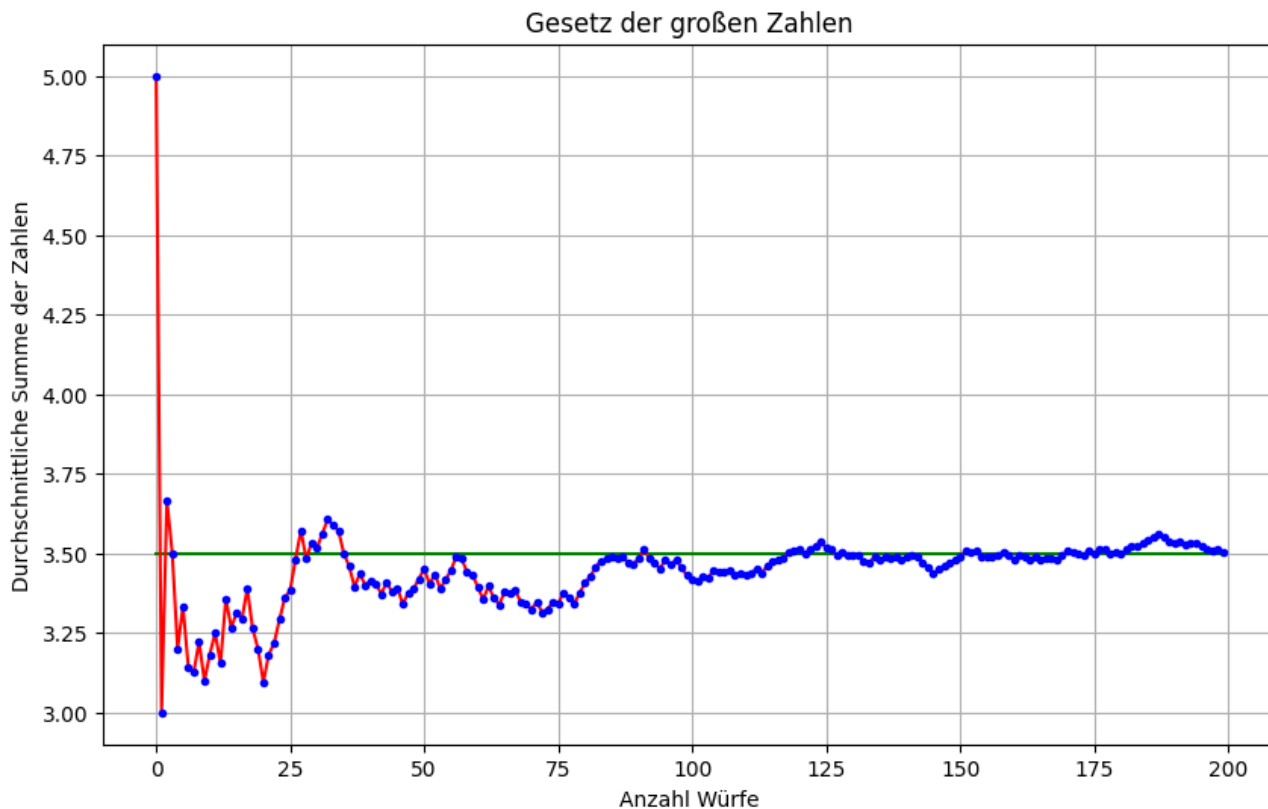
$$P\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} \frac{1}{n} (X_1(\omega) + \dots + X_n(\omega)) = m\right\}\right) = 1,$$

d.h.

$$\frac{1}{n} (X_1 + \dots + X_n) \xrightarrow{f.s.} m.$$

In der Praxis: $\frac{1}{n} (X_1 + \dots + X_n) \approx m$, wenn n hinreichend groß ist! Siehe Formel (2).

Bemerkung: Das Gesetz der großen Zahl hat bei Versicherungen eine große praktische Bedeutung. Es erlaubt eine ungefähre Vorhersage über den künftigen Schadensverlauf. Je größer die Zahl der versicherten Personen, Güter und Sachwerte, die von der gleichen Gefahr bedroht sind, desto geringer ist der Einfluss des Zufalls. Das



Simulation SGGZ für $N = 200$ Würfelwürfe

Gesetz der großen Zahl kann aber nichts darüber aussagen, wer im Einzelnen von einem Schaden getroffen wird.

Beispiel: Würfeln: Seien $X_1, \dots, X_n, \dots \sim Unid(6)$ unabhängige ZG; man erhält $E(X_n) = \frac{1+2+3+4+5+6}{6} = 3.5$ für alle $n \geq 1$. Laut Satz 16 folgt, dass $(X_n)_n$ erfüllt das **SGGZ**, d.h. $\frac{1}{n}(X_1 + \dots + X_n) \xrightarrow{f.s.} 3.5$

Beispiel: Seien $X_1, \dots, X_n, \dots \sim Unif[-1, 1]$ unabhängige ZG. Zu welchem Wert konvergiert f.s. die Folge

$$Z_n = \frac{1}{n}(X_1^2 + \dots + X_n^2), \quad n \in \mathbb{N}^* \quad ?$$

Lsg.: Wir wenden Satz 16 für die Folge von unabhängigen ZG $(X_n^2)_n \Rightarrow Z_n \xrightarrow{f.s.} E(X_1^2)$. Wir berechnen

$$E(X_1^2) = \int_{-1}^1 t^2 \frac{1}{1 - (-1)} dt = \frac{1}{2} \cdot \frac{t^3}{3} \Big|_{-1}^1 = \frac{1}{3}.$$

$$\Rightarrow Z_n \xrightarrow{f.s.} \frac{1}{3}.$$



Relative und absolute Häufigkeit (siehe Def.1): Sei A ein zufälliges Ereignis das in einem Experiment auftaucht; man wiederholt das Experiment n mal (unter denselben gegebenen Bedingungen); man bezeichnet mit $k_n(A)$ wie oft das Ereignis A auftaucht; die **relative Häufigkeit** des Ereignisses A ist die Zahl $h_n(A) = \frac{k_n(A)}{n}$; die **absolute Häufigkeit** des Ereignisses A ist die Zahl $k_n(A)$.

Experiment: Man wirft n mal eine Münze; A : man erhält Zahl

n Anzahl Durchführungen Exp.	absolute Häufigkeit $k_n(A)$	relative Häufigkeit $h_n(A)$
100	48	0.48
1000	497	0.497
10000	5005	0.5005

Es gilt: $h_n(A) \approx \frac{1}{2}$ wenn n hinreichend groß ist! (siehe Satz 17)

Satz 17. Sei A ein zufälliges Ereignis das in einem Experiment auftaucht; man wiederholt das Experiment n mal (unter denselben gegebenen Bedingungen und unabhängig voneinander).

Das Gesetz der großen Zahlen: je öfter man das Zufallsexperiment durchführt (also je größer n), desto besser approximiert die relative Häufigkeit $h_n(A)$ des Ereignisses A seine echte Wahrscheinlichkeit $P(A)$:

$$P\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} h_n(A) = P(A)\right\}\right) = 1,$$

d.h. $h_n(A) \xrightarrow{f.s.} P(A)$.

In der Praxis: $h_n(A) \approx P(A)$, wenn n hinreichend groß ist!

Beweis für Satz 17: Man wendet Satz 16 an für die folgende Folge von unabhängigen Ereignissen $(X_n)_n$

$$X_n = \begin{cases} 1, & \text{wenn } A \text{ in der } n\text{-ten Durchführung des Experimentes auftaucht} \\ 0, & \text{wenn } \bar{A} \text{ in der } n\text{-ten Durchführung des Experimentes auftaucht} \end{cases}$$

$$\Rightarrow X_n \sim \begin{pmatrix} 0 & 1 \\ 1 - P(A) & P(A) \end{pmatrix} \Rightarrow X_n \sim \text{Bernoulli}(P(A))$$

$$\Rightarrow E(X_n) = 0 \cdot (1 - P(A)) + 1 \cdot P(A) = P(A) \quad \forall n \in \mathbb{N}^*.$$

$$\text{Satz 16} \Rightarrow \frac{1}{n}(X_1 + \dots + X_n) \xrightarrow{f.s.} P(A).$$

Zusätzlich, gilt $\frac{1}{n}(X_1 + \dots + X_n) = h_n(A)$ (relative Häufigkeit von A), also $h_n(A) \xrightarrow{f.s.} P(A)$. □

Statistik

Statistik = wissenschaftliche Disziplin, deren Gegenstand die Entwicklung und Anwendung formaler Methoden zur Gewinnung, Beschreibung, Analyse und Beurteilung von Daten (Beobachtungen) ist.

Aufgaben und Ziele der Statistik

► Design von Experimenten:

Wie sollen die Daten gewonnen werden?

► Beschreibende (deskriptive) Statistik:

Wie sollen große Datensätze beschrieben werden, um die Gesetzmäßigkeiten und Strukturen in ihnen entdecken zu können?

► Schließende Statistik:

Welche Schlußfolgerungen kann man aus den Daten ziehen?

Arbeitsweise in der Statistik

- Datenerhebung (Beobachtung, Befragung, Experiment)
- Visualisierung und beschreibende Datenanalyse: graphische Präsentation, Zusammenfassung (Darstellung der Daten in einer Tabelle oder Darstellung mit Hilfe von Grafiken)
- Explorative Datenanalyse (man sucht nach Gesetzmäßigkeiten in den Daten)
- Modellierung der Daten
- Modellanpassung (Schätzung von Modellparametern)
- Modellvalidierung (Wie gut war die Modellanpassung?)

Anwendungen der Statistik:

- in politischen Umfragen: z.B. Befragung zur Beliebtheit von Politikern oder einer Partei
- in der Analyse von Finanzmarktdaten: z.B. Analyse von Aktien-, Zinskursen
- in klinischen und epidemiologischen Studien (Medizin und Pharmazie)
- in der Technik: z.B. die Lebensdaueranalyse oder die Zuverlässigkeit von elektronischen Systemen

Grundbegriffe der Statistik

- **statistische Einheiten** = Objekte an denen interessierende Größen erfaßt werden
z.B. Bevölkerung einer Stadt; Schüler einer bestimmten Schule; Patienten einer Klinik
- **Grundgesamtheit** (Population) = Menge aller statistischen Einheiten über die man Aussagen erhalten will
z.B. wahlberechtigte Bevölkerung einer Stadt; Schüler der 10. Klassen einer bestimmten Schule; Patienten einer bestimmten Station einer Klinik
- **statistisches Merkmal** = Eigenschaft einer statistischen Einheit für die man sich bei einer statistischen Untersuchung interessiert
z.B. Alter; Geschlecht; Wert BMW Aktie 24.11.2016, 17 Uhr;
- **Merkmalsausprägung** = konkreter Wert des Merkmals
z.B. 25 Jahre; weiblich; 82.5 Euro
- **Stichprobe** = tatsächlich untersuchte Teilmenge der Grundgesamtheit

Grundbegriffe der Statistik

Ein Merkmal einer Grundgesamtheit ist eine Zufallsgröße X . Ziel der Statistik ist das Gesetz (die Verteilung) von X zu finden (zu schätzen), anhand der statistischen Daten.

Die Verteilung von X kann

1) vollständig spezifiziert sein

z.B. $X \sim \text{Exp}(3)$, $X \sim \text{Bino}(10, 0.3)$, $X \sim N(0, 1)$, $X \sim \text{Unif}[2, 5]$

2) spezifiziert sein, aber von einem oder mehreren unbekannten Parametern abhängen

z.B. $X \sim \text{Exp}(\lambda)$, $X \sim \text{Bino}(10, p)$, $X \sim N(\mu, \sigma^2)$, $X \sim \text{Unif}[a, b]$

3) unbekannt sein, $X \sim ???$

► in den Fällen 2) und 3) werden der unbekannte Parameter oder die unbekannte Verteilung

↪ geschätzt → Schätztheorie

↪ getestet → Testen von statistischen Hypothesen

Schätztheorie

Sei X die zufällige Variable, welche das untersuchte statistische Merkmal darstellt. Seien x_1, \dots, x_n *statistische Daten* (Beobachtungen, Stichprobenwerte) für das Merkmal X , die anhand einer Stichprobe erhalten wurden. Die Daten x_1, \dots, x_n können als Werte von n zufälligen Variablen X_1, \dots, X_n betrachtet werden, welche unabhängige ZG sind mit derselben Verteilung wie X ; X_1, \dots, X_n heißen *Stichprobenvariablen*.

Def. 22. Seien X_1, \dots, X_n Stichprobenvariablen und $g: \mathbb{R}^n \rightarrow \mathbb{R}$, so dass $g(X_1, \dots, X_n)$ eine ZG ist.

► $g(X_1, \dots, X_n)$ heißt *Schätzfunktion* (ist eine ZG)

► $g(x_1, \dots, x_n)$ heißt *Schätzwert* (ist ein Wert)

Beispiele von Schätzfunktionen

► *Stichprobenmittel* (empirischer Mittelwert)

$$\bar{X}_n = g(X_1, \dots, X_n) = \frac{1}{n} (X_1 + \dots + X_n)$$

► *Wert des Stichprobenmittels*

$$\bar{x}_n = g(x_1, \dots, x_n) = \frac{1}{n} (x_1 + \dots + x_n)$$

► *Stichprobenvarianz* (empirische Varianz)

$$S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

► *Wert der Stichprobenvarianz*

$$s_n^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x}_n)^2$$

► *empirische Standardabweichung*

$$S_n = \left(\frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2 \right)^{\frac{1}{2}}$$

► *Wert der empirischen Standardabweichung*

$$s_n = \left(\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x}_n)^2 \right)^{\frac{1}{2}}$$

► empirisches zentriertes Moment zweiter Ordnung

$$M_n = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2 = \frac{n-1}{n} S_n^2$$

► Wert des empirischen zentrierten Moments zweiter Ordnung

$$m_n = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x}_n)^2 = \frac{n-1}{n} s_n^2$$

► empirische Verteilungsfunktion $\mathcal{F}_n : \mathbb{R} \times \Omega \rightarrow [0, 1]$

$$\mathcal{F}_n(x, \omega) = \frac{\#\{i \in \{1, \dots, n\} : X_i(\omega) \leq x\}}{n}, x \in \mathbb{R}$$

► Wert der empirischen Verteilungsfunktion $\mathcal{F}_n : \mathbb{R} \rightarrow [0, 1]$

$$\mathcal{F}_n(x) = \frac{\#\{i \in \{1, \dots, n\} : x_i \leq x\}}{n}, x \in \mathbb{R}$$

Sei X eine ZG (das untersuchte statistische Merkmal); die Verteilung von X hängt von einem unbekannten Parameter θ oder mehreren unbekannten Parametern $\theta_1, \dots, \theta_r$ ab; man möchte diese Parameter schätzen!

► Seien $x_1, \dots, x_n \hookrightarrow$ statistische Daten (Beobachtungen, Stichprobenwerte) für das Merkmal X , die anhand einer Stichprobe erhalten wurden.

► Seien X_1, \dots, X_n sind Stichprobenvariablen, d.h. sie sind unabhängige ZG mit derselben Verteilung wie X .

► $g(X_1, \dots, X_n) \hookrightarrow$ Schätzfunktion (ist eine ZG; ist eine Funktion die von den Stichprobenvariablen abhängt)

► $g(x_1, \dots, x_n) \hookrightarrow$ Schätzwert (ist ein Wert; Wert der Schätzfunktion).

Def. 23. Die Schätzfunktion $g(X_1, \dots, X_n)$ ist **erwartungstreu** für den unbekannten Parameter θ , wenn

$$E(g(X_1, \dots, X_n)) = \theta.$$

➡ Ein Schätzer ist dann erwartungstreu, wenn sein Erwartungswert gleich dem zu schätzenden Parameter ist.

Def. 24. Die Schätzfunktion $g(X_1, \dots, X_n)$ ist **konsistent** für den unbekannten Parameter θ , wenn

$$g(X_1, \dots, X_n) \xrightarrow{f.s.} \theta.$$

Eigenschaften der Schätzfunktionen:

► das Stichprobenmittel (empirischer Mittelwert)

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

ist ein **erwartungstreuer** und **konsistenter** Schätzer für den theoretische Erwartungswert $E(X)$ des Merkmals X

➡ in Simulationen $E(X) \approx \bar{x}_n$; in Python: `numpy.mean(d)`, wobei `d = [x1, ..., xn]` der Vektor der statistischen Daten ist;

► die Stichprobenvarianz (empirische Varianz)

$$S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

ist ein **erwartungstreuer** und **konsistenter** Schätzer für die theoretische Varianz $V(X)$ des Merkmals X
 ➡ in Simulationen $V(X) \approx s_n^2$; in Octave: `numpy.var(d, ddof=1)`, wobei d der Vektor der statistischen Daten ist;

► das empirische zentrierte Moment zweiter Ordnung

$$M_n = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

ist **kein erwartungstreuer**, aber ein **konsistenter** Schätzer für die theoretische Varianz $V(X)$ des Merkmals X

➡ man kann zeigen, dass $M_n = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2$;

➡ in Simulationen $V(X) \approx m_n$; in Octave: `numpy.var(d)` (auch `numpy.var(d, ddof=0)`), wobei d der Vektor der statistischen Daten ist;

► die empirische Standardabweichung

$$S_n = \left(\frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2 \right)^{\frac{1}{2}}$$

ist **kein erwartungstreuer**, aber ein **konsistenter** Schätzer für die theoretische Standardabweichung $Std(X)$ des Merkmals X

➡ in Simulationen $Std(X) \approx s_n$; in Octave: `numpy.std(d, ddof=1)`, wobei d der Vektor der statistischen Daten ist;

► die empirische Verteilungsfunktion $\mathcal{F}_n : \mathbb{R} \times \Omega \rightarrow \mathbb{R}$

$$\mathcal{F}_n(x, \omega) = \frac{\#\{i \in \{1, \dots, n\} : X_i(\omega) \leq x\}}{n}, x \in \mathbb{R}$$

ist ein **erwartungstreuer** und **konsistenter** Schätzer für $F(x)$ (wobei $F(x) = P(X \leq x)$, $x \in \mathbb{R}$, ist die theoretische Verteilungsfunktion von X)

Methoden zur Berechnung von Schätzfunktionen

Wir diskutieren Methoden, die ein systematisches Vorgehen für die Wahl einer Schätzfunktion ermöglichen, um den unbekannten Parameter θ auf geeignete Weise zu schätzen.

► Momenten-Methode

► Maximum-Likelihood-Methode

Momenten-Methode für das Schätzen unbekannter Parameter $\theta = (\theta_1, \dots, \theta_r)$ der Verteilung des beobachteten Merkmals X

$X \sim Exp(\lambda)$ unbekannter Parameter: $\theta = \lambda$ [$r=1$]

$X \sim N(\mu, \sigma^2)$ unbekannte Parameter: $(\theta_1, \theta_2) = (\mu, \sigma^2)$ [$r=2$]

► x_1, \dots, x_n statistische Daten;

► X_1, \dots, X_n seien Stichprobenvariablen für das Merkmal X

Man löst das System

$$\begin{cases} E(X^k) = \frac{1}{n} \sum_{i=1}^n x_i^k \\ k \in \{1, \dots, r\} \end{cases}$$

mit Unbekannten $\theta_1, \dots, \theta_r$.

Die Lösung des Systems $(\hat{\theta}_1, \dots, \hat{\theta}_r)$ ergeben die Schätzwerte für die unbekannten Parameter $(\theta_1, \dots, \theta_r)$.

Beispiel 1: Mit Hilfe der Momenten-Methode schätze man den unbekannten Parameter $\theta = a$ für $X \sim Unif[0, a]$ (gleichmäßige Verteilung auf $[0, a]$); die statistischen Daten sind: 0.1, 0.3, 0.9, 0.49, 0.12, 0.31, 0.98, 0.73, 0.13, 0.62;

Lsg.: Seien $n = 10$ und X_1, \dots, X_n Stichprobenvariablen. Für: $r = 1$, berechnet man $E(X) = \frac{a}{2}$ und $\bar{x}_n = 0.468$. Man löst

$$E(X) = \frac{1}{n} \sum_{i=1}^n x_i \implies \frac{a}{2} = \frac{1}{n} \sum_{i=1}^n x_i$$

Der Schätzwert für den unbekannten Parameter a ist

$$\hat{a}(x_1, \dots, x_n) = \frac{2}{n} \sum_{i=1}^n x_i = 0.936,$$

und die Schätzfunktion ist

$$\hat{a}(X_1, \dots, X_n) = \frac{2}{n} \sum_{i=1}^n X_i.$$

Der unbekannte Parameter a wird durch den Wert 0.936 geschätzt.

► Ist $\hat{a}(X_1, \dots, X_n)$ ein erwartungstreuer Schätzer für den Parameter a ?

Lsg.: Ja. Man zeigt $E(\hat{a}(X_1, \dots, X_n)) = a$. ▲

Beispiel 2: Mit Hilfe der Momenten-Methode schätze man die unbekannten Parameter $\theta_1 := \mu$ und $\theta_2 = \sigma^2$ für $X \sim N(\mu, \sigma^2)$; gegeben sind die statistischen Daten:

$$0.831, 0.71, -0.2, -0.04, 2.08, -1.2, 0.448, -0.18, -0.27, -0.55.$$

Lsg.: Seien $n = 10$ und X_1, \dots, X_n Stichprobenvariablen. Man hat $r = 2$, und berechnet $E(X) = \mu$, $E(X^2) = V(X) + E^2(X) = \sigma^2 + \mu^2$ (siehe Beispiel auf Seite 28). Man löst

$$\begin{cases} \mu &= \frac{1}{n} \sum_{i=1}^n x_i \\ \sigma^2 + \mu^2 &= \frac{1}{n} \sum_{i=1}^n x_i^2 \end{cases} \implies \text{hat die Lösung} \begin{cases} \hat{\mu}(x_1, \dots, x_n) &= \frac{1}{n} \sum_{i=1}^n x_i \\ \hat{\sigma}^2(x_1, \dots, x_n) &= \frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2. \end{cases}$$

Die Schätzfunktionen sind

$$\hat{\mu}(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n \text{ (Stichprobenmittel),}$$

$$\hat{\sigma}^2(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2 = M_n \text{ (empirisches zentrales Moment zweiter Ordnung, siehe Seite 37).}$$

Die Schätzwerte sind

$$\hat{\mu}(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}_n = 0.1629,$$

$$\hat{\sigma}^2(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}_n^2 = m_n = 0.7346.$$



Maximum-Likelihood-Methode zur Berechnung von Schätzfunktionen

↪ die Verteilung des beobachteten Merkmals X hängt von einem unbekannten Parameter θ oder von mehreren unbekannten Parametern $(\theta_1, \dots, \theta_r)$ ab

↪ man konstruiert den Maximum-Likelihood-Schätzer für den unbekannten Parameter θ , bzw. für die unbekannten Parameter $(\theta_1, \dots, \theta_r)$

z.B.:

$X \sim \text{Exp}(\lambda)$ unbekannter Parameter: $\theta = \lambda$ [$r=1$]

$X \sim N(\mu, \sigma^2)$ unbekannte Parameter: $(\theta_1, \theta_2) = (\mu, \sigma^2)$ [$r=2$]

Maximum-Likelihood-Methode (MLM) für das Schätzen unbekannter Parameter der Verteilung des beobachteten Merkmals X :

Seien:

- ▶ x_1, \dots, x_n statistische Daten;
- ▶ X_1, \dots, X_n seien Stichprobenvariablen für das Merkmal X
- ▶ θ unbekannter Parameter

$$\text{▶ } L(x_1, \dots, x_n; \theta) = \begin{cases} P(X = x_1) \cdot \dots \cdot P(X = x_n), & X \text{ diskret} \\ f_X(x_1) \cdot \dots \cdot f_X(x_n), & X \text{ stetig mit Dichtefunktion } f_X \end{cases}$$

heißt **Likelihood-Funktion** des Parameters θ für die statistischen Daten x_1, \dots, x_n .

▶ Man wählt für die statistischen Daten x_1, \dots, x_n als Parameterschätzung $\hat{\theta}$ denjenigen Parameter als Schätzwert für den die Likelihood-Funktion maximal ist

$$(4) \quad L(x_1, \dots, x_n; \hat{\theta}) = \max_{\theta} L(x_1, \dots, x_n; \theta),$$

$\hat{\theta}$ ist *globaler Maximumpunkt* für die Likelihood-Funktion L .

In der Praxis: man löst $\frac{\partial L}{\partial \theta} = 0$ und zeigt $\frac{\partial^2 L}{\partial \theta^2} < 0$. Oft ist es leichter die äquivalente Form zu lösen

$$\frac{\partial \ln L}{\partial \theta} = 0 \text{ und zeigt } \frac{\partial^2 \ln L}{\partial \theta^2} < 0.$$

Es gibt Beispiele, in welchen (4) durch andere Methoden gelöst wird; z.B. wenn $\frac{\partial L}{\partial \theta} = 0$ keine Lösung hat (oder äquivalent $\frac{\partial \ln L}{\partial \theta} = 0$ hat keine Lösung).

Erinnerung: für $a, b > 0$, gelten die Eigenschaften:

$$\ln(a \cdot b) = \ln a + \ln b, \ln(a^b) = b \cdot \ln a, \ln\left(\frac{a}{b}\right) = \ln a - \ln b.$$

Beispiel: Man schätze den Parameter $\theta := p \in (0, 1)$ der Bernoulli Verteilung mit der MLM!

Lsg.: $X \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$, mit den statistischen Daten: 0,1,1,0,0,0,1,0,1,0.

$\Rightarrow n = 10, x_1 = 0, x_2 = 1, x_3 = 1, x_4 = 0 \dots;$

Es gilt: $P(X = x) = p^x(1 - p)^{1-x}, x \in \{0, 1\}$

$$\Rightarrow L(x_1, \dots, x_n; p) = P(X = x_1) \cdot \dots \cdot P(X = x_n) = p^{x_1 + \dots + x_n} (1 - p)^{n - (x_1 + \dots + x_n)}$$

$$\Rightarrow \ln L(x_1, \dots, x_n; p) = (x_1 + \dots + x_n) \ln(p) + (n - (x_1 + \dots + x_n)) \ln(1 - p)$$

$$\frac{\partial \ln L}{\partial p} = 0 \Rightarrow p = \frac{1}{n}(x_1 + \dots + x_n)$$

Man kann zeigen:

$$\frac{\partial^2 \ln L}{\partial p^2} < 0.$$

Der ML-Schätzer ist

$$\hat{p}(X_1, \dots, X_n) = \frac{1}{n}(X_1 + \dots + X_n) = \bar{X}_n$$

und hat den Schätzwert

$$\hat{p}(x_1, \dots, x_n) = \frac{1}{n}(x_1 + \dots + x_n) = \bar{x}_n = \frac{4}{10} = 0.4$$

Übung: Ist \hat{p} erwartungstreu und konsistenter Schätzer?

Bemerkung: Wenn die Verteilung von r unbekannten Parametern $(\theta_1, \dots, \theta_r)$ abhängt: eine Methode ist das folgende System zu lösen

$$\frac{\partial \ln L}{\partial \theta_j} = 0, j = \overline{1, r} \text{ und man zeigt dass die Matrix } \left(\frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j} \right)_{1 \leq i, j \leq r}$$

negativ definiert ist.

Hinweis: Eine Matrix M ist negativ definiert, wenn $y^t M y < 0$ für alle $y \in \mathbb{R}^r \setminus \{0_r\}$. ▲

Def. 25. Sei $\alpha \in (0, 1)$. Das **Quantil der Ordnung** α für die Verteilung des beobachteten Merkmals X ist der Wert $z_\alpha \in \mathbb{R}$ für welchen gilt

$$P(X < z_\alpha) \leq \alpha \leq P(X \leq z_\alpha).$$

$z_{0.5}$ heißt **Median**.

• Falls X eine stetige ZG ist, dann ist z_α Quantil der Ordnung $\alpha \iff P(X \leq z_\alpha) = \alpha \iff F_X(z_\alpha) = \alpha$.

► ist F_X bijektiv, dann gilt $z_\alpha = F_X^{-1}(\alpha)$

• $\alpha \cdot 100\%$ der Werte von X sind kleiner oder gleich mit z_α .

Beispiel: 1) Sei $X \sim \begin{pmatrix} 1 & 3 & 5 & 7 \\ 0.2 & 0.35 & 0.35 & 0.1 \end{pmatrix}$ diskrete ZG

$$\Rightarrow P(X < 3) = 0.2 \leq 0.5 \leq P(X \leq 3) = 0.2 + 0.35 = 0.55 \Rightarrow z_{0.5} = 3 \text{ ist der Median.}$$

2) Für $X \sim \text{Exp}(\lambda) \Rightarrow F_X(x) = 1 - e^{-\lambda x}$ für $x \geq 0$

$$\Rightarrow \text{Quantil der } \text{Exp}(\lambda) \text{ Verteilung } z_\alpha = F_X^{-1}(\alpha) = -\frac{\ln(1-\alpha)}{\lambda} \text{ für } \alpha \in (0, 1). \quad \diamond$$

Verteilungen die für statistische Tests benutzt werden und ihre Quantile:

➡ `from scipy.stats import norm, t, chi2`

• Normalverteilung $N(0, 1)$

\hookrightarrow Verteilungsfunktion $F_{N(0,1)}(x) = \text{norm.cdf}(x, 0, 1)$;
 Quantil $z_\alpha = \text{norm.ppf}(\alpha, 0, 1)$, d.h. $F_{N(0,1)}(z_\alpha) = \alpha$;
 • Student Verteilung mit n Freiheitsgraden $T(n)$
 \hookrightarrow Verteilungsfunktion $F_{T(n)}(x) = \text{t.cdf}(x, n)$;
 Quantil $t_\alpha = \text{t.ppf}(\alpha, n)$, d.h. $F_{T(n)}(t_\alpha) = \alpha$;
 • χ^2 -Verteilung mit n Freiheitsgraden $\chi^2(n)$
 \hookrightarrow Verteilungsfunktion $F_{\chi^2(n)}(x) = \text{chi2.cdf}(x, n)$;
 Quantil $c_\alpha = \text{chi2.ppf}(\alpha, n)$, d.h. $F_{\chi^2(n)}(c_\alpha) = \alpha$.
 \hookrightarrow ppf - percent point function (inverse of cdf)

```

#Beispiel Berechnung von Quantilen
from scipy.stats import norm, t, chi2
alfa=0.01
z1=norm.ppf(alfa,0,1)
z2=norm.ppf(1-alfa,0,1)
print("Quantile der N(0,1) Verteilung:", "z_alfa=",z1,"z_{1-alfa}=",z2)
n=10
alfa=0.05
t1=t.ppf(alfa,n)
t2=t.ppf(1-alfa,n)
print(f"Quantile der Student Verteilung T({n}):", "t_alfa=",t1,"t_{1-alfa}=",t2)
c1=chi2.ppf(alfa,n)
c2=chi2.ppf(1-alfa,n)
print(f"Quantile der Chi-Quadrat({n}) Verteilung:", "c_alfa=",c1,"c_{1-alfa}=",c2)
  
```

Beispiel:

$\alpha = 0.01$: $z_\alpha = \text{norm.ppf}(0.01, 0, 1) \approx -2.3263$, $z_{1-\alpha} = \text{norm.ppf}(1 - 0.01, 0, 1) \approx 2.3263$,
 $\alpha = 0.05$: $t_\alpha = \text{t.ppf}(0.05, 10) \approx -1.8125$, $t_{1-\alpha} = \text{t.ppf}(1 - 0.05, 10) \approx 1.8125$,
 $\alpha = 0.05$: $c_\alpha = \text{chi2.ppf}(0.05, 10) \approx 3.9403$, $c_{1-\alpha} = \text{chi2.ppf}(1 - 0.05, 10) \approx 18.307$.

Satz 18. Für die Quantile der *standard Normalverteilung* $N(0, 1)$ gilt $z_\alpha = -z_{1-\alpha}$ für alle $\alpha \in (0, 1)$.
 Für die Quantile der *Studentverteilung* $T(n)$ ($n \in \mathbb{N}^*$) gilt $t_\alpha = -t_{1-\alpha}$ für alle $\alpha \in (0, 1)$.

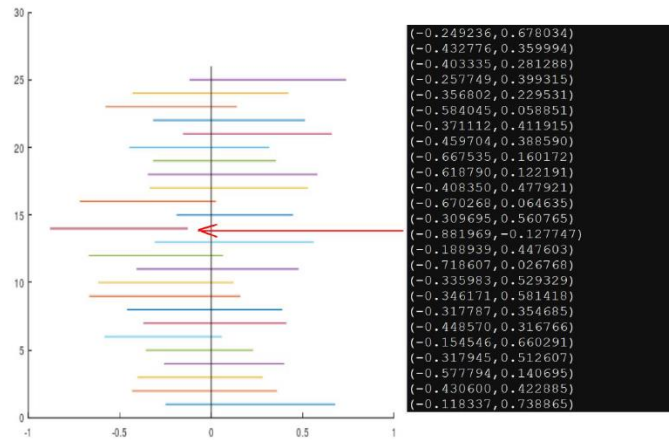
Konfidenzintervalle (Vertrauensintervalle)

Konfidenzintervalle basieren auf Stichproben und decken zu einer bestimmten Wahrscheinlichkeit den wahren (aber unbekannten) Parameter ab.

Im Gegensatz zum geschätzten Parameter, der anhand der statistischen Daten einer Stichprobe berechnet wird, kann der wahre Wert eines unbekannten Parameters selten exakt bestimmt werden. **Konfidenzintervalle bieten aber die Möglichkeit, mit einer gewissen Erfolgswahrscheinlichkeit diesen genauer lokalisieren zu können.**
 Z.B.: man forscht das statistische Merkmal X = Wartezeit in Minuten an einem Schalter bei einer bestimmten Bank; der unbekannte Parameter ist $\theta = E(X)$. Man hat anhand einer Stichprobe den Wert für das Stichprobenmittel $\bar{x}_{200} = 10$ (Minuten) erhalten. Wenn man eine andere Stichprobe wählt, erhält man einen anderen Wert z.B. $\bar{x}_{300} = 9$ Minuten und 10 Sekunden. Kann man ein zufälliges Intervall konstruieren, welches den wahren Wert des unbekannten Parameters θ enthält mit einer vorgegebenen Wahrscheinlichkeit (z.B.) 0.95? Anhand konkreter statistischer Daten wird dieses ein numerisches Intervall.

• Seien:

- ▶ x_1, \dots, x_n statistische Daten;
- ▶ X_1, \dots, X_n seien Stichprobenvariablen für das Merkmal X
- ▶ θ unbekannter Parameter



Simulation: aus 25 Konfidenzintervallen, ein Intervall enthält nicht den *wahren Wert* 0; der unbekannte Parameter ist hier θ =der theoretische Erwartungswert; die statistischen Daten wurden mit `normr.rvs(0, 1)` generiert

• Vorgegeben ist entweder $\alpha \in (0, 1)$, das *Signifikanzniveau*, oder $1 - \alpha$, das *Konfidenzniveau* (die *Vertrauenswahrscheinlichkeit* oder *Überdeckungswahrscheinlichkeit*).

Man sucht zwei Schätzer $g_1(X_1, \dots, X_n)$ und $g_2(X_1, \dots, X_n)$ so dass

$$P\left(g_1(X_1, \dots, X_n) < \theta < g_2(X_1, \dots, X_n)\right) = 1 - \alpha$$

- $\left(g_1(X_1, \dots, X_n), g_2(X_1, \dots, X_n)\right)$ heißt **zweiseitiges Konfidenzintervall** für den unbekannten Parameter θ
- $\left(g_1(x_1, \dots, x_n), g_2(x_1, \dots, x_n)\right)$ ist der **Wert zweiseitigen Konfidenzintervalls**
- $g_1(X_1, \dots, X_n)$ ist die untere Grenze des Konfidenzintervalls, sein Wert ist $g_1(x_1, \dots, x_n)$
- $g_2(X_1, \dots, X_n)$ ist die obere Grenze des Konfidenzintervalls, sein Wert ist $g_2(x_1, \dots, x_n)$
- $1 - \alpha$ (das Konfidenzniveau) ist die Wahrscheinlichkeit, dass der unbekannte Parameter θ im Intervall $\left(g_1(X_1, \dots, X_n), g_2(X_1, \dots, X_n)\right)$ ist
- es gibt auch **einseitige Konfidenzintervalle** für den unbekannten Parameter θ : $\left(-\infty, g_3(X_1, \dots, X_n)\right)$, $\left(g_4(X_1, \dots, X_n), \infty\right)$, wobei g_3 und g_4 so bestimmt werden, dass

$$P\left(\theta < g_3(X_1, \dots, X_n)\right) = 1 - \alpha, \text{ bzw. } P\left(g_4(X_1, \dots, X_n) < \theta\right) = 1 - \alpha$$

- $\left(-\infty, g_3(x_1, \dots, x_n)\right)$ $\left(g_4(x_1, \dots, x_n), \infty\right)$ sind die Werte der einseitigen Konfidenzintervalle
- $1 - \alpha$ (das Konfidenzniveau) ist die Wahrscheinlichkeit, dass der unbekannte Parameter θ im Intervall $\left(-\infty, g_3(X_1, \dots, X_n)\right)$ ist, bzw. im Intervall $\left(g_4(X_1, \dots, X_n), \infty\right)$ ist.

Bemerkung: Das Konfidenzintervall ist ein zufälliges Intervall, seine konkrete Berechnung hängt von der Stichprobe ab. Die Interpretation von $1 - \alpha$ (das Konfidenzniveau) ist: wenn wir mehrere Stichproben (welche viele statistische Daten enthalten) durchführen und für jede Stichprobe das zugehörige Konfidenzintervall berechnen (mit demselben Konfidenzniveau $1 - \alpha$), dann enthalten *im Mittel* $(1 - \alpha) \cdot 100\%$ dieser berechneten Intervalle den wahren Wert des unbekannten des Parameters θ (siehe das obige Bild für $\alpha = 0.05$). Das Konfidenzintervall gibt den Bereich an, der bei vielen Wiederholungen eines Zufallsexperiments mit einer gewissen Wahrscheinlichkeit

(dem Konfidenzniveau) die wahre Lage des Parameters einschließt.

Verwendung von Konfidenzintervallen in der Künstlichen Intelligenz (KI):

► **Modellevaluation:** Konfidenzintervalle können verwendet werden, um die Leistung eines KI-Modells zu bewerten. Zum Beispiel, bei der Bewertung der Genauigkeit eines Klassifikationsmodells gibt ein Konfidenzintervall einen Bereich an, innerhalb dessen die wahre Genauigkeit des Modells mit einer bestimmten Sicherheit (z.B. 95%) liegt.

► **Parameterschätzung:** Bei der Schätzung von Parametern eines Modells können Konfidenzintervalle den Bereich angeben, der wahrscheinlich den wahren Parameterwert enthält. Dies ist entscheidend, um die Zuverlässigkeit der Modellparameter zu verstehen.

► **Vergleich von Modellen:** Konfidenzintervalle können verwendet werden, um die Leistung verschiedener Modelle zu vergleichen. Wenn sich die Konfidenzintervalle der Leistungskennzahlen (wie Genauigkeit oder Fehlerrate) zweier Modelle überschneiden, weist dies darauf hin, dass es möglicherweise keinen signifikanten Unterschied in ihrer Leistung gibt.

Wiederholung (Bezeichnungen)

Stichprobenvariablen für das Merkmal X X_1, \dots, X_n sind unabhängige ZG, haben dieselbe Verteilung wie X	statistische Daten für das Merkmal X x_1, \dots, x_n sind (numerische) Werte der ZG X_1, \dots, X_n
Schätzfunktion	Wert der Schätzfunktion
Stichprobenmittel $\bar{X}_n = \frac{1}{n} (X_1 + \dots + X_n)$	Wert des Stichprobenmittels $\bar{x}_n = \frac{1}{n} (x_1 + \dots + x_n)$
Stichprobenvarianz $S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$	Wert der Stichprobenvarianz $s_n^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x}_n)^2$
empirische Standardabweichung $S_n = \left(\frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2 \right)^{\frac{1}{2}}$	Wert der empirischen Standardabweichung $s_n = \left(\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x}_n)^2 \right)^{\frac{1}{2}}$

Satz 19. Seien X_1, \dots, X_n Stichprobenvariablen für das Merkmal $X \sim N(\mu, \sigma^2)$; für das Stichprobenmittel gilt $\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$.

Zur Erinnerung: $X \sim N(\mu, \sigma^2) \implies E(X) = \mu, V(X) = \sigma^2$ (siehe die Rechnungen auf Seite 28).

Beispiel: Seien $X_1, \dots, X_{100} \sim N(0, 1)$ unabhängige ZG. Man berechne $P(|X_1 + \dots + X_{100}| \leq 10)$.

Hinweis: `norm.cdf` in Python.

Lsg.: Wir benutzen Satz 19 ($\mu = 0, \sigma = 1, n = 100$) $\implies \frac{\bar{X}_{100} - 0}{\frac{1}{\sqrt{100}}} \sim N(0, 1)$

$$\begin{aligned}
 P(|X_1 + \dots + X_{100}| \leq 10) &= P(-10 \leq X_1 + \dots + X_{100} \leq 10) = P\left(-1 \leq \frac{\bar{X}_{100} - 0}{\frac{1}{\sqrt{100}}} \leq 1\right) \\
 &= F_{N(0,1)}(1) - F_{N(0,1)}(-1) = \text{norm.cdf}(1, 0, 1) - \text{norm.cdf}(-1, 0, 1) \approx 0.6827.
 \end{aligned}$$



Satz 20. (Zentraler Grenzwertsatz) Seien X_1, \dots, X_n Stichprobenvariablen für das Merkmal X , mit $\mu = E(X) = E(X_k)$ und $\sigma^2 = V(X) = V(X_k) > 0$ für alle $k \in \{1, \dots, n\}$. Es gilt

$$\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1) \text{ wenn } n \text{ hinreichend groß ist (} n > 30 \text{)}.$$

➡ Der zentrale Grenzwertsatz besagt, dass das Stichprobenmittel als Zufallsvariable (d.h. das arithmetische Mittel identisch verteilter Zufallsvariablen einer beliebigen Verteilung mit endlicher Varianz) näherungsweise standard normalverteilt ist, insbesondere wegen der additiven Überlagerung vieler unabhängiger Zufallseinflüsse.

➡ Praktisch, wie wird dieses Resultat benutzt?

$P\left(a < \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} < b\right) \approx F_{N(0,1)}(b) - F_{N(0,1)}(a) = \text{norm.cdf}(b, 0, 1) - \text{norm.cdf}(a, 0, 1)$ für n hinreichend groß ($n > 30$), für $\forall a, b \in \mathbb{R}, a < b$, wobei $F_{N(0,1)}$ die Verteilungsfunktion der $N(0, 1)$ ist.

Beispiel: Seien $(X_n)_{1 \leq n \leq 100}$ Stichprobenvariablen für das Merkmal $X \sim \text{Bernoulli}(0.5)$; man schätze anhand des Zentralen Grenzwertsatzes (Satz 20)

$$P(0.35 < \bar{X}_{100} < 0.65).$$

Lsg.: Man berechnet $\mu = E(X_n) = 0.5$, $\sigma = \sqrt{V(X_n)} = 0.5$ und man schreibt

$$P(0.35 < \bar{X}_{100} < 0.65) = P\left(-3 < \frac{\bar{X}_{100} - 0.5}{\frac{0.5}{\sqrt{100}}} < 3\right).$$

Laut Satz 20

$$\Rightarrow P\left(-3 < \frac{\bar{X}_{100} - 0.5}{\frac{0.5}{\sqrt{100}}} < 3\right) \approx \text{norm.cdf}(3, 0, 1) - \text{norm.cdf}(-3, 0, 1) = 0.9973$$

$$\Rightarrow P(\bar{X}_{100} \in (0.35, 0.65)) \approx 0.9973,$$

d.h. für ein Merkmal vom Typ $\text{Bernoulli}(0.5)$ gehört das Stichprobenmittel \bar{X}_{100} mit großer Wahrscheinlichkeit dem Intervall $(0.35, 0.65)$ an. ♣

➡ **Konfidenzintervall für den theoretischen Erwartungswert $\mu = E(X)$ des beobachteten Merkmals X , wenn die Varianz $\sigma^2 = V(X)$ des Merkmals X bekannt ist:**

Beispiel: Ein Mathelehrer hat während mehrerer Jahre die Ergebnisse seiner Schüler bei einem bestimmten Test beobachtet. Die Punktezahl beim Test eines Schülers ist eine ZG $X \in (0, 100)$, mit der Standardabweichung 10. Der Wert des Stichprobenmittels für 144 Schüler ist 68. Sei $\alpha = 0.05$ das Signifikanzniveau; man gebe ein zweiseitiges Konfidenzintervall an für den theoretischen Erwartungswert $E(X)$.

► Gegeben sind: $\alpha \in (0, 1)$, $\sigma^2 = V(X)$, die statistischen Daten x_1, \dots, x_n

► seien X_1, \dots, X_n Stichprobenvariablen für das Merkmal X

► man sucht ein Konfidenzintervall für den unbekannten Parameter $\mu = E(X)$ (der theoretische Erwartungswert), wenn die Varianz $\sigma^2 = V(X)$ des Merkmals X bekannt ist

► falls $X \sim N(\mu, \sigma^2)$ oder $n > 30$ und X hat eine beliebige Verteilung, dann

$$(5) \quad \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

(siehe Satz 19, bzw. Satz 20)

► man berechnet die Quantile der normalen Verteilung $N(0, 1)$:

$$z_{1-\frac{\alpha}{2}} = \text{norm.ppf}(1 - \frac{\alpha}{2}, 0, 1), z_{1-\alpha} = \text{norm.ppf}(1 - \alpha, 0, 1), z_{\alpha} = \text{norm.ppf}(\alpha, 0, 1)$$

• ein *zweiseitiges Konfidenzintervall* für $\mu = E(X)$ ist $\left(\bar{X}_n - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}, \bar{X}_n + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}} \right)$, d.h.

$$P\left(\bar{X}_n - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}} < \mu < \bar{X}_n + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

weil:

$$\begin{aligned} P\left(\bar{X}_n - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}} < \mu < \bar{X}_n + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}\right) &= P\left(-z_{1-\frac{\alpha}{2}} < \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} < z_{1-\frac{\alpha}{2}}\right) \\ &\stackrel{(5)}{=} F_{N(0,1)}(z_{1-\frac{\alpha}{2}}) - F_{N(0,1)}(-z_{1-\frac{\alpha}{2}}) = F_{N(0,1)}(z_{1-\frac{\alpha}{2}}) - F_{N(0,1)}(z_{\frac{\alpha}{2}}) \stackrel{\text{siehe Satz 18}}{=} 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha \end{aligned}$$

• *einseitige Konfidenzintervalle* für $\mu = E(X)$ sind: $\left(-\infty, \bar{X}_n - \frac{\sigma}{\sqrt{n}} \cdot z_{\alpha}\right)$, $\left(\bar{X}_n - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha}, \infty\right)$, d.h.

$$P\left(\mu < \bar{X}_n - \frac{\sigma}{\sqrt{n}} \cdot z_{\alpha}\right) = 1 - \alpha, \quad P\left(\bar{X}_n - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha} < \mu\right) = 1 - \alpha.$$

Konfidenzintervall für den Erwartungswert $\mu = E(X)$, wenn die Varianz $\sigma^2 = V(X)$ des Merkmals X bekannt ist	Ausdruck des Konfidenzintervalls anhand der Daten
zweiseitig	$\left(\bar{x}_n - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}, \bar{x}_n + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}\right)$
einseitig (mit oberer Grenze)	$\left(-\infty, \bar{x}_n - \frac{\sigma}{\sqrt{n}} \cdot z_{\alpha}\right)$
einseitig (mit unterer Grenze)	$\left(\bar{x}_n - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha}, \infty\right).$

Beispiel: Ein Mathelehrer hat während mehrerer Jahre die Ergebnisse seiner Schüler bei einem bestimmten Test beobachtet. Die Punktezahl beim Test eines Schülers ist eine ZG $X \in (0, 100)$, mit der Standardabweichung 10. Der Wert des Stichprobenmittels für 144 Schüler ist 68. Sei $\alpha = 0.05$ das Signifikanzniveau; man gebe ein zweiseitiges Konfidenzintervall an für den theoretischen Erwartungswert $E(X)$.

Lsg.:

$$\left(\bar{x}_n - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}, \bar{x}_n + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}\right)$$

wobei $n = 144, \sigma = 10, \bar{x}_n = 68, \alpha = 0.05, z_{1-\frac{\alpha}{2}} = \text{norm.ppf}(1 - \frac{0.05}{2}, 0, 1) \approx 1.96$. Der Wert des zweiseitigen Konfidenzintervalls ist $(66.367, 69.633)$. ♣

Übung: Wie verändert sich das eingeführte zweiseitige Konfidenzintervall für den Erwartungswert bei:

- Vergrößerung der Standardabweichung in der Grundgesamtheit?
- Vergrößerung der Vertrauenswahrscheinlichkeit?
- Vergrößerung des Stichprobenumfangs?

Satz 21. Seien X_1, \dots, X_n Stichprobenvariablen für das Merkmal X (wobei $X \sim N(\mu, \sigma^2)$ oder $n > 30$). Für das Stichprobenmittel und die empirische Standardabweichung gilt $\frac{\bar{X}_n - \mu}{\frac{S_n}{\sqrt{n}}} \sim T(n-1)$.

► **Konfidenzintervall für den theoretischen Erwartungswert $\mu = E(X)$ des beobachteten Merkmals X , wenn die Varianz des Merkmals X unbekannt ist:**

Beispiel: Der Wert des Stichprobenmittels für das Merkmal X = der Durchmesser von einem Schraubentyp (von einer bestimmten Firma hergestellt) ist $\bar{x}_{100} = 15.5$ mm und die Stichprobenvarianz ist $s_{100}^2 = 0.09$ mm². Sei $1 - \alpha = 0.99$ das Konfidenzniveau; man gebe ein zweiseitiges Konfidenzintervall an für den theoretischen Erwartungswert $E(X)$.

- Gegeben sind: $\alpha \in (0, 1)$, die statistischen Daten x_1, \dots, x_n
- seien X_1, \dots, X_n Stichprobenvariablen für das Merkmal X
- man sucht ein Konfidenzintervall für den unbekannten Parameter $\mu = E(X)$ (der theoretische Erwartungswert), wenn die Varianz des Merkmals X unbekannt ist
- falls $X \sim N(\mu, \sigma^2)$ oder $n > 30$ und X hat eine beliebige Verteilung, dann

$$(6) \quad \frac{\bar{X}_n - \mu}{\frac{S_n}{\sqrt{n}}} \sim T(n-1)$$

(siehe Satz 21)

- man berechnet die Quantile der Studentverteilung $T(n-1)$ mit $n-1$ Freiheitsgraden:

$$t_{1-\frac{\alpha}{2}} = t \cdot \text{ppf}(1 - \frac{\alpha}{2}, n-1), t_{1-\alpha} = t \cdot \text{ppf}(1 - \alpha, n-1), t_\alpha = t \cdot \text{ppf}(\alpha, n-1)$$

- ein **zweiseitiges Konfidenzintervall** für $\mu = E(X)$ ist $\left(\bar{X}_n - \frac{S_n}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}, \bar{X}_n + \frac{S_n}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}} \right)$, d.h.

$$P\left(\bar{X}_n - \frac{S_n}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}} < \mu < \bar{X}_n + \frac{S_n}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

weil:

$$P\left(\bar{X}_n - \frac{S_n}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}} < \mu < \bar{X}_n + \frac{S_n}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}\right) = P\left(-t_{1-\frac{\alpha}{2}} < \frac{\bar{X}_n - \mu}{\frac{S_n}{\sqrt{n}}} < t_{1-\frac{\alpha}{2}}\right)$$

$$\stackrel{(6)}{=} F_{T(n-1)}(t_{1-\frac{\alpha}{2}}) - F_{T(n-1)}(-t_{1-\frac{\alpha}{2}}) = F_{T(n-1)}(t_{1-\frac{\alpha}{2}}) - F_{T(n-1)}(t_{\frac{\alpha}{2}}) \stackrel{\text{siehe Satz 41}}{=} 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha$$

- **einseitige Konfidenzintervalle** für $\mu = E(X)$ sind: $\left(-\infty, \bar{X}_n - \frac{S_n}{\sqrt{n}} \cdot t_\alpha\right)$, $\left(\bar{X}_n - \frac{S_n}{\sqrt{n}} \cdot t_{1-\alpha}, \infty\right)$, d.h.

$$P\left(\mu < \bar{X}_n - \frac{S_n}{\sqrt{n}} \cdot t_\alpha\right) = 1 - \alpha, \quad P\left(\bar{X}_n - \frac{S_n}{\sqrt{n}} \cdot t_{1-\alpha} < \mu\right) = 1 - \alpha.$$

Konfidenzintervall für den Erwartungswert $\mu = E(X)$, wenn die Varianz des Merkmals X unbekannt ist	Ausdruck des Konfidenzintervalls anhand der Daten
zweiseitig	$\left(\bar{x}_n - \frac{s_n}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}, \bar{x}_n + \frac{s_n}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}\right)$
einseitig (mit oberer Grenze)	$\left(-\infty, \bar{x}_n - \frac{s_n}{\sqrt{n}} \cdot t_\alpha\right)$
einseitig (mit unterer Grenze)	$\left(\bar{x}_n - \frac{s_n}{\sqrt{n}} \cdot t_{1-\alpha}, \infty\right)$

Beispiel: Der Wert des Stichprobenmittels für das Merkmal X = der Durchmesser von einem Schraubentyp (von einer bestimmten Firma hergestellt) ist $\bar{x}_{100} = 15.5$ mm und die Stichprobenvarianz ist $s_{100}^2 = 0.09$ mm². Sei $1 - \alpha = 0.99$ das Konfidenzniveau; man gebe ein zweiseitiges Konfidenzintervall an für den theoretischen Erwartungswert $E(X)$.

Lsg.: Wert des zweiseitigen Konfidenzintervalls für $E(X)$, wenn die Varianz des Merkmals unbekannt ist:

$$\left(\bar{x}_n - \frac{s_n}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}, \bar{x}_n + \frac{s_n}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}} \right) = (15.421, 15.579)$$

wobei $\bar{x}_n = 15.5$, $s_n = 0.3$ ($s_n^2 = 0.09$), $\alpha = 0.01$, $t_{1-\frac{\alpha}{2}} = \text{t.ppf}(0.995, 99) = 2.6264$, $\sqrt{n} = 10$. ♣

Satz 22. Seien X_1, \dots, X_n Stichprobenvariablen für das Merkmal $X \sim N(\mu, \sigma^2)$; für die Stichprobenvarianz gilt $\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1)$.

Beispiel: Die Zeit die ein Zentralprozessor (CPU) benötigt, um bestimmte Operationen durchzuführen, hat eine normale Verteilung $N(\mu, \sigma^2)$ mit $\mu = 20$ Sekunden und $\sigma = 3$ Sekunden. In einer Stichprobe mit 25 Operationen, welches ist die Wahrscheinlichkeit, dass die Stichprobenvarianz (der Zeit um die studierten Operationen durchzuführen) mehr als 12 Sekunden ist?

Lsg.: Wir werden Satz 22 benutzen. Es gilt

$$P(S_{25}^2 > 12) = P\left(\frac{25-1}{3^2} S_{25}^2 > \frac{24}{9} \cdot 12\right) = 1 - P\left(\frac{24}{9} S_{25}^2 \leq 32\right).$$

Aber $\frac{24}{9} S_{25}^2 \sim \chi^2(25-1)$ (laut Satz 22)

$$\implies P(S_{25}^2 > 12) = 1 - F_{\chi^2(24)}(32) = 1 - \text{chi2.cdf}(32, 24) \approx 1 - 0.87301 = 0.12699.$$

➡ Konfidenzintervall für die theoretische Varianz $\sigma^2 = V(X)$ des beobachteten Merkmals X

Beispiel: Der Wert des Stichprobenmittels für das Merkmal X = der Durchmesser von einem Schraubentyp (von einer bestimmten Firma hergestellt) ist $\bar{x}_{100} = 15.5$ mm und die Stichprobenvarianz ist $s_{100}^2 = 0.09$ mm². Sei $1 - \alpha = 0.99$ das Konfidenzniveau; man gebe ein zweiseitiges Konfidenzintervall an für die theoretische Varianz $V(X)$. Ist die Varianz zu groß (d.h. mehr als 0.099 mm²), muss das Gerät, welches die Schrauben produziert, technisch angepasst werden. Man nimmt an, dass der Durchmesser einer solchen Schraube (von dieser Firma hergestellt) eine Normalverteilung hat.

► Gegeben sind: $\alpha \in (0, 1)$, die statistischen Daten x_1, \dots, x_n

► seien X_1, \dots, X_n Stichprobenvariablen für das Merkmal X

► man sucht ein Konfidenzintervall für den unbekannten Parameter $\sigma^2 = V(X)$ (die theoretische Varianz)

► falls $X \sim N(\mu, \sigma^2)$, dann folgt laut Satz 22, dass $\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1)$

► man berechnet die Quantile der $\chi^2(n-1)$ (Chi-Quadrat Verteilung mit $n-1$ Freiheitsgraden):

$$c_{1-\frac{\alpha}{2}} = \text{chi2.ppf}(1 - \frac{\alpha}{2}, n-1), c_{\frac{\alpha}{2}} = \text{chi2.ppf}(\frac{\alpha}{2}, n-1), c_{1-\alpha} = \text{chi2.ppf}(1 - \alpha, n-1), c_{\alpha} = \text{chi2.ppf}(\alpha, n-1)$$

• ein zweiseitiges Konfidenzintervall für $\sigma^2 = V(X)$ ist $\left(\frac{n-1}{c_{1-\frac{\alpha}{2}}} \cdot S_n^2, \frac{n-1}{c_{\frac{\alpha}{2}}} \cdot S_n^2 \right)$, d.h.

$$P\left(\frac{n-1}{c_{1-\frac{\alpha}{2}}} \cdot S_n^2 < \sigma^2 < \frac{n-1}{c_{\frac{\alpha}{2}}} \cdot S_n^2\right) = 1 - \alpha$$

- *einseitige Konfidenzintervalle* für $\sigma^2 = V(X)$ sind: $\left(0, \frac{n-1}{c_\alpha} \cdot S_n^2\right), \left(\frac{n-1}{c_{1-\alpha}} \cdot S_n^2, \infty\right)$, d.h.

$$P\left(\sigma^2 < \frac{n-1}{c_\alpha} \cdot S_n^2\right) = 1 - \alpha, \quad P\left(\frac{n-1}{c_{1-\alpha}} \cdot S_n^2 < \sigma^2\right) = 1 - \alpha.$$

Konfidenzintervall für die Varianz $\sigma^2 = V(X)$	Ausdruck des Konfidenzintervalls anhand der Daten
zweiseitig	$\left(\frac{n-1}{c_{1-\frac{\alpha}{2}}} \cdot S_n^2, \frac{n-1}{c_{\frac{\alpha}{2}}} \cdot S_n^2\right)$
einseitig (mit oberer Grenze)	$\left(0, \frac{n-1}{c_\alpha} \cdot S_n^2\right)$
einseitig (mit unterer Grenze)	$\left(\frac{n-1}{c_{1-\alpha}} \cdot S_n^2, \infty\right).$

Konfidenzintervall für die Standardabweichung $\sigma = Std(X) = \sqrt{V(X)}$	Ausdruck des Konfidenzintervalls anhand der Daten
zweiseitig	$\left(\sqrt{\frac{n-1}{c_{1-\frac{\alpha}{2}}}} \cdot s_n, \sqrt{\frac{n-1}{c_{\frac{\alpha}{2}}}} \cdot s_n\right)$
einseitig (mit oberer Grenze)	$\left(0, \sqrt{\frac{n-1}{c_\alpha}} \cdot s_n\right)$
einseitig (mit unterer Grenze)	$\left(\sqrt{\frac{n-1}{c_{1-\alpha}}} \cdot s_n, \infty\right).$

Beispiel: Der Wert des Stichprobenmittels für das Merkmal X = der Durchmesser von einem Schraubentyp (von einer bestimmten Firma hergestellt) ist $\bar{x}_{100} = 15.5$ mm und die Stichprobenvarianz ist $s_{100}^2 = 0.09$ mm². Sei $1 - \alpha = 0.99$ das Konfidenzniveau; man gebe ein zweiseitiges Konfidenzintervall an für die theoretische Varianz $V(X)$. Ist die Varianz zu groß (d.h. mehr als 0.099 mm²), muss das Gerät, welches die Schrauben produziert, technisch angepasst werden. Man nimmt an, dass der Durchmesser einer solchen Schraube (von dieser Firma hergestellt) eine Normalverteilung hat.

Lsg.: Wert des zweiseitigen Konfidenzintervalls für die theoretische Varianz $V(X)$ ist

$$\left(\frac{n-1}{c_{1-\frac{\alpha}{2}}} \cdot s_n^2, \frac{n-1}{c_{\frac{\alpha}{2}}} \cdot s_n^2\right)$$

wobei $\bar{x}_n = 15.5, s_n^2 = 0.09, \alpha = 0.01, c_{1-\frac{\alpha}{2}} = \text{chi2.ppf}(0.995, 99) = 138.99,$

$c_{\frac{\alpha}{2}} = \text{chi2.ppf}(0.005, 99) = 66.510$. Der Wert des zweiseitigen Konfidenzintervalls ist $(0.064107, 0.133965)$. Dieses Intervall enthält Werte, die größer als 0.099 sind, d.h. das Gerät muss technisch angepasst werden! ♣

Übung: Ein Hersteller behauptet, dass die Funktionszeit eines hergestellten Batterietyps 520 Stunden ist. Man hat 64 Batterien getestet und das Stichprobenmittel von 525 Stunden und Standardabweichung 25 Stunden erhalten. Man gebe ein zweiseitiges Konfidenzintervall an (mit $1 - \alpha = 0.99$) für

- den (theoretischen) Erwartungswert der Funktionszeit;
- die (theoretische) Varianz der Funktionszeit;
- die (theoretische) Standardabweichung der Funktionszeit.

Man setzt voraus dass die Funktionszeit eine Normalverteilung hat!

Hinweis: Für die Berechnung der Quantile benutze man Python, siehe das Python-Beispiel für Quantile auf Seite 41.

Statistische Tests

- Eine Grundgesamtheit wird bezüglich des Merkmals X untersucht, die Verteilung von X hängt vom unbekannten Parameter θ ab.
- Statistischer Test: Überprüfung von Hypothesen bezüglich θ , anhand einer Stichprobe

Beispiel: Servicebewertung für eine bestimmte Hotelkette - Umfrage zur Kundenzufriedenheit: *Wenn Sie an Ihre Erfahrungen mit unserem Hotel denken, wie würden Sie die Qualität des erhaltenen Kundenservice bewerten?*

Meinung des Kunden	Wert (der Meinung)	absolute Häufigkeit
Ausgezeichnet	2	30
Sehr zufriedenstellend	1	152
Neutral	0	180
Unbefriedigend	-1	172
Sehr schwach	-2	22

Kann man behaupten, dass die Kunden dieser Hotelkette durchschnittlich eine neutrale Meinung über dieses Hotel haben? (d.h. *durchschnittlich ist der Wert der Meinung gleich 0*)

↪ zu einer gegebenen Irrtumswahrscheinlichkeit α (z.B. $\alpha = 0.05$), führt man einen statistischen Test durch!

- $x_1, \dots, x_n \hookrightarrow$ *statistische Daten* (Beobachtungen, Stichprobenwerte) für das Merkmal X
- $X_1, \dots, X_n \hookrightarrow$ *Stichprobenvariablen* sind *unabhängige ZG* mit *derselben Verteilung* wie X .

Seien gegeben $\alpha \in (0, 1)$ das Signifikanzniveau (Irrtumswahrscheinlichkeit) und der Wert θ_0 , sowie die statistischen Daten x_1, \dots, x_n .

H_0 : Nullhypothese, H_1 : alternative Hypothese

- I. $H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0$
- II. $H_0 : \theta \geq \theta_0, \quad H_1 : \theta < \theta_0$
- III. $H_0 : \theta \leq \theta_0, \quad H_1 : \theta > \theta_0$

Die Funktion $G(X_1, \dots, X_n)$ der Stichprobenvariablen X_1, \dots, X_n heißt Testgröße oder Prüfgröße. Ihre Verteilung ist bekannt. Für die konkrete Stichprobe x_1, \dots, x_n ergibt sich der Wert $G(x_1, \dots, x_n)$ als Realisation der Zufallsgröße $G(X_1, \dots, X_n)$.

Man sucht einen Ablehnungsbereich $U \subset \mathbb{R}$, so dass für gegebenes Signifikanzniveau α gilt

$$P(G(X_1, \dots, X_n) \notin U | H_0) = 1 - \alpha.$$

Der Ablehnungsbereich U ist ein Bereich von Werten, die so unwahrscheinlich sind, dass man die Nullhypothese ablehnen kann. Wenn die Teststatistik einen Wert ergibt, der innerhalb des Ablehnungsbereichs liegt, wird die Nullhypothese verworfen und die Alternative angenommen. Der Ablehnungsbereich wird durch das Signifikanzniveau des Tests α bestimmt.

Schlussfolgerung des Tests:

$G(x_1, \dots, x_n) \notin U \Rightarrow H_0$ wird angenommen

$G(x_1, \dots, x_n) \in U \Rightarrow$ man lehnt H_0 ab, zugunsten von H_1

Man testet eine Grundgesamtheit bezüglich des Merkmals X .

- ▶ Test für den theoretischen Erwartungswert $E(X)$
 - ▷ wenn die Varianz bekannt ist: Gauß Test (Z-Test)
 - ▷ wenn die Varianz unbekannt ist: Student Test (T-Test)
- ▶ Test für die theoretische Standardabweichung $\sqrt{V(X)}$ oder Varianz $V(X)$: χ^2 -Test
- ▶ Test für den Anteilswert (approximativer Gauß Test)
- ▶ Unabhängigkeitstest

Bei statistischen Tests geht man schrittweise vor:

- ▶ Welcher Parameter soll getestet werden?
- ▶ Was ist die Hypothese und was die Alternative?
- ▶ Welcher ist der Wert von α ?
- ▶ Welcher Test ist geeignet?
- ▶ Berechnen der Schätzfunktion anhand der statistischen Daten
- ▶ Schlussfolgerung

➡ **Test für den Erwartungswert $\mu = E(X)$ des beobachteten Merkmals X , wenn die Varianz $\sigma^2 = V(X)$ des Merkmals X bekannt ist: Gauß Test, Z-Test**

I. $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$; **II.** $H_0: \mu \geq \mu_0, H_1: \mu < \mu_0$; **III.** $H_0: \mu \leq \mu_0, H_1: \mu > \mu_0$

- ▶ Gegeben sind: $\alpha \in (0, 1)$, $\mu_0 \in \mathbb{R}$, $\sigma > 0$, die statistischen Daten x_1, \dots, x_n
- ▶ seien X_1, \dots, X_n Stichprobenvariablen für das Merkmal X
- ▶ falls $X \sim N(\mu, \sigma^2)$ oder $n > 30$ und X hat eine beliebige Verteilung, dann folgt aus Satz 19, bzw. Satz 20

$$\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

- ▶ anhand der statistischen Daten x_1, \dots, x_n berechnet man $z = \frac{\bar{x}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}$

- ▶ Quantile der normalen Verteilung $N(0, 1)$:

$$z_{1-\frac{\alpha}{2}} = \text{norm.ppf}(1 - \frac{\alpha}{2}, 0, 1), z_{\alpha} = \text{norm.ppf}(\alpha, 0, 1), z_{1-\alpha} = \text{norm.ppf}(1 - \alpha, 0, 1)$$

- ➡ statistischer Test:

	I. $H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	II. $H_0: \mu \geq \mu_0$ $H_1: \mu < \mu_0$	III. $H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$
Man akzeptiert H_0 , wenn	$ z < z_{1-\frac{\alpha}{2}}$	$z > z_{\alpha}$	$z < z_{1-\alpha}$
Man lehnt H_0 ab, zugunsten von H_1 , wenn	$ z \geq z_{1-\frac{\alpha}{2}}$	$z \leq z_{\alpha}$	$z \geq z_{1-\alpha}$

- ➡ Statistische Tests und Konfidenzintervalle:

I. $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$

man akzeptiert $H_0 \iff |z| < z_{1-\frac{\alpha}{2}} \iff \bar{x}_n - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}} < \mu_0 < \bar{x}_n + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}$, d.h. μ_0 gehört dem zweiseitigen Konfidenzintervall an (siehe die Tabelle auf Seite 45)

II. $H_0: \mu \geq \mu_0, H_1: \mu < \mu_0$

man akzeptiert $H_0 \iff z > z_{\alpha} \iff \mu_0 < \bar{x}_n - \frac{\sigma}{\sqrt{n}} \cdot z_{\alpha}$, d.h. μ_0 gehört dem einseitigen Konfidenzintervall an (siehe die Tabelle auf Seite 45)

III. $H_0: \mu \leq \mu_0, H_1: \mu > \mu_0$

man akzeptiert $H_0 \iff z < z_{1-\alpha} \iff \bar{x}_n - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha} < \mu_0$, d.h. μ_0 gehört dem einseitigen Konfidenzintervall an (siehe die Tabelle auf Seite 45)

Beispiel: Ein Autohersteller behauptet, dass der Benzinverbrauch für einen neuen Autotyp *im Mittel* 5.5 Liter ist (pro 100 km). Dabei kann er davon ausgehen, dass der Verbrauch normalverteilt ist mit $\sigma = 0.3$ Liter. Eine Verbraucherzentrale vermutet, dass der Hersteller einen zu niedrigen Mittelwert angegeben hat und überprüft 20 Autos des neuen Typs auf ihren Verbrauch und berechnet einen empirischen Mittelwert von 5.8 Liter. Kann hiermit die Behauptung des Herstellers widerlegt werden? Als Signifikanzniveau nehme man $\alpha = 0.01$.

Lsg.: $H_0: \mu = 5.5$ mit $H_1: \mu \neq 5.5$, Varianz ist bekannt $\sigma^2 = 0.09$ (Gauß Test), $n = 20$, $\bar{x}_n = 5.8$, $\mu_0 = 5.5$

$$z = \frac{\bar{x}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{5.8 - 5.5}{\frac{0.3}{\sqrt{20}}} \approx 4.4721 > z_{1-\frac{\alpha}{2}} = \text{norm.ppf}(0.995, 0, 1) \approx 2.58$$

$\Rightarrow H_0$ wird abgelehnt \Rightarrow anhand dieser Stichprobe, stimmt die Behauptung des Herstellers *nicht* (die Behauptung des Herstellers kann widerlegt werden). ♣

➡ Test für den Erwartungswert $\mu = E(X)$ des Merkmals X , wenn die Varianz des Merkmals X unbekannt ist: Student Test, T-Test

I. $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$; II. $H_0: \mu \geq \mu_0, H_1: \mu < \mu_0$; III. $H_0: \mu \leq \mu_0, H_1: \mu > \mu_0$

► Gegeben sind: $\alpha \in (0, 1)$, $\mu_0 \in \mathbb{R}$, die statistischen Daten x_1, \dots, x_n

► seien X_1, \dots, X_n Stichprobenvariablen für das Merkmal X

► falls $X \sim N(\mu, \sigma^2)$ oder $n > 30$ und X hat eine beliebige Verteilung, dann $\frac{\bar{X}_n - \mu}{\frac{S_n}{\sqrt{n}}} \sim T(n-1)$

(siehe Satz 21)

► anhand der statistischen Daten x_1, \dots, x_n berechnet man $t = \frac{\bar{x}_n - \mu_0}{\frac{s_n}{\sqrt{n}}}$

► Quantile der Studentverteilung mit $n-1$ Freiheitsgraden $T(n-1)$:

$$t_{1-\frac{\alpha}{2}} = \text{t.ppf}(1 - \frac{\alpha}{2}, n-1), t_{\frac{\alpha}{2}} = \text{t.ppf}(\frac{\alpha}{2}, n-1), t_{1-\alpha} = \text{t.ppf}(1 - \alpha, n-1)$$

➡ statistischer Test:

	I. $H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	II. $H_0: \mu \geq \mu_0$ $H_1: \mu < \mu_0$	III. $H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$
Man akzeptiert H_0 , wenn	$ t < t_{1-\frac{\alpha}{2}}$	$t > t_\alpha$	$t < t_{1-\alpha}$
Man lehnt H_0 ab, zugunsten von H_1 , wenn	$ t \geq t_{1-\frac{\alpha}{2}}$	$t \leq t_\alpha$	$t \geq t_{1-\alpha}$

➡ Statistische Tests und Konfidenzintervalle:

I. $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$

man akzeptiert $H_0 \iff |t| < t_{1-\frac{\alpha}{2}} \iff \bar{x}_n - \frac{s_n}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}} < \mu_0 < \bar{x}_n + \frac{s_n}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}$, d.h. μ_0 gehört dem zweiseitigen Konfidenzintervall an (siehe die Tabelle auf Seite 46)

II. $H_0: \mu \geq \mu_0, H_1: \mu < \mu_0$

man akzeptiert $H_0 \iff t > t_\alpha \iff \mu_0 < \bar{x}_n - \frac{s_n}{\sqrt{n}} \cdot t_\alpha$, d.h. μ_0 gehört dem einseitigen Konfidenzintervall an (siehe die Tabelle auf Seite 46)

III. $H_0: \mu \leq \mu_0, H_1: \mu > \mu_0$

man akzeptiert $H_0 \iff t < t_{1-\alpha} \iff \bar{x}_n - \frac{s_n}{\sqrt{n}} \cdot t_{1-\alpha} < \mu_0$, d.h. μ_0 gehört dem einseitigen Konfidenzintervall an (siehe die Tabelle auf Seite 46)

Beispiel: Die Anleitungen eines Medikaments geben an, dass jede Tablette durchschnittlich 2.4 g aktive Substanzen enthält. 100 zufällig gewählte Tabletten werden untersucht und man stellt fest, dass im Mittel 2.5 g aktive Substanzen enthalten mit einer empirischen Standardabweichung von 0.2 g. Kann man behaupten, dass das Medikament die Angaben respektiert? ($\alpha = 0.01$)

Lsg.: $H_0: \mu = 2.4$ mit $H_1: \mu \neq 2.4$, Varianz ist unbekannt (Student Test), $n = 100$, $\bar{x}_n = 2.5$, $s_n = 0.2$, $\mu_0 = 2.4$

$$\Rightarrow t = \frac{\bar{x}_n - \mu_0}{\frac{s_n}{\sqrt{n}}} = \frac{2.5 - 2.4}{\frac{0.2}{\sqrt{100}}} = 5 > t_{1-\alpha/2} = t \cdot \text{ppf}(1 - \alpha/2, n - 1) \approx 2.6264$$

$\Rightarrow H_0$ wird abgelehnt \Rightarrow die Angaben werden nicht respektiert. ▲

➡ **Test für Standardabweichung $\sigma = \sqrt{V(X)}$ des beobachteten Merkmals X : Chi-Quadrat Test**

I. $H_0: \sigma = \sigma_0, H_1: \sigma \neq \sigma_0$; **II.** $H_0: \sigma \geq \sigma_0, H_1: \sigma < \sigma_0$; **III.** $H_0: \sigma \leq \sigma_0, H_1: \sigma > \sigma_0$

► Gegeben: $\alpha \in (0, 1)$, $\sigma_0 > 0$, x_1, \dots, x_n statistische Daten

► seien X_1, \dots, X_n Stichprobenvariablen für das Merkmal X

► wenn $X \sim N(\mu, \sigma^2)$, dann $\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2(n-1)$ (siehe Satz 22)

► anhand der statistischen Daten x_1, \dots, x_n berechnet man $c = \frac{n-1}{\sigma_0^2} \cdot s_n^2$

► Quantile der χ^2 Verteilung mit $n-1$ Freiheitsgraden:

$$c_{1-\frac{\alpha}{2}} = \chi^2 \cdot \text{ppf}(1 - \frac{\alpha}{2}, n-1), c_{\frac{\alpha}{2}} = \chi^2 \cdot \text{ppf}(\frac{\alpha}{2}, n-1), c_{1-\alpha} = \chi^2 \cdot \text{ppf}(1 - \alpha, n-1),$$

$$c_{\alpha} = \chi^2 \cdot \text{ppf}(\alpha, n-1)$$

➡ statistischer Test:

	I. $H_0: \sigma = \sigma_0$ $H_1: \sigma \neq \sigma_0$	II. $H_0: \sigma \geq \sigma_0$ $H_1: \sigma < \sigma_0$	III. $H_0: \sigma \leq \sigma_0$ $H_1: \sigma > \sigma_0$
Man akzeptiert H_0 , wenn	$c_{\frac{\alpha}{2}} < c < c_{1-\frac{\alpha}{2}}$	$c > c_{\alpha}$	$c < c_{1-\alpha}$
Man lehnt H_0 ab, zugunsten von H_1 , wenn	$c \notin (c_{\frac{\alpha}{2}}, c_{1-\frac{\alpha}{2}})$	$c \leq c_{\alpha}$	$c \geq c_{1-\alpha}$

➡ Statistische Teste und Konfidenzintervalle:

I. $H_0: \sigma = \sigma_0, H_1: \sigma \neq \sigma_0$

man akzeptiert $H_0 \iff c_{\frac{\alpha}{2}} < c < c_{1-\frac{\alpha}{2}} \iff \sqrt{\frac{n-1}{c_{1-\frac{\alpha}{2}}}} \cdot s_n < \sigma_0 < \sqrt{\frac{n-1}{c_{\frac{\alpha}{2}}}} \cdot s_n$, d.h. σ_0 gehört dem zweiseitigen Konfidenzintervall an (siehe die Tabelle auf Seite 48)

II. $H_0: \sigma \geq \sigma_0, H_1: \sigma < \sigma_0$

man akzeptiert $H_0 \iff c > c_{\alpha} \iff \sigma_0 < \sqrt{\frac{n-1}{c_{\alpha}}} \cdot s_n$, d.h. σ_0 gehört dem einseitigen Konfidenzintervall an (siehe die Tabelle auf Seite 48)

III. $H_0: \sigma \leq \sigma_0, H_1: \sigma > \sigma_0$

man akzeptiert $H_0 \iff c < c_{1-\alpha} \iff \sqrt{\frac{n-1}{c_{1-\alpha}}} \cdot s_n < \sigma_0$, d.h. σ_0 gehört dem einseitigen Konfidenzintervall an (siehe die Tabelle auf Seite 48)

Beispiel: Bei der Produktion eines Werkstückes wurde die Bearbeitungszeit untersucht. Für die als normalverteilt angesehene zufällige Bearbeitungszeit wurden 25 Werte (in Minuten) erfasst. Aus diesen Werten erhält man einen Mittelwert von 8.1 Minuten und Standardabweichung 1.6 Minuten. Testen Sie zum Niveau $\alpha = 0.05$ mit Hilfe der Konstruktion von entsprechenden Konfidenzintervallen, ob a) die Standardabweichung der Bearbeitungszeit 1.5 Minuten ist; b) der Erwartungswert der Bearbeitungszeit 7.8 Minuten ist.

Lsg.: a) $H_0: \sigma = 1.5, H_1: \sigma \neq 1.5$

$n = 25, \sigma_0 = 1.5, s_{25} = 1.6, \alpha = 0.05, c_{0.975} = \text{chi2.ppf}(0.975, 24) = 39.3641,$

$c_{0.025} = \text{chi2.ppf}(0.025, 24) = 12.4012$

Wert des zweiseitigen Konfidenzintervalls für theoretische Standardabweichung:

$$\left(\sqrt{\frac{n-1}{c_{1-\frac{\alpha}{2}}}} \cdot s_n, \sqrt{\frac{n-1}{c_{\frac{\alpha}{2}}}} \cdot s_n \right) = (1.249325, 2.225843) \implies \sigma_0 = 1.5 \in (1.249325, 2.225843)$$

$\implies H_0$ wird akzeptiert, d.h. anhand der Daten kann behauptet werden: die Standardabweichung der Bearbeitungszeit ist 1.5 Minuten.

b) $H_0: \mu = 7.8, H_1: \mu \neq 7.8$

$n = 25, \mu_0 = 7.8, \bar{x}_{25} = 8.1, s_{25} = 1.6, \alpha = 0.05, t_{0.975} = \text{t.ppf}(0.975, 24) = 2.0639$

Wert des zweiseitigen Konfidenzintervalls für theoretischen Erwartungswert:

$$\left(\bar{x}_n - \frac{s_n}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}, \bar{x}_n + \frac{s_n}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}} \right) = (7.439552, 8.760448) \implies m_0 = 7.8 \in (7.439552, 8.760448)$$

$\implies H_0$ wird akzeptiert, d.h. anhand der Daten kann behauptet werden: der Erwartungswert der Bearbeitungszeit ist 7.8 Minuten. ♣

Bemerkung: Tests für die Varianz $\sigma^2 = V(X)$ des beobachteten Merkmals X werden analog durchgeführt, weil

$$\sigma^2 = \sigma_0^2 \iff \sigma = \sigma_0$$

$$\sigma^2 \neq \sigma_0^2 \iff \sigma \neq \sigma_0$$

$$\sigma^2 < \sigma_0^2 \iff \sigma < \sigma_0$$

$$\sigma^2 > \sigma_0^2 \iff \sigma > \sigma_0.$$

➡ **Test für Anteilswert p des beobachteten Merkmals $X \sim \text{Bernoulli}(p)$: Approximativer Gauß Test**

I. $H_0: p = p_0, H_1: p \neq p_0$; **II.** $H_0: p \geq p_0, H_1: p < p_0$; **III.** $H_0: p \leq p_0, H_1: p > p_0$

► Gegeben sind: $\alpha \in (0, 1), p_0 \in (0, 1)$, die statistischen Daten $x_1, \dots, x_n \in \{0, 1\}$

► seien X_1, \dots, X_n Stichprobenvariablen für das Merkmal X

► falls $X \sim \text{Bernoulli}(p)$ und $np(1-p) \geq 10$, dann $\frac{\bar{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0, 1)$ (siehe Satz 20)

► anhand der statistischen Daten x_1, \dots, x_n berechnet man $z = \frac{\bar{x}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ und verifiziert ob

zusätzlich $np_0(1-p_0) \geq 10$

► Quantile der normalen Verteilung $N(0, 1)$:

$z_{1-\frac{\alpha}{2}} = \text{norm.ppf}(1 - \frac{\alpha}{2}, 0, 1), z_{1-\alpha} = \text{norm.ppf}(1 - \alpha, 0, 1), z_\alpha = \text{norm.ppf}(\alpha, 0, 1)$

► dieser Test kann durchgeführt werden, wenn $np_0(1-p_0) \geq 10$:

	I. $H_0: p = p_0$ $H_1: p \neq p_0$	II. $H_0: p \geq p_0$ $H_1: p < p_0$	III. $H_0: p \leq p_0$ $H_1: p > p_0$
Man akzeptiert H_0 , wenn	$ z < z_{1-\frac{\alpha}{2}}$	$z > z_\alpha$	$z < z_{1-\alpha}$
Man lehnt H_0 ab, zugunsten von H_1 , wenn	$ z \geq z_{1-\frac{\alpha}{2}}$	$z \leq z_\alpha$	$z \geq z_{1-\alpha}$

Beispiel: Eine Münze wurde 100-mal geworfen und man erhielt 61-mal “Kopf”. Wir sollen testen, ob die Münze fair ist, in dem Sinne, dass die Wahrscheinlichkeit für das Auftreten von “Kopf” gleich der Wahrscheinlichkeit

für das Auftreten von "Zahl" ist, d.h. man führt einen Test bzgl. dem Anteilswert p durch, zu testen ist $p = 0.5$, gegen $p \neq 0.5$. Man wählt die Irrtumswahrscheinlichkeit $\alpha = 0.05$.

Lsg.: $np_0(1 - p_0) = 100 \cdot 0.5 \cdot 0.5 \geq 10$

$H_0 : p = 0.5, H_1 : p \neq 0.5$, Test für den Anteilswert p

$$z = \frac{\bar{x}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{61}{100} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{100}}} = 2.2; |z| = z > z_{1-\frac{\alpha}{2}} = \text{norminv}(1 - \frac{0.05}{2}, 0, 1) = 1.96 \implies H_0 \text{ wird}$$

abgelehnt; anhand der Daten schlußfolgert man, dass die Münze *nicht* fair ist. ■

➡ Unabhängigkeitstest zweier diskreter Merkmale X und Y

► sei $\alpha \in (0, 1)$ Signifikanzniveau (Irrtumswahrscheinlichkeit)

► X nimmt die Werte $\{a_1, \dots, a_r\}$ an, bzw. Y die Werte $\{b_1, \dots, b_s\}$

► gegeben sind die statistischen Daten $(x_i, y_j), i \in I_X, j \in I_Y$, für (X, Y) (I_X, I_Y Indexmengen)

► Nullhypothese / alternative Hypothese

$H_0 : X$ und Y sind unabhängig $H_1 : X$ und Y sind nicht unabhängig

► man berechnet:

- $n_{ij} = \#\{(k, l) \in I_X \times I_Y : x_k = a_i \text{ und } y_l = b_j\}$

- $n_{i.} := \sum_{j=1}^s n_{ij}, \quad n_{.j} := \sum_{i=1}^r n_{ij}, \quad n := \sum_{i=1}^r \sum_{j=1}^s n_{ij}$

- $c = \sum_{i=1}^r \sum_{j=1}^s \frac{\left(n_{ij} - \frac{n_{i.} \cdot n_{.j}}{n}\right)^2}{\frac{n_{i.} \cdot n_{.j}}{n}}$

- das Quantil $c_{1-\alpha} = \text{chi2.ppf}(1 - \alpha, (r - 1)(s - 1))$

➡ Schlußfolgerung des Tests:

→ wenn $c \leq c_{1-\alpha} \Rightarrow H_0$ wird akzeptiert

→ wenn $c > c_{1-\alpha} \Rightarrow H_0$ wird abgelehnt.

Beispiel: Gegeben sind die statistischen Daten bzgl. den Urlaubs-Präferenzen von Männern (M) und Frauen (F):

	Meer	Gebirge
M	209	280
F	225	248

. Sind die Urlaubs-Präferenzen geschlechtsunabhängig? (sei $\alpha = 0.05$)

Lsg.: Man führt den Unabhängigkeitstest zweier diskreter Merkmale durch:

X : Geschlecht (Werte: M, F), $r = 2$;

Y : Urlaubs-Präferenzen (Werte: $Meer, Gebirge$), $s = 2$.

▷ aus der Tabelle: $n_{11} = 209, n_{12} = 280, n_{21} = 225, n_{22} = 248$

$\Rightarrow n_{1.} = 489, n_{.1} = 434, n_{2.} = 473, n_{.2} = 528, n = 962$

$$c = \sum_{i=1}^2 \sum_{j=1}^2 \frac{\left(n_{ij} - \frac{n_{i.} \cdot n_{.j}}{n}\right)^2}{\frac{n_{i.} \cdot n_{.j}}{n}} \approx 2.2622$$

▷ es gilt $\text{chi2.ppf}(1 - 0.05, 1) = 3.8415 > c$

$\Rightarrow H_0$ wird akzeptiert, d.h. anhand der statistischen Daten kann man behaupten: die Urlaubs-Präferenzen sind *unabhängig* vom Geschlecht! ♣

Statistische Fehlentscheidungen

$$P(\text{Fehler erster Art}) = P(\text{man lehnt } H_0 \text{ ab} | H_0 \text{ ist wahr}) = \alpha$$

$$P(\text{Fehler zweiter Art}) = P(\text{man akzeptiert } H_0 | H_1 \text{ ist wahr}) = \beta$$

Entscheidung \ Realität	H_0 ist wahr	H_1 ist wahr
man lehnt H_0 ab	Fehler 1. Art	richtige Entscheidung
man akzeptiert H_0	richtige Entscheidung	Fehler 2. Art

Analogie zu Strafverfahren - Realität: Der Angeklagte ist schuldig / unschuldig; die Entscheidung wird getroffen:
Der Angeklagte ist schuldig / unschuldig.

Entscheidung \ Realität	Der Angeklagte ist schuldig	Der Angeklagte ist unschuldig
"Der Angeklagte ist unschuldig"	Fehler 1. Art	richtige Entscheidung
"Der Angeklagte ist schuldig"	richtige Entscheidung	Fehler 2. Art