



TECHNISCHE UNIVERSITÄT CHEMNITZ

Stochastik für Informatiker

Dr. D. Uhlig-Düvelmeyer - SS 2008

Copyright © 2008 Tobias Doerffel

Diese privaten Mitschriften der o.g. Vorlesung erheben weder den Anspruch auf Vollständigkeit noch auf Fehlerfreiheit. Die Verwendung der hier vorliegenden Informationen geschieht auf eigene Gefahr! Korrekturhinweise an [tobias.doerffel \)at\(informatik.tu-chemnitz.de](mailto:tobias.doerffel@informatik.tu-chemnitz.de) werden dankend entgegengenommen.

Weitere Informationen auf <http://www.tu-chemnitz.de/~doto/>

Chemnitz, 8. April 2009

Inhaltsverzeichnis

1	Einführung in die Wahrscheinlichkeitsrechnung	3
1.1	Zufälle Ereignisse und deren Wahrscheinlichkeit	3
1.1.1	Rechnen mit zufälligen Ereignissen	5
1.1.2	Relative Häufigkeiten	6
1.1.3	Wahrscheinlichkeit	7
1.1.4	Geometrische Wahrscheinlichkeit	10
1.1.5	Bedingte Wahrscheinlichkeit	10
1.1.6	Stochastische Unabhängigkeit von Ereignissen	14
1.2	Zufallsgrößen	15
1.2.1	Die Verteilungsfunktion	15
1.2.2	Diskrete Zufallsgrößen	17
1.2.3	Stetige Zufallsgrößen	18
1.2.4	Erwartungswert und Varianz	19
1.2.5	Spezielle Verteilungen	23
1.3	Das Gesetz der großen Zahlen und Grenzwertsätze	32
1.4	Mehrdimensionale Verteilungen	35
2	Einführung in die mathematische Statistik	40
2.1	Grundbegriffe der beschreibenden Statistik	40
2.2	Grundbegriffe der beurteilenden Statistik	44
2.2.1	Parameterschätzung	45
2.2.2	Intervallschätzungen (Bereichsschätzungen)	49
2.2.3	Hauptsatz der mathematischen Statistik	55
2.2.4	Prüfen statistischer Hypothesen (Tests)	56

1 Einführung in die Wahrscheinlichkeitsrechnung

Stochastik: griech. „Kunst des Mutmaßens“ → Mathematik des Zufalls

Ziel:

Wissen über die Unsicherheit zu gewinnen (wie verhält sich der Zufall?)

Aufteilung des mathematischen Gebiets:

- Wahrscheinlichkeitsrechnung: Theorie zufälliger Ereignisse, Beschreibung von Modellen
- beschreibende Stochastik (deskriptive Statistik): Strukturierung und Zusammenfassung von Daten (operative Datenanalyse)
- beurteilende Statistik (induktive/schließende Statistik): induktiver Schluss anhand einer Zufallsgröße auf die zugrundeliegende Grundgesamtheit
- Spezialgebiete: stochastische Prozesse, Bedienungstheorie

Anwendungsgebiete

- statistische Qualitätskontrolle
- Versicherungswesen
- Warteschlangentheorie
- Medizin
- Meinungsforschung
- Finanzmathematik

1.1 Zufälle Ereignisse und deren Wahrscheinlichkeit

Definition:

Ein stochastischer Vorgang heißt (ideales) *Zufallsexperiment*, wenn Folgendes gilt:

- das Experiment wird unter genau festgelegten Bedingungen (Versuchsbedingungen) durchgeführt
- die Menge der möglichen Ergebnisse (Ausgänge) ist vor dem Experiment bekannt
- das Experiment kann zumindest prinzipiell beliebig oft wiederholt werden (unter gleichen Bedingungen)

Definition:

Die Menge der möglichen Ereignisse eines idealen Zufallsexperiments heißt Ergebnisraum (Ereignisraum, Grundmenge, Merkmalraum) und wird mit Ω bezeichnet.

Liegt ein endlicher Ergebnisraum vor, schreiben wir $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$

Im Falle eines abzählbar-unendlichen Ergebnisraumes setzen wir $\Omega = \{\omega_j, \dots, j \in \mathbb{N}\}$

Es gibt auch Ergebnisräume mit überabzählbar vielen Elementen, z.B. $\Omega = [0, T]$.

Beispiel: wiederholter Würfelwurf, solange bis '6' fällt. Beobachten Anzahl der benötigten Würfe. $\Omega = \{\omega_1, \omega_2, \dots\}$ mit $\omega_j =$ 'i-Würfe' bzw. $\omega_i = i$.

Beispiel: Anzahl der Anrufe in einer Telefonzentrale zwischen 8:00 und 9:00 Uhr.

- Möglichkeit 1: $\Omega = \{\omega_0, \omega_1, \dots\}$, $\omega_i =$ 'i Anrufe'
- Möglichkeit 2: $\Omega = \{\omega_0, \omega_1, \dots, \omega_{100}\}$ mit $\omega_i = \begin{cases} i \leq 99 & \text{abzählbar-unendlich} \\ i = 100 & \text{endlich} \end{cases}$

Oft interessiert nicht das konkrete Ergebnis eines Zufallsexperiments, sondern nur, ob der Ausgang des Experiments zu einer gewissen Menge gehört.

Beispiel: Download-Zeit nicht 'genau 60 Sekunden', sondern 'nicht mehr als 60 Sekunden'.

Definition:

Ein zufälliges Ereignis A ist eine Teilmenge des Ergebnisraumes Ω . Man sagt das Ereignis A tritt ein, wenn ein Ergebnis $\omega \in \Omega$ mit $\omega \in A$ beobachtet wird.

Bemerkung: Nach der Durchführung des Experiments kann man stets sagen, ob A eingetreten ist ($\omega \in A$) oder nicht ($\omega \notin A$). Nicht jede Teilmenge von Ω muss als zufälliges Ereignis sinnvoll sein (aber alle zufälle Ereignisse sind Teilmengen von Ω).

Spezielle Ereignisse:

$A = \emptyset \subset \Omega \rightarrow$ unmögliches Ereignis, weil $\omega \in \emptyset$ nie eintritt

$A = \Omega \subset \Omega \rightarrow$ sicheres Ereignis, weil $\omega \in \Omega$ immer eintritt

$A = \{\omega\}$ für $\omega \in \Omega \rightarrow$ Elementarereignis

Definition:

Die Menge aller betrachteten Ereignisse heißt Ereignissystem \mathfrak{a} (σ -Algebra, Ereignisalgebra), wenn folgende 3 Bedingungen erfüllt sind.

- $\Omega \in \mathfrak{a}$
- aus $A \in \mathfrak{a}$ folgt $\bar{A} \in \mathfrak{a}$ (Komplementäreignis, Gegenereignis) $\forall A \in \mathfrak{a}$
- $A_1, A_2, \dots \in \mathfrak{a} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathfrak{a} \quad \forall A_1, A_2, \dots \in \mathfrak{a}$

Bemerkung:

- \mathfrak{a} Ereignissystem: $\emptyset \in \mathfrak{a}$ ($\emptyset = \bar{\Omega}$), $A, B \in \mathfrak{a} \Rightarrow A \cap B \in \mathfrak{a}$ ($A \cap B = \overline{\bar{A} \cup \bar{B}}$, $A \setminus B \in \mathfrak{a}$)
- sinnvolle Operationen führen aus \mathfrak{a} nicht heraus
- Falls Ω endlich ist, wählt man der Einfachheit halber oft die Menge aller Teilmengen von Ω als Ereignissystem (d.h. $\mathfrak{a} = \mathcal{P}(\Omega)$). Dann ist tatsächlich jede Teilmenge von Ω ein Ereignis. Für überabzählbare Mengen kann man jedoch „wilde“ Teilmengen von Ω konstruieren, denen keine vernünftige Wahrscheinlichkeit mehr zugeordnet werden kann. Deswegen konzentriert man sich auf Ereignissysteme.
- Ereignisse sind Mengen ($\subset \Omega$), so dass mit Ereignissen wie mit Mengen gerechnet werden kann.

1.1.1 Rechnen mit zufälligen Ereignissen

- Durchschnitt: $A \cap B = \{\omega \in \Omega : \omega \in A \text{ und } \omega \in B\}$
- Vereinigung: $A \cup B = \{\omega \in \Omega : \omega \in A \text{ oder } \omega \in B\}$
- Differenz von B und A : $B \setminus A = B \cap \bar{A} = \{\omega \in \Omega, \omega \in B \text{ und } \omega \notin A\}$
- Komplementäreignis: $\bar{A} = \{\omega \in \Omega : \omega \notin A\} = \Omega \setminus A$ (A tritt nicht ein)
- A und B sind unvereinbar, wenn $A \cap B = \emptyset$
- Gleichheit: $A = B \Leftrightarrow A \subseteq B$ und $B \subseteq A$

Rechenregeln

- $A \cup B = B \cup A$, $A \cap B = B \cap A$ (Kommutativgesetz)
- $(A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C$, $(A \cap B) \cap C = A \cap (B \cap C) = A \cap B \cap C$ (Assoziativgesetz)
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (Distributivgesetz)

- $\overline{A \cap B} = \overline{A} \cup \overline{B}, \overline{A \cup B} = \overline{A} \cap \overline{B}$
 $\bigcap_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} \overline{A} \bigcup_{i=1}^{\infty} A_i = \bigcap_{i=1}^{\infty} \overline{A}$ (Regeln von De Morgan)

1.1.2 Relative Häufigkeiten

Wir betrachten ein Zufallsexperiment, das durch seinen Ergebnisraum Ω und sein Ereignissystem \mathfrak{a} beschrieben ist. Sei $A \in \mathfrak{a}$ ein Ereignis. Wir wiederholen das Experiment unter gleichen Bedingungen n mal und notieren den Ausgang a_j (=Ergebnis im j -ten Experiment ($j = 1, \dots, n$)) und erhalten einen n -Tupel $a = (a_1, \dots, a_n)$

→ relative Häufigkeit $h_{n,a}(A) = \frac{\text{Anzahl der } j \in \{1, \dots, n\} \text{ mit } a_j \in A}{n} = \frac{n_A}{n}$

(„Gewissheitsgrad“ für das Eintreten von A)

$h_{n,a}(A)$ hängt ab von

- Ereignis A
- Anzahl n
- konkreten Beobachtungen \mathfrak{a}

Beispiel:

Wurf eines Würfels $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathfrak{a} = \mathcal{P}(\Omega)$, $A = \{2, 4, 6\}$ =gerade Augenzahl, $n = 10$

$a_1 = 3, a_2 = 3, a_3 = 2, a_4 = 5, a_5 = 4, a_6 = 1, a_7 = 6, a_8 = 2, a_9 = 4, a_{10} = 5$

$a = \{3, 3, 2, 5, 4, 1, 6, 2, 4, 5\} \rightarrow h_{n,a}(A) = \frac{5}{10} = \frac{1}{2}$

Eigenschaften der relativen Häufigkeit

- $0 \leq h_{n,a}(A) \leq 1$
- $h_{n,a}(\Omega) = 1, h_{n,a}(\emptyset) = 0$
- $h_{A \cup B} = h_{n,a}(A) + h_{n,a}(B)$ falls $A \cap B = \emptyset$ ($h_{n,a}(A \cup B) = \frac{n_A + n_B}{n} = \frac{n_A}{n} + \frac{n_B}{n}$)

Wir interpretieren $h_{n,a}(A)$ als empirischen Gewissheitsgrad für das Eintreten von A . Bei erneuter n -maliger Durchführung tritt i.A. ein anderes n -Tupel $b = (b_1, \dots, b_n)$ ein und es gilt $h_{n,a}(A) \neq h_{n,b}(A)$

Oftmals stabilisieren sich die relativen Häufigkeiten für wachsende $n \rightarrow$ Wahrscheinlichkeit=„Grenzwert“ von $h_{n,a}(A)$ für $n \rightarrow \infty$.

aber: Schwierigkeiten, da Konvergenzbegriff nicht im klassischen Sinne verwendbar ist. \Rightarrow Definition von Wahrscheinlichkeit nicht über relative Häufigkeit, sondern implizit durch axiomatische Festlegung ihrer Eigenschaften.

1.1.3 Wahrscheinlichkeit

Definition: (Axiomensystem von KOLGOMOROV (1933))

Sei Ω ein Ergebnisraum und \mathfrak{a} eine σ -Algebra von Ereignissen über Ω . Sei $P(\dots)$ eine für beliebige $A \subset \mathfrak{a}$ definierte reellwertige Funktion, die folgenden Axiomen genügt:

- $P(A) \geq 0 \quad \forall A \in \mathfrak{a}$
- $P(\Omega) = 1$
- $P\left(\underbrace{\bigcup_{i=1}^{\infty} A_i}_{\in \mathfrak{a}}\right) = \sum_{i=1}^{\infty} P(A_i) \quad \forall A_i \in \mathfrak{a} \text{ mit } A_i \cap A_j = \emptyset \ (i \neq j)$

Dann heißt P Wahrscheinlichkeitsmaß auf \mathfrak{a} und die Zahl $P(A)$ für $A \in \mathfrak{a}$ heißt Wahrscheinlichkeit des Ereignisses A .

Definition:

Das Tripel $(\Omega, \mathfrak{a}, P)$ heißt Wahrscheinlichkeitsraum oder Wahrscheinlichkeitsmodell.

Satz: Sei $(\Omega, \mathfrak{a}, P)$ ein Wahrscheinlichkeitsraum. Dann gilt

- $P(\emptyset) = 0$
- $P(A \cup B) = P(A) + P(B) \quad \forall A, B \in \mathfrak{a} \text{ mit } A \cap B = \emptyset$
- $P(\bar{A}) = 1 - P(A)$
- $P(A) \leq 1 \quad \forall A \in \mathfrak{a}$
- aus $A \subseteq B$ folgt $P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \forall A, B \in \mathfrak{a}$

Beweis:

1. $\Omega = \Omega \cup \emptyset$ sowie $\Omega \cap \emptyset = \emptyset \Rightarrow P(\Omega) = P(\Omega) + P(\emptyset) \Rightarrow 1 = 1 + P(\emptyset) \Rightarrow P(\emptyset) = 0$
2. klar
3. $\Omega = A \cup \bar{A}$ sowie $A \cap \bar{A} = \emptyset \Rightarrow 1 = P(\Omega) = P(A) + P(\bar{A}) \Rightarrow P(\bar{A}) = 1 - P(A)$
4. $P(A) = 1 - P(\bar{A}) \leq 1$

1.1 Zufälle Ereignisse und deren Wahrscheinlichkeit

5. $A \subseteq B \Rightarrow B = A \cup (B \setminus A)$ mit $A \cap (B \setminus A) = \emptyset \Rightarrow P(B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A) \geq P(A)$
6. $A \cup B = (A \setminus B) \cup (A \cap B) \cup (B \setminus A) \Rightarrow P(A \cup B) = P(A \setminus B) + P(A \cap B) + P(B \setminus A)$
ferner gilt $A = (A \cap B) \cup (A \cap \bar{B}) = (A \cap B) \cup (A \setminus B)$
 $\Rightarrow P(A) = P(A \cap B) + P(A \setminus B)$ $P(B) = P(A \cap B) + P(B \setminus A)$
 $\Rightarrow P(A \cup B) = P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B) = P(A) + P(B) - P(A \cap B)$

Beispiel: Qualitätskontrolle

$\Omega = \{\omega_1, \omega_2\} = \{0, 1\}$ ($\omega_1 = 0 =$ Produkt defekt, $\omega_1 = 1 =$ Produkt nicht defekt)

$\mathfrak{a} = \mathcal{P}(\Omega) = \{\emptyset, \{0\}, \{1\}, \Omega\}$

- \mathcal{P} muss für alle $A \in \mathfrak{a}$ definiert werden, so dass die Axiome erfüllt sind $\Rightarrow P(\emptyset) = 0, P(\Omega) = 1$ muss gelten
- Sei $P(\{0\}) = p \rightarrow 0 \leq p \leq 1$ $P(\bar{A}) = 1 - P(A), \{0\} = \bar{\{1\}} \Rightarrow P(\{1\}) = P(\{0\}) = 1 - P(\{0\}) = 1 - p$

Damit ist das Wahrscheinlichkeitsmodell durch die Angabe einer Zahl p mit $0 \leq p \leq 1$ charakterisiert.

Definition:

Ein Zufallsexperiment mit genau 2 möglichen Ausgängen heißt Bernoulli-Experiment. Als Ergebnisraum benutzt man $\Omega = \{0, 1\}$. Das Wahrscheinlichkeitsmaß wird durch $P(\{1\}) = p$ mit $0 \leq p \leq 1$ festgelegt. Der resultierende Wahrscheinlichkeitsraum $(\Omega, \mathfrak{a}, P)$ heißt Bernoulli-Modell.

Definition:

Ein Zufallsexperiment habe endlich viele Versuchsausgänge, d.h. $\Omega = \{\omega_1, \dots, \omega_N\}$, die alle gleichmögich sind (d.h. alle Elementarereignisse ω_i haben die gleiche Wahrscheinlichkeit)

$$\rightarrow \text{wegen } 1 = P(\Omega) = P\left(\bigcup_{i=1}^{\infty} \{\omega_i\}\right) = \sum_{i=1}^N P(\{\omega_i\}) = N \cdot P(\{\omega_i\}) \quad i = 1 \dots N$$

$$\Rightarrow P(\{\omega_i\}) = \frac{1}{N} \quad i = 1 \dots N$$

Dann heißt der Wahrscheinlichkeitsraum $(\Omega, \mathcal{P}(\Omega), P)$ Laplacescher Wahrscheinlichkeitsraum der Ordnung N , das Zufallsexperiment heißt Laplace-Experiment, das Wahrscheinlichkeitsmaß P heißt Laplace-Verteilung oder diskrete Gleichverteilung auf Ω .

Folgerung: Sei $A \in \mathcal{P}(\Omega) = \mathfrak{a}$ ein beliebiges Ereignis. Dann gilt offenbar

$$P(A) = \frac{\text{Anzahl der für } A \text{ günstigen Elementarereignisse}}{\text{Anzahl aller Elementarereignisse}} = \frac{\text{Anzahl } j \in \{1, \dots, N\} \text{ mit } \omega_j \in A}{N}$$

Bemerkung: Diese Vorgehensweise funktioniert nur in Laplaceschen Wahrscheinlichkeitsräumen! Sie heißt auch *klassische Methode zur Bestimmung von Wahrscheinlichkeiten*.

Beispiel:

Wurf mit fairem Würfel $\rightarrow \Omega = \{\omega_1, \dots, \omega_6\}$ mit $\omega_i = i$, $P(\{\omega_i\}) = \frac{1}{6}$, $\mathfrak{a} = \mathcal{P}(\Omega)$

A=„Wurf einer gerade Augenzahl“ = $\{\omega_2, \omega_4, \omega_6\} \in \mathfrak{a} \rightarrow P(A) = \frac{3}{6} = \frac{1}{2}$

Hilfsmittel aus der Kombinatorik

Zählen aller Möglichkeiten, aus einer n -elementigen Menge k Elemente auszuwählen:

\rightarrow kommt es auf die Reihenfolge der ausgewählten Elemente an?

\rightarrow dürfen Elemente mehrfach ausgewählt werden?

1. Reihenfolge wichtig, Wiederholung möglich: Variation von n Elementen zur k -ten Klasse mit Wiederholungen:

$$\overline{V}_n^k = n^k$$

2. Reihenfolge wichtig, Wiederholung nicht möglich: Variation von n Elementen zur k -ten Klasse ohne Wiederholung:

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

3. Reihenfolge nicht wichtig, Wiederholung möglich: Kombination von n Elementen zur k -ten Klasse mit Wiederholung

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

4. Reihenfolge nicht wichtig, Wiederholung nicht möglich: Kombination von n Elementen zur k -ten Klasse ohne Wiederholung

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

1.1.4 Geometrische Wahrscheinlichkeit

klassische Methode (Laplace-Experiment): endliche Anzahl möglicher Versuchsausgänge \rightarrow jetzt geometrische Objekte (Kurven, Flächen, Volumen) mit endlichem Inhalt beinhalten aber unendlich viele Punkte, d.h. zufällig gewählter Punkt hat immer die Wahrscheinlichkeit 0 \rightarrow Spezialfall der klassischen Wahrscheinlichkeit unter folgenden Bedingungen:

- Ω lässt sich als geometrisches Objekt mit endlichem Inhalt darstellen
- Wahrscheinlichkeit für ein Ereignis A ist proportional zur Größe dieser Teilmenge

Dann gilt:

$$P(A) = \frac{\text{Inhalt von } A}{\text{Inhalt von } \Omega}$$

Beispiel: Stellung des Minutenzeigers einer Uhr (Zeiger Länge r)

klar: $|\Omega| = r^2 \cdot \pi$ $|B| = \frac{\alpha}{360^\circ} \pi r^2$ $\alpha = \frac{360^\circ}{30} = 12^\circ$

B =Minutenzeiger zwischen 10 und 12 $\Rightarrow P(B) = \frac{|B|}{|\Omega|} = \frac{\frac{\alpha}{360^\circ} \pi r^2}{\pi r^2} = \frac{1}{30}$

Beispiel: 2 Personen vereinbaren Treffen zwischen 12:00 und 13:00 Uhr. Jeder trifft unabhängig vom anderen zu einem zufälligen Zeitpunkt zwischen 12:00 und 13:00 Uhr ein.

A =Person, die zuerst eintrifft, muss max. 10 Minuten warten

$$P(A) = \frac{60 \cdot 60 - 2 \frac{50 \cdot 50}{2}}{60 \cdot 60} = \frac{1100}{3600} = \frac{11}{36} = 0,3056$$

1.1.5 Bedingte Wahrscheinlichkeit

Verschiedene Ereignisse sind oft nicht unabhängig voneinander. Die Wahrscheinlichkeit kann unterschiedlich sein, je nachdem, ob ein anderes Ereignis eingetreten ist oder nicht.

Beispiel: Skatenspiel (3 Spieler, jeder je 10 Karten, 2 Karten im Skat)

A =Grün-As bei Spieler 1 $\rightarrow P(A) = \frac{10}{32}$

B =Grün-As im Skat $\rightarrow P(B) = \frac{2}{32}$

C =Spieler 2 hat Grün-As nicht (Zusatzinfo)

$\Rightarrow P(A|C) = \frac{10}{22}, P(B|C) = \frac{2}{22} \rightarrow$ Wahrscheinlichkeit von Ereignis A unter Bedingung C (Zusatzinformation/Vorabinformation)

es gilt: $P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{\frac{10}{32}}{\frac{22}{32}} = \frac{10}{22}$

Definition:

Es sei $(\Omega, \mathfrak{a}, P)$ ein Wahrscheinlichkeitsraum und A und B Ereignisse ($A, B \in \mathfrak{a}$) mit $P(B) > 0$. Dann heißt $P(A|B) = \frac{P(A \cap B)}{P(B)}$ die *bedingte Wahrscheinlichkeit* des Ereignisses A unter der Bedingung B . Manchmal wird $P(A|B) = P_B(A)$ geschrieben.

Beispiel: Zweimaliges Würfeln mit idealem Würfel.

$A = \text{„erster Wurf 6“} = \{(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\} \rightarrow P(A) = \frac{6}{36} = \frac{1}{6}$

$B = \text{„Augensumme beider Würfe ist 8“} = \{(6, 2), (2, 6), (5, 3), (3, 5), (4, 4)\}$
 $\rightarrow P(B) = \frac{5}{36}$

$A \cap B = \{(6, 2)\} \Rightarrow P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{36}}{\frac{5}{36}} = \frac{1}{5}$

Satz: (Rechenregeln für bedingte Wahrscheinlichkeit)

Es sei $(\Omega, \mathfrak{a}, P)$ ein Wahrscheinlichkeitsraum und $A_1, A_2, B \in \mathfrak{a}$ mit $P(B) > 0$. Dann gilt

1. $P(B|B) = 1$ und $P(\emptyset|B) = 0$
2. $A \cap B = \emptyset \Rightarrow P(A|B) = 0$
3. $P(A|B) = 1 - P(\bar{A}|B)$
4. $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B)$
5. $B \subset A \Rightarrow P(A|B) = 1$
6. $A \subset B \Rightarrow P(A|B) = \frac{P(A)}{P(B)}$ ($\Leftrightarrow P(A) = P(B) \cdot P(A|B)$)

Bemerkung:

- Bedingung B bleibt immer „fest“, d.h. wird nicht variiert
- Satz macht keine Aussage für sich verändernde Bedingungen
- i.A. gilt nicht:
 - $P(B|A) = P(A|B)$
 - $P(A|B) = 1 - P(A|\bar{B})$
- Umstellen der Formel für bedingte Wahrscheinlichkeit liefert die sogenannte Multiplikationsregel $P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$

Satz: (erweiterte Multiplikationsregel)

Sei $(\Omega, \mathfrak{a}, P)$ ein Wahrscheinlichkeitsraum und $A_1, \dots, A_n \in \mathfrak{a}$ mit $P(A_1 \cap \dots \cap A_n) > 0$. Dann gilt

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

Beweis: vollständige Induktion

- Induktionsstart: $P(A_1 \cap A_2) = P(A_2|A_1)P(A_1)$ ($n = 2$)
- Induktionsvoraussetzung für $n - 1$, d.h. $P(A_1 \cap \dots \cap A_{n-1}) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\dots P(A_{n-1}|A_1 \cap \dots \cap A_{n-2})$
- Induktionsschritt: $P(A_1 \cap \dots \cap A_n) = P(A_1 \cap \dots \cap A_{n-1} \cap A_n) = P(A_n|A_1 \cap \dots \cap A_{n-1})P(A_1 \cap \dots \cap A_{n-1})$
 $= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\dots P(A_{n-1}|A_1 \cap \dots \cap A_{n-1})P(A_n|A_1 \cap \dots \cap A_{n-1})$

Definition:

Sei $(\Omega, \mathfrak{a}, P)$ ein Wahrscheinlichkeitsraum. Die Ereignisse $A_1, \dots, A_n \in \mathfrak{a}$ bilden vollständiges Ereignissystem, falls $A_i \cap A_j = \emptyset$ ($i \neq j$) und $\bigcup_{i=1}^n A_i = \Omega$ gilt.

Satz: (Satz von der totalen Wahrscheinlichkeit)

Sei $(\Omega, \mathfrak{a}, P)$ ein Wahrscheinlichkeitsraum, $A_1, \dots, A_n \in \mathfrak{a}$ ein vollständiges Ereignissystem und $B \in \mathfrak{a}$. Dann gilt

$$P(B) = \sum_{i=1}^n P(B|A_i) \cdot P(A_i)$$

Satz: (Satz von BAYES)

Sei $(\Omega, \mathfrak{a}, P)$ ein Wahrscheinlichkeitsraum, $A_1, \dots, A_n \in \mathfrak{a}$ ein vollständiges Ereignissystem und $B \in \mathfrak{a}$ mit $P(B) > 0$. Dann gilt für $i = 1, \dots, n$:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^n P(B|A_j)P(A_j)}$$

Beispiel: Haschisch=Einstiegsdroge?

aus Polizeibericht bekannt: von 300 Personen haben 96 Haschisch geraucht, 10 waren heroinabhängig und 6 Personen haben Haschisch geraucht und heroinabhängig.

A_1 =Haschisch, A_2 =kein Haschisch, B =heroinabhängig

$$P(A_1) = \frac{96}{300}, P(A_2) = \frac{204}{300} \quad P(B|A_1) = \frac{6}{96}, P(B|A_2) = \frac{4}{204}$$

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} = \frac{\frac{6}{96} \frac{96}{300}}{\frac{6}{96} \frac{96}{300} + \frac{4}{204} \frac{204}{300}} = \frac{6}{10} = 0,6$$

aber: $P(B|A_1) = \frac{6}{96} = 0,0625$

Beispiel: Spam-Filter

Analyse:

	Anteil Spam	Anteil gesamt
„***“ im Betreff = A_1	95% = $P(S A_1)$	5% = $P(A_1)$
„Sex“ im Text = A_2	68% = $P(S A_2)$	13% = $P(A_2)$
weder „***“ im Betreff noch „Sex“ im Text = A_3	18% = $P(S A_3)$	82% = $P(A_3)$

→ vollständiges Ereignissystem

gesucht: Wahrscheinlichkeit, dass eine Spam-Mail weder „***“ im Betreff noch „Sex“ im Text stehen hat.

$$P(A_3|S) = \frac{P(A_3 \cap S)}{P(S)} = \frac{P(S|A_3) \cdot P(A_3)}{P(S)} = 0,52$$

$$P(S) = P(S|A_1)P(A_1) + P(S|A_2)P(A_2) + P(S|A_3)P(A_3) = 0,2835$$

1.1.6 Stochastische Unabhängigkeit von Ereignissen

Definition:

Zwei Ereignisse $A, B \in \mathfrak{a}$ heißen *stochastisch unabhängig* (bezüglich des Maßes P), falls gilt $P(A \cap B) = P(A) \cdot P(B)$

Bemerkung:

1. Wenn A und B stochastisch unabhängig sind, so gilt

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

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

2. Man unterscheidet strikt zwischen Unvereinbarkeit ($A \cap B = \emptyset$) und Unabhängigkeit ($P(A \cap B) = P(A)P(B)$).

Satz: (Eigenschaften unabhängiger Ereignisse)

Sei $(\Omega, \mathfrak{a}, P)$ ein Wahrscheinlichkeitsraum. Dann gilt

1. \emptyset und Ω sind zu jedem $A \in \mathfrak{a}$ unabhängig
2. falls A und B unabhängig sind, so sind auch \bar{A} und B , A und \bar{B} sowie \bar{A} und \bar{B} unabhängig
3. Unvereinbare Ereignisse A und B mit $P(A) > 0$ und $P(B) > 0$ sind stets abhängig

Definition

Die zufälligen Ereignisse $A_1, \dots, A_n \in \mathfrak{a}$ heißen *paarweise unabhängig*, falls $P(A_i \cap A_j) = P(A_i)P(A_j) \quad \forall i, j \quad i \neq j$ gilt. $A_1, \dots, A_n \in \mathfrak{a}$ heißen unabhängig in der Gesamtheit, falls für $k \leq n$ und alle $i_1, i_2, \dots, i_k \in \{1, \dots, n\}$ gilt.

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

Bemerkung: aus der Unabhängigkeit in Gesamtheit folgt die paarweise Unabhängigkeit. Die Umkehrung gilt i.A. nicht!

1.2 Zufallsgrößen

Häufig sind Ergebnisse von Zufallsversuchen Zahlenwerte. Auch in anderen Fällen möchte man zur Charakterisierung der Ergebnisse von Zufallssituationen Zahlenwerte verwenden. Dies geschieht mit Hilfe von Zufallsgrößen X , indem jedem Ergebnis $\omega \in \Omega$ eine reelle Zahl $X(\omega)$ als Wert der Zufallsgröße zugeordnet wird.

Definition

Es sei $(\Omega, \mathfrak{a}, P)$ ein Wahrscheinlichkeitsraum zu einer festen Zufallssituation. Dann heißt eine Abbildung $X : \Omega \rightarrow \mathbb{R}$ Zufallsgröße (oder Zufallsvariable) aber $(\Omega, \mathfrak{a}, P)$, wenn für alle Intervalle $I \subset \mathbb{R}$ gilt.

$$\{\omega \in \Omega \mid X(\omega) \in I\} \in \mathfrak{a}$$

Wir schreiben $P(X \in I)$ für die Wahrscheinlichkeit $P(\{\omega \in \Omega : X(\omega) \in I\})$

Bemerkung: Bedingung der Messbarkeit, d.h. $\{\omega \in \Omega : X(\omega) \in I\} \in \mathfrak{a}$ sichert, dass die Wahrscheinlichkeit für dieses Ereignis auch definiert ist.

Beispiel: Roulette-Spiel, mehrfach, betrachten nur Rot/Schwarz (Null=Grün)

$$\Omega = \{(a_1, a_2, \dots) \text{ wobei } a_j \in \{„r“, „s“, „g“\}, j = 1, 2, \dots, \}$$

$$X(\omega) = \min\{j : a_j = „r“\} \quad (X = 5 \Rightarrow \text{beim 5. Versuch 1 mal „rot“})$$

1.2.1 Die Verteilungsfunktion

Definition:

Sei $(\Omega, \mathfrak{a}, P)$ ein Wahrscheinlichkeitsraum und X eine Zufallsgröße. Die Funktion $F_X(x) = P(\{\omega \in \Omega : X(\omega) < x\})$ die für alle $x \in (-\infty, \infty) = \mathbb{R}$ definiert ist, heißt *Verteilungsfunktion* der Zufallsgröße X .

Bemerkung:

- $F_X : \mathbb{R} \rightarrow \mathbb{R}$
- $P(\{\omega \in \Omega : X(\omega) < x\}) = P(\{\omega \in \Omega : X(\omega) \in (-\infty, x)\})$ ist wegen Messbarkeitsforderung $\{\omega \in \Omega : X(\omega) \in I\} \in \mathfrak{a}$ immer definiert.
- in Literatur auch manchmal: $F_X(x) = P(\{\omega \in \Omega : X(\omega) \leq x\}) \rightarrow$ andere Eigenschaften
- $F_X(x_0)$ gibt die Wahrscheinlichkeit an, dass X Werte (echt) kleiner als x_0 annimmt

1.2 Zufallsgrößen

Beispiel: (idealer) Würfel $\Omega = \{\omega_1, \dots, \omega_6\}$ mit $w_i = i, i = 1 \dots 6, \mathfrak{a} = \mathcal{P}(\Omega)$

$$F_X(-3) = P(X < -3) = 0$$

$$F_X(2) = P(X < 2) = 0$$

$$F_X(2, 1) = P(X < 2, 1) = P(\{\omega_1\}) = \frac{1}{6}$$

$$F_X(4, 01) = P(X < 4, 01) = P(\{\omega_1, \omega_2\}) = \frac{2}{6} = \frac{1}{3}$$

Satz: (Eigenschaften der Verteilungsfunktion)

Sei $(\Omega, \mathfrak{a}, P)$ ein Wahrscheinlichkeitsraum und X eine Zufallsgröße. Dann gilt für $a, b \in \mathbb{R}$ ($a < b$)

1.
 - $P(X < b) = P(\{\omega \in \Omega : X(\omega) < b\}) = F_X(b)$
 - $P(X \geq a) = P(\{\omega \in \Omega : X(\omega) \geq a\}) = 1 - F_X(a)$
 - $P(a \leq X < b) = P(\{\omega \in \Omega : a \leq X(\omega) < b\}) = F_X(b) - F_X(a)$
2. $0 \leq F_X(x) \leq 1 \quad \forall x \in \mathbb{R}$
3. $F(x_1) \leq F(x_2) \quad \forall x_1, x_2$ mit $x_1 \leq x_2$, d.h. F ist monoton wachsend
4. $\lim_{x \rightarrow -\infty} F(x) = 0$ und $\lim_{x \rightarrow \infty} F(x) = 1$
5. F ist linksseitig stetig, d.h. $\lim_{x \rightarrow x_0 - 0} F(x) = F(x_0)$

Satz:

Eine beliebige Funktion $F(x)$ mit den Eigenschaften 2-5 (vorheriger Satz) ist eine Verteilungsfunktion einer gewissen Zufallsgröße X .

Bemerkung:

- In vielen Situationen ist man gar nicht an der konkreten Abbildung $X : \Omega \rightarrow \mathbb{R}$ interessiert (oder diese Abbildung ist nicht zugänglich oder bekannt) und man verwendet nur die Verteilungsfunktion zur Beschreibung der Zufallsgröße
- bei Definition mit $F(x) = P(X \leq x)$ ist Verteilungsfunktion F rechtsseitig stetig (restliche Eigenschaften 2-5 bleiben)

1.2.2 Diskrete Zufallsgrößen

Definition:

Eine Zufallsgröße X heißt diskret, wenn sie nur endlich oder abzählbar unendlich viele Werte annehmen kann. Offenbar kann eine diskrete Zufallsgröße beschrieben werden, wenn die Werte von X und die zugehörigen Wahrscheinlichkeiten bekannt sind.

Definition:

Sei X eine diskrete Zufallsgröße mit den Werten x_1, x_2, x_3, \dots . Die Funktion

$$p(x) = P(\{\omega \in \Omega : X(\omega) = x\}) = P(X = x)$$

heißt Wahrscheinlichkeitsfunktion von X .

Satz:

Sei X eine diskrete Zufallsgröße mit den Werten x_1, x_2, \dots und den entsprechenden Wahrscheinlichkeiten $p_i = p(x_i) = P(X = x_i)$. Dann gilt für die Verteilungsfunktion

$$F(x) = P(X < x) = \sum_{x_i < x} p_i$$

Bemerkung: Die Verteilungsfunktion einer diskreten Zufallsgröße ist eine (monoton wachsende) Treppenfunktion, die an den Stellen x_i Sprünge der Höhe $p_i = p(x_i) = P(X = x_i)$ aufweist.

Beispiel: $\Omega = \{\omega_1, \omega_2\}$, $\mathfrak{a} = \mathcal{P}(\Omega) = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \Omega\}$, $P(\{\omega_1\}) = \frac{1}{2} = P(\{\omega_2\})$

Wahrscheinlichkeitsfunktion: $p_1 = p(x_1) = p(0) = P(X = 0) = \frac{1}{2}$, $p_2 = p(x_2) = p(1) = P(X = 1) = \frac{1}{2}$

Verteilungsfunktion: $F(x) = P(X < x) = \sum_{i: x_i < x} p_i$

$$\rightarrow \begin{cases} x \in (-\infty, 0] & F(x) = 0 & (= P(\emptyset)) \\ x \in (0, 1] & F(x) = p_1 = \frac{1}{2} & (= P(\{\omega_1\})) \\ x \in (1, \infty) & F(x) = p_1 + p_2 = 1 & (= P(\Omega)) \end{cases}$$

1.2.3 Stetige Zufallsgrößen

Eine Zufallsgröße heißt stetig, wenn ihre Verteilungsfunktion stetig ist.

Definition:

Eine Zufallsgröße heißt (absolut) stetig, wenn eine integrierbare Funktion $f(t)$ ($t \in \mathbb{R}$) existiert, so dass für die Verteilungsfunktion $F_X(x)$ von X gilt

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

Die Funktion $f(t)$ heißt Dichtefunktion oder Verteilungsdichte der Zufallsgröße X . Für (absolut) stetige Zufallsgrößen gilt also

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

Beispiel: Gleichverteilung

$a < b$, $a, b \in \mathbb{R}$ fest. Die Zufallsgröße heißt gleichverteilt auf dem Intervall (a, b) , wenn ihre Dichtefunktion $f(t)$ mittels $f(t) = \begin{cases} \frac{1}{b-a} & t \in [a, b] \\ 0 & \text{sonst} \end{cases}$ gegeben ist. Dann gilt für die Verteilungsfunktion

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

Satz:

Die Dichtefunktion $f(t)$ einer stetigen Zufallsgröße X hat die Eigenschaften

- $f(t) \geq 0$
- $\int_{-\infty}^{\infty} f(t) dt = 1$

Satz: (Umkehrung)

Jede integrierbare Funktion f , die die Eigenschaften aus dem vorherigen Satz hat, ist Dichtefunktion einer gewissen stetigen Zufallsgröße.

Bemerkung:

Ist $f(t)$ an der Stelle t_0 stetig, dann gilt: $F'(t_0) = f(t_0)$. Allerdings muss $f(t)$ nicht stetig sein.

Für eine stetige Zufallsgröße gilt

- $P(a \leq X < b) = F(b) - F(a) = \int_{-\infty}^b f(t) dt - \int_{-\infty}^a f(t) dt = \int_a^b f(t) dt$
- $P(X = a) = P(a \leq X \leq a) = \int_a^a f(t) dt = 0$
- $P(a \leq X < b) = P(a < X < b) = P(a \leq X \leq b) = P(a < X \leq b)$

1.2.4 Erwartungswert und Varianz

Allgemein ist der Erwartungswert als Integral (Lebesgue-Integral) bezüglich dem Wahrscheinlichkeitsmaß P definiert.

$$EX = \int_{\Omega} X dP = \int_{\Omega} X(\omega)P(d\omega)$$

Für diskrete und (absolut) stetige Zufallsgrößen hat der Erwartungswert folgende Form:

Definition:

Sei X eine diskrete Zufallsgröße mit den Werten x_1, x_2, \dots mit Wahrscheinlichkeitsfunktion p ($p_x = p(x_k) = P(X = x_k)$). Dann heißt

$$EX = \sum_k x_k \cdot p(x_k) = \sum_k x_k \cdot p_k = \sum_k x_k \cdot P(X = x_k)$$

Erwartungswert der diskreten Zufallsgröße X . Der Erwartungswert ist eine endliche reelle Zahl, falls gilt $\sum_k |x_k|p_k < \infty$

Definition:

Sei X eine (absolut) stetige Zufallsgröße mit Dichte $f(x)$. Dann heißt

$$EX = \int_{-\infty}^{\infty} x f(x) dx$$

Erwartungswert der stetigen Zufallsgröße X . Der Erwartungswert ist eine endlich reelle Zahl, falls $\int_{-\infty}^{\infty} |x|f(x) dx < \infty$

1.2 Zufallsgrößen

Satz: (Eigenschaften des Erwartungswertes)

Seien X und Y Zufallsgrößen (beliebig) und $c \in \mathbb{R}$ (konstant). Dann gilt:

1. $Ec = c$
2. $E(cX) = cEX$
3. $E(X + Y) = EX + EY$

Beweis: (nur für diskrete Zufallsgrößen)

1. $c \in \mathbb{R}$ entspricht (entartet) Zufallsgröße mit nur einem Wert, d.h. $\frac{z_n}{p(z_n)} \mid \frac{c}{1}$
 $\Rightarrow Ec = c \cdot 1 = c$

2. X habe Wahrscheinlichkeit p , $z := cX$

X	x_1	x_2	x_3	\dots	Z	$z_1 = cx_1$	$z_2 = cx_2$	$z_3 = cx_3$	\dots
$p(x_n)$	p_1	p_2	p_3	\dots	$p(z_n)$	p_1	p_2	p_3	\dots

$$\Rightarrow E(cX) = EZ = \sum_k z_k \cdot p_k = \sum_k c \cdot x_k \cdot p_k = c \cdot \underbrace{\sum_k x_k p_k}_{=EX} = c \cdot EX$$

- 3.

$$\begin{aligned} E(X + Y) &= \sum_i \sum_j (x_i + y_j) P(X = x_i \wedge Y = y_j) \\ &= \sum_i x_i \sum_j P(X = x_i \wedge Y = y_j) + \sum_j y_j \sum_i P(X = x_i \wedge Y = y_j) \\ &= \sum_i x_i P(X = x_i) + \sum_j y_j P(Y = y_j) \\ &= EX + EY \end{aligned}$$

Satz:

Sei X eine Zufallsgröße und $g : \mathbb{R} \rightarrow \mathbb{R}$ eine Funktion. Dann ist $Y = g(X)$ eine Zufallsgröße und es gilt

$$EY = Eg(X) = \begin{cases} \sum_k g(x_k) \cdot P(X = x_k) = \sum_k g(x_k) \cdot p(x_k) & \text{falls } X \text{ diskrete Zufallsgröße} \\ \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx & \text{falls } X \text{ stetige Zufallsgröße} \end{cases}$$

Der Erwartungswert einer Zufallsgröße gibt an, um welchen Wert die Realisierungen der Zufallsgröße schwanken.

Definition:

Sei X eine Zufallsgröße mit $EX < \infty$. Der Erwartungswert von $(X - EX)^2$ heißt Varianz der Zufallsgröße X und wird mit $D^2X = E(X - EX)^2$ bezeichnet. Die Zahl $\delta = \sqrt{D^2X}$ heißt Standardabweichung der Zufallsgröße X .

Bemerkung: aus vorherigem Satz folgt mit $\mu = EX$ und $g(X) = (X - \mu)^2$

$$D^2X = Eg(X) = \begin{cases} \sum_k (x_k - \mu)^2 \cdot p(x_k) = \sum_k (x_k - EX)^2 p(x_k) & X \text{ diskret} \\ \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f_X(x) dx = \int_{-\infty}^{\infty} (x - EX)^2 f(x) dx & X \text{ stetig} \end{cases}$$

Satz: (Eigenschaften der Varianz)

Sei X eine Zufallsgröße und $c \in \mathbb{R}$ (konstant). Dann gilt

1. $D^2X = EX^2 - (EX)^2$
2. $D^2c = 0$
3. $D^2(cX) = c^2 D^2X$
4. $D^2(c + X) = D^2(X)$

Bemerkung: i.A. gilt nicht $D^2(X + Y) = D^2X + D^2Y$ (gilt nur für unabhängige Zufallsgrößen)

$$D^2(X + Y) = E(X + Y - E(X + Y))^2 = E((X - EX) + (Y - EY))^2 = E((X - EX)^2 + (Y - EY)^2 + 2(X - EX)(Y - EY)) = D^2XD^2Y + 2E(X - EX)(Y - EY)$$

Definition:

Sei X eine Zufallsgröße. Dann heißt $m_k = EX^k$, $k = 1, 2, \dots$ k -te Momente der Zufallsgröße X und $\mu_k = E(X - EX)^k$, $k = 1, 2, \dots$ k -te zentrale Momente der Zufallsgröße X .

Bemerkung:

- $\mu_1 = 0$ (1. zentrales Moment)
- $m_1 = EX$
- $\mu_2 = E(X - EX)^2 = D^2X = EX^2 - (EX)^2 = m_2 - m_1^2$
- um die Asymmetrie von Verteilungen zu untersuchen, verwendet man oft

$$\gamma_1 = \frac{\mu_3}{\sqrt{m_2^3}} = \frac{E(X - EX)^3}{(D^2X)^{\frac{3}{2}}}$$

bzw. $\gamma_2 = \beta_2 - 3$ (Exzess) (Maß für die Wölbung)

Definition:

Sei X eine Zufallsgröße und $p \in (0, 1)$ eine reelle Zahl. Als Quantil der Ordnung p bezeichnet man die Zahl Q_p mit

$$F_X(Q_p) \leq p \leq F_X(Q_p + 0)$$

Bemerkung:

- X diskrete Zufallsgröße: linksseitig stetige Verteilungsfunktion
- i.A. Median \neq Erwartungswert
- X stetige Zufallsgröße, F stetig gilt $F(Q_p) = p$

Definition:

Eine Zufallsgröße X heißt standardisiert, falls $EX = 0$ und $D^2X = 1$ gilt.

Bemerkung:

Die Zufallsgröße $Y = \frac{X - EX}{\sqrt{D^2X}}$ für $D^2X \neq 0$ ($EX < \infty$, $D^2X < \infty$) ist standardisiert (normiert).

$$EY = E\left(\frac{X - EX}{\sqrt{D^2X}}\right) = \frac{1}{\sqrt{D^2X}}(EX - EX) = 0$$

$$D^2Y = D^2\left(\frac{X - EX}{\sqrt{D^2X}}\right) = \frac{1}{D^2X}D^2(X - EX) = \frac{1}{D^2X}D^2X = 1$$

Satz: (Tschebyscheff-Ungleichung)

Sei X eine Zufallsgröße. Dann gilt

$$P(|X - EX| > \varepsilon) < \frac{D^2X}{\varepsilon^2} \quad \forall \varepsilon > 0$$

Folgerung:

Für jedes $k > 0$ gilt $P(|X - EX| > k\sigma) < \frac{D^2X}{k^2\sigma^2} = \frac{1}{k^2}$ bzw. $P(|X - EX| \leq k\sigma) \geq 1 - \frac{1}{k^2}$

z.B. $k = 3$

$$P(EX - 3\sigma \leq X \leq EX + 3\sigma) \geq 1 - \frac{1}{9} = \frac{8}{9} \approx 0,889$$

Interpretation: die Wahrscheinlichkeit dafür, dass eine beliebige Zufallsgröße (mit $EX < \infty$, $D^2X < \infty$) einen Wert annimmt, der um höchstens 3σ von ihrem Erwartungswert abweicht, beträgt mindestens 89% (allgemeine 3σ -Regel)

Bemerkung:

Die Tschebyscheff-Ungleichung gilt für beliebige Zufallsgrößen. Es genügt EX und D^2X zu kennen, um die Wahrscheinlichkeit der Abweichung abzuschätzen. Hat man genauere Verteilungsaussagen, sind genauere Abschätzungen zu erwarten oder $P(|X - EX| > \varepsilon)$ kann sogar exakt berechnet werden.

1.2.5 Spezielle Verteilungen

Diskrete Zufallsgröße:

Binomialverteilung:

n -malige Wiederholung von Bernoulli-Experiment. Bernoulli k -ter Versuch:

$$X_k(\omega) = \begin{cases} 1 & \text{falls } \omega = \omega_1 \\ 0 & \text{falls } \omega = \omega_2 \end{cases}$$

$$P(X_k = 1) = p, P(X_k = 0) = 1 - p$$

Wir definieren $X = \sum_{k=1}^n X_k = \text{Anzahl der Erfolge bei } n \text{ Wiederholungen} \rightarrow X = k \Leftrightarrow$

Erfolg tritt genau k mal bei n Versuchen ein. Offenbar gibt es $\binom{n}{k}$ Möglichkeiten

$$\text{dafür} \Rightarrow P(X = k) = \binom{n}{k} \cdot p^k (1 - p)^{n-k}$$

Definition:

Eine Zufallsgröße X mit der Wahrscheinlichkeitsfunktion

$$P(X = k) = \binom{n}{k} \cdot p^k (1 - p)^{n-k} \quad (k = 0, 1, \dots, n)$$

heißt binomialverteilt mit den Parametern n (Freiheitsgrade) und p (Erfolgswahrscheinlichkeit oder Fehlerquote)

Schreibweise: $X \sim B(n, p)$

Bemerkung:

wesentlich beim Bernoullischem Versuchsschema sind:

- die Unabhängigkeit der n Versuche
- die stets gleichbleibende (Erfolgs-)Wahrscheinlichkeit p

Satz:

Sei $X \sim B(n, p)$. Dann gilt $EX = n \cdot p$ und $D^2X = n \cdot p \cdot (1 - p)$

Beispiel:

- Prüfung mit 10 Fragen: jeweils 3-mögliche Antworten (genau eine jeweils richtig)
- Prüfung bestanden, wenn mindestens 5 richtige Antworten

gesucht: Wahrscheinlichkeit für Bestehen, wenn zufällig angekreuzt wird

$$X_k = \begin{cases} 1 & k\text{-te Frage richtig} \\ 0 & \text{sonst} \end{cases} \Rightarrow p = P(X_k = 1) = \frac{1}{3}$$

$$X = \sum_{k=1}^{10} X_k = \text{Anzahl richtiger Antworten} \quad X \sim B(n, p), \quad n = 10, \quad p = \frac{1}{3}$$

$$P(X = k) = \binom{10}{k} \cdot \left(\frac{1}{3}\right)^k \cdot \left(\frac{2}{3}\right)^{n-k}$$

$$\Rightarrow P(X \geq 5) = P(X = 5) + \dots + P(X = 10) = 0,2131$$

Geometrische Verteilung:

$$\text{Bernoulli-Experiment: } X_p(\omega) = \begin{cases} 1 & \omega = \omega_1 \quad p = P(X_k = 1) \\ 0 & \omega = \omega_2 \end{cases}$$

$X = k \Leftrightarrow$ beim k -ten Versuch tritt Erfolg erstmalig ein, d.h. $\{X_1 = 0\} \cap \dots \cap \{X_{k-1} = 0\} \cap \{X_k = 1\}$

$$\Rightarrow P(X = k) = (1 - p)^{k-1} \cdot p, \quad k = 1, 2, \dots$$

$$\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} p \cdot (1 - p)^{k-1} = p \cdot \sum_{k=0}^{\infty} (1 - p)^k = p \cdot \frac{1}{1 - (1 - p)} = 1$$

Definition:

Eine Zufallsgröße X mit der Wahrscheinlichkeit $P(X = k) = (1-p)^{k-1} \cdot p$, $k = 1, 2, \dots$ für $p \in (0, 1]$ heißt geometrisch verteilt, mit dem Parameter p .

Satz:

Sei X geometrisch verteilt mit $p \in (0, 1]$. Dann gilt $EX = \frac{1}{p}$ und $D^2X = \frac{1-p}{p^2}$

Hypergeometrische Verteilung:

Referenzmodell Urne mit N Kugeln, davon M schwarze, $N - M$ weiße \rightarrow ziehen ohne Zurücklegen der n Kugeln

X = Anzahl schwarzer Kugeln

$$\Rightarrow P(X = m) = \frac{\binom{M}{m} \cdot \binom{N-M}{n-m}}{\binom{N}{n}} \quad m = \max \{0, n - (N - M)\}, \dots, \min \{n, M\}$$

Definition:

Eine Zufallsgröße mit obiger Wahrscheinlichkeitsfunktion heißt hypergeometrisch verteilt. Schreibweise: $X \sim H(n, N, M)$

Satz:

Sei $X \sim H(n, N, M)$. Dann gilt $EX = n \cdot \frac{M}{N}$ und $D^2X = n \cdot \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}$.

Bemerkung:

$Y \sim B(n, p) \rightarrow EY = p \cdot n$, $D^2Y = n \cdot p \cdot (1-p)$ = Urnenmodell mit Zurücklegen

$X \sim H(n, N, M) \rightarrow p = \frac{M}{N}$ = beim 1. Ziehen Wahrscheinlichkeit, dass schwarze Kugel gezogen wird

$$\Rightarrow EX = n \cdot p, \quad D^2X = n \cdot p \cdot (1-p) \frac{N-n}{N-1} \quad \text{Grenzfall } n = N \Rightarrow D^2X = 0$$

\rightarrow Unterschied groß, wenn $N \approx n$, da dann Faktor $\frac{N-n}{N-1}$ sehr klein. Für N sehr groß, kaum Unterschiede zwischen hypergeometrischer Verteilung und Binomialverteilung $H(n, N, M) \xrightarrow[N \rightarrow \infty]{p = \frac{M}{N}} B(n, p)$, d.h. für große N (und entsprechend kleine n) ist eine hypergeometrisch verteilte Zufallsgrößen annähernd binomialverteilt.

Poisson-Verteilung:

Definition:

Eine diskrete Zufallsgröße X mit Wahrscheinlichkeitsfunktion $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$
 $k = 0, 1, \dots$ heißt Poisson-verteilt mit dem Parameter $\lambda > 0$ $X \sim \pi(\lambda)$

Satz:

Sei $X \sim \pi(\lambda)$. Dann gilt $EX = \lambda$ und $D^2X = \lambda$.

Bemerkung:

1. $X \sim B(n, p)$, wenn n groß und p sehr klein, wird Binomialverteilung schnell unhandlich. Poisson-Verteilung ist aber gute Näherung in solchen Fällen ($p = \frac{1}{N}$, $\lambda = \frac{n}{N} = n \cdot p$) \rightarrow Grenzwertsätze
2. Poisson-Prozesse

Stetige Zufallsgrößen:

Gleichverteilung: (siehe Beispiel im Abschnitt 1.2.3)

Definition:

Eine stetige Zufallsgröße X mit Dichtefunktion

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

heißt gleichverteilt mit den Parametern a und b . $X \sim Gl[a, b]$.

Satz:

Für $X \sim Gl[a, b]$ gilt

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a, b] \\ 1 & x > b \end{cases} \quad EX = \frac{a+b}{2} \quad D^2X = \frac{(a-b)^2}{12}$$

Exponentialverteilung**Definition:**

Eine stetige Zufallsgröße X mit der Dichtefunktion

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

heißt exponentialverteilt mit dem Parameter $\lambda > 0$. $X \sim Exp(\lambda)$

Satz:

Sei $X \sim Exp(\lambda)$. Dann gilt

$$\begin{cases} 0 & x \leq 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases} \quad EX = \frac{1}{\lambda} \quad D^2 X = \frac{1}{\lambda^2}$$

Definition:

Ein Zufallsprozess $\{X_t, t \in [0, \infty)\}$, der zum Zeitpunkt t zählt, wie oft ein gewisses Ereignis eingetreten ist und obige Voraussetzung erfüllt, heißt *Poisson-Prozess* mit Intensität μ .

Satz:

Für einen Poisson-Prozess gilt:

$$P(X_t = k) = \frac{(\mu t)^k}{k!} \cdot e^{-\mu t} \quad k = 0, 1, \dots \quad (\lambda = \mu \cdot t)$$

d.h. $X_t \sim \pi(\mu t)$ ($\lambda_t = \mu \cdot t$)

Eigenschaften:

- $EX = \mu t = \lambda_t$
- $D^2 X_t = \mu t = \lambda_t$

Satz:

Sei $\{X_t, t \in [0, \infty)\}$ ein Poisson-Prozess mit Intensität μ . Die Zufallsgröße bezeichne die Zeit zwischen dem Auftreten von 2 Ereignissen. Dann ist Y exponentialverteilt, d.h. $Y \sim \text{Exp}(\mu)$.

Beweis:

Sei $y \geq 0$

$$\begin{aligned} F_Y(y) &= P(Y < y) = 1 - P(Y \geq y) \\ &= 1 - P(X_y = 0) = 1 - e^{-\mu y} \frac{(\mu y)^0}{0!} = 1 - e^{-\mu y} \\ \Rightarrow F_Y(y) &= \begin{cases} 1 - e^{-\mu y} & y \geq 0 \\ 0 & y < 0 \end{cases} \end{aligned}$$

Beispiel: Ein Server erhält im Mittel 120 Anfragen pro Stunde. Er ist überlastet, wenn in einer Minute 5 oder mehr Anfragen eintreffen. Gesucht ist die Wahrscheinlichkeit, das innerhalb einer Minute der Server überlastet ist.

X_t =Anzahl Serveranfragen in t Minuten \rightarrow Annahmen 1-3 (idealerweise) erfüllt \rightarrow Poisson-Prozess Intensität μ bekannt: $EX_{60} = 120 = \mu \cdot t = \mu \cdot 60 \Rightarrow \mu = 2$

$$\begin{aligned} P(X_1 \geq 5) &= 1 - P(X_1 < 5) = 1 - (P(X_1 = 0) + P(X_1 = 1) + \dots + P(X_1 = 4)) \\ &= 1 - e^{-\mu \cdot 1} \left(\frac{(\mu \cdot 1)^0}{0!} + \frac{(\mu \cdot 1)^1}{1!} + \dots + \frac{(\mu \cdot 1)^4}{4!} \right) \\ &= 1 - 7e^{-2} = 0,053 \end{aligned}$$

Wie groß ist die Wahrscheinlichkeit, dass zwischen 2 Anfragen mindestens 1 Minute liegt?

Y =Zeit zwischen 2 Anfragen $\sim \text{Exp}(2)$

$$P(Y \geq 1) = 1 - P(Y < 1) = 1 - F_Y(1) = 1 - (1 - e^{-\mu \cdot 1}) = e^{-\mu} = e^{-2} = 0,135$$

Bemerkung:

- Aufgrund des Zusammenhangs zwischen Poisson-Prozess und Exponentialverteilung sind viele zufällige Zeiten, wie z.B. Wartezeiten, Reparaturzeiten, Lebensdauer von (Bau)Teilen exponentialverteilt.
- Bei Lebensdauer ist zu beachten, dass keine „Spätausfälle“ aufgrund von Alterserscheinungen modelliert werden können (Grund: Annahmen 1-3)

Normalverteilung

Definition: Eine stetige Zufallsgröße X heißt normalverteilt (oder Gauß-verteilt) mit den Parametern μ und $\sigma^2 > 0$, wenn ihre Dichtefunktion der Gleichung

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

genügt. Schreibweise: $X \sim N(\mu, \sigma^2)$. Eine $N(0, 1)$ -verteilte Zufallsgröße heißt standardnormalverteilt.

Bemerkung: Für $N(0, 1)$ -verteilte Zufallsgrößen wird die Dichtefunktion häufig mit φ und die Verteilungsfunktion Φ bezeichnet, d.h.

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

sowie

$$\Phi(x) = \int_{-\infty}^x \varphi(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

Satz:

Sei $X \sim N(\mu, \sigma^2)$. Dann gilt

$$EX = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \mu \quad \text{und} \quad D^2X = \sigma^2$$

Bemerkung:

- Die Verteilungsfunktion einer $N(\mu, \sigma^2)$ -verteilten Zufallsgröße ist durch

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt$$

gegeben

- Die analytische Berechnung von $F(x)$ ist unmöglich (da sich die Stammfunktion zu $\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$ nicht durch elementare Funktionen aufschreiben lässt) \Rightarrow numerische Berechnung und Tabellierung
 \Rightarrow numerische Berechnung und Tabellierung für $\Phi(x)$ (d.h. $N(0, 1)$), da nicht für jedes Paar (μ, σ^2) möglich.

Für $X \sim N(\mu, \sigma^2)$ gilt

$$\begin{aligned} F_X(x) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\frac{x-\mu}{\sigma}} \sigma \exp\left(-\frac{\tilde{t}^2}{2}\right) d\tilde{t} \\ &= \Phi\left(\frac{x-\mu}{\sigma}\right) \end{aligned}$$

- oft ist $\Phi(x)$ nur für $x \geq 0$ tabelliert, da $\Phi(x) = 1 - \Phi(-x)$ gilt

Erinnerung: $Y = \frac{X - EX}{\sqrt{D^2X}} = \frac{X - \mu}{\sigma}$ ist normiert, d.h. $EY = 0$, $D^2Y = 1 \rightarrow$ falls X normalverteilt ist, so ist auch Y normalverteilt.

Satz:

Sei $X \sim N(\mu, \sigma^2)$. Dann ist $Y = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

Bemerkung: $X \sim N(\mu, \sigma^2)$

$$P(a < Y < b) = F_X(b) - F_X(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Bemerkung:

Als normalverteilt können Zufallsgrößen angesehen werden, die durch Überlagerungen einer großen Anzahl von (unabhängigen) Einflüssen entstehen, wobei jede Einflussgröße nur einen im Verhältnis zur Gesamtsumme unbedeutenden Betrag liefert. Diese Problematik werden wir später in Form des zentralen Grenzwertsatzes präzisieren.

Beispiele normalverteilter Zufallsgrößen sind

- zufällige Beobachtungen oder Messfehler
- zufällige Abweichungen vom Nennmaß bei der Fertigung von Werkstücken

Summe normalverteilter Zufallsgrößen**Satz** (Additionssatz)

Seien $X_i \sim N(\mu_i, \sigma_i^2)$ für $i = 1, \dots, n$ vollständig unabhängige normalverteilte Zufallsgrößen. Dann gilt

$$Z = \sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$$

Bemerkung: Zufallsgrößen X und Y heißen stochastisch unabhängig, falls $P(A \cap B) = P(A) \cdot P(B)$ mit $A = \{\omega \in \Omega : X(\omega) < x\}$ und $B = \{\omega \in \Omega : Y(\omega) < y\}$

$$\Leftrightarrow P(X < x, Y < y) = P(X < x) \cdot P(Y < y) \Leftrightarrow F_{(X,Y)}(x, y) = F_X(x) \cdot F_Y(y)$$

Beispiel:

Kern eines Transformators: 25 Bleche und dazwischen jeweils Isolierschichten (24 Stück)

- X_i =Dicke des i -ten Bleches, $X_i \sim N(0,8, 0,04^2)$
- Y_j =Dicke der j -ten Isolierschicht, $Y_j \sim N(0,2, 0,03^2)$

gesucht: Wahrscheinlichkeit, dass 2 Bleche und 1 Isolierschicht dicker als 1,85 mm sowie Wahrscheinlichkeit, dass Kern zu dick für Spulenöffnung (25,3 mm)

$$Z = X_1 + Y_1 + X_2 \sim (1,8, 0,0041)$$

$$\begin{aligned} \Rightarrow P(Z > 1,85) &= P\left(\frac{Z - \mu}{\sigma} > \frac{1,85 - \mu}{\sigma}\right) = P\left(\frac{Z - \mu}{\sigma} > \frac{0,05}{\sqrt{0,0041}}\right) \\ &= 1 - \Phi\left(\frac{0,05}{\sqrt{0,0041}}\right) = 1 - \Phi(0,78) = 1 - 0,78230 = 0,21770 \end{aligned}$$

$$Z = \sum_{i=1}^{25} X_i + \sum_{j=1}^{24} Y_j \sim N(24,8, 0,0616)$$

$$\Rightarrow P(Z > 25,3) = P\left(\frac{Z - \mu}{\sigma} > \frac{25,3 - 24,8}{\sqrt{0,0616}}\right) = \dots = 1 - \Phi(2,015) = 0,0222$$

1.3 Das Gesetz der großen Zahlen und Grenzwertsätze

Häufig treten Folgen Zufallsgrößen X_1, X_2, \dots, X_n und Linearkombinationen $Y_n = a_1 X_1 + a_2 X_2 + \dots + a_n X_n$ auf. Frage: Verteilung von Y_n ?

Definition:

Die Zufallsgrößen X_1, X_2, \dots, X_n heißen unabhängig und identisch verteilt (kurz: vom Typ i.i.d.) wenn sie vollständig unabhängig sind, identische Verteilungen aufweisen sowie Erwartungswert und Varianzen existieren (endlich sind). Für X_1, \dots, X_n i.i.d gilt also:

- $EX_1 = EX_2 = \dots = EX_n = \mu \in \mathbb{R}$
- $D^2 X_1 = D^2 X_2 = \dots = D^2 X_n = \sigma^2 < \infty$

arithmetische Mittel:

- $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \Rightarrow E\bar{X}_n = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n EX_i = \mu$
- $D^2 \bar{X}_n = D^2 \left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \left(\sum_{i=1}^n D^2 X_i\right) = \frac{1}{n^2} \sum_{i=1}^n D^2 X_i = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

Das Gesetz der großen Zahlen

Satz: (Schwache Gesetze der großen Zahlen)

Seien X_1, \dots, X_n vom Typ i.i.d mit $\mu = EX_i = E\bar{X}_n$. Dann gilt $\forall \varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \leq \varepsilon) = 1$$

d.h. das arithmetische Mittel \bar{X}_n konvergiert für wachsende n im Sinne der Wahrscheinlichkeit gegen den Erwartungswert der Zufallsgröße X_i .

Spezialfall:

relative Häufigkeit des Ereignisses A : $H_n(A) = \bar{X}_n$ mit

$$X_i = \begin{cases} 1 & \text{falls } A \text{ im } i\text{-ten Versuch eintritt} \\ 0 & \text{falls } A \text{ nicht im } i\text{-ten Versuch eintritt} \end{cases}$$

wobei $p = P(A) \Rightarrow EX_i = p = \mu$ d.h. der letzte Satz besagt

$$\lim_{n \rightarrow \infty} P(|H_n(A) - p| \leq \varepsilon) = 1$$

Grenzwertsätze

Satz: (Grenzwertsatz von Moivre-Laplace)

Seien X_1, \dots, X_n Zufallsgrößen mit

$$X_i = \begin{cases} 1 & A \text{ tritt im } i\text{-ten Versuch ein} \\ 0 & \text{sonst} \end{cases}$$

und $P(X_i = 1) = p$ und $P(X_i = 0) = 1 - p$. Dann ist $Y_n = X_1 + \dots + X_n \sim B(n, p)$ und es gilt für die zugehörige normierte Zufallsgröße

$$\tilde{Y}_n = \frac{Y_n - EY_n}{\sqrt{D^2Y_n}} = \frac{Y_n - np}{\sqrt{np(1-p)}}$$

$$\lim_{n \rightarrow \infty} P(\tilde{Y}_n < y) = \lim_{n \rightarrow \infty} F_n(y) = \Phi(y) \quad \forall y \in \mathbb{R}$$

→ für große n ist \tilde{Y}_n näherungsweise $N(0, 1)$ verteilt und somit Y_n näherungsweise $N(np, np(1-p))$ (Schreibweise: $Y_n \approx N(np, np(1-p))$)

Bemerkung:

1. praktische Umsetzung: wann ist n groß? Faustregel:
 - $np(1-p) > 9 \Rightarrow$ gute Näherung
 - $np(1-p) > 4 \Rightarrow$ brauchbare Näherung
2. $Y_n \sim B(n, p)$, dann $Y_n \approx N(np, p(1-p)n)$, offenbar

$$\begin{aligned} P(a \leq Y_n \leq b) &= P(a-1 < Y_n < b+1) \\ &\approx \Phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-np}{\sqrt{np(1-p)}}\right) \\ &\neq \Phi\left(\frac{a-1-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{b+1-np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

Es zeigt sich, dass die Methode der Stetigkeitskorrektur bessere Näherungen liefert

$$P(A \leq Y_n \leq b) = P\left(a - \frac{1}{2} < Y_n < b + \frac{1}{2}\right) \approx \Phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

1.3 Das Gesetz der großen Zahlen und Grenzwertsätze

analog:

$$P(a < Y_n < b) = P(a+1 \leq Y_n \leq b-1) = P \approx \Phi\left(\frac{b - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

Faustregel: $np(1-p) = 18,75 > 9 \Rightarrow$ GWS von Moivre-Laplace liefert gute Näherung $\Rightarrow Y_n \approx N(25, 18,75)$

exakt (Computer):

$$P(15 \leq Y_n \leq 30) = \sum_{k=15}^{30} \binom{100}{k} \cdot 0,25^k \cdot 0,75^{100-k} = 0,8908$$

GWS ohne Stetigkeitskorrektur

$$P(15 \leq Y_n \leq 30) \approx \Phi\left(\frac{30 - 25}{\sqrt{18,75}}\right) - \Phi\left(\frac{15 - 25}{\sqrt{18,75}}\right) = 0,8645$$

GWS mit Stetigkeitskorrektur

$$P(14,5 < Y_n < 30,5) \approx \Phi\left(\frac{30,5 - 25}{\sqrt{18,75}}\right) - \Phi\left(\frac{14,5 - 25}{\sqrt{18,75}}\right) = 0,8903$$

Zusammenfassung

$$Y_i = \sum_{i=1}^n X_i \sim B(n, p) \approx N(np, np(1-p))$$

wobei X_1, \dots, X_n i.i.d mit $X_i = \begin{cases} 1 & \text{falls } A \text{ eintritt} \\ 0 & \text{sonst} \end{cases} \quad p = P(A)$

Satz: (Zentraler GWS von Lindberg/Levy)

Sei X_1, \dots, X_n eine Folge von Zufallsgrößen vom Typ i.i.d mit $\mu = EX_i \in \mathbb{R}$ und $\sigma^2 = D^2X_i \in (0, \infty)$ und $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ sowie

$$\bar{Y}_n = \frac{\bar{X}_n - E\bar{X}_n}{\sqrt{D^2\bar{X}_n}} = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n} \frac{(\bar{X}_n - \mu)}{\sigma}$$

und $F_n(y) = P(\bar{Y}_n < y)$ die Verteilungsfunktion des standardisierten arithmetischen Mittels. Dann gilt

$$\lim_{n \rightarrow \infty} F_n(y) = \Phi(y) \quad \forall y \in \mathbb{R}$$

Bemerkung:

- Das arithmetische Mittel einer Folge von beliebig verteilten Zufallsgrößen (i.i.d) kann in guter Näherung als normalverteilt angenommen werden. Somit motiviert der zentrale GWS die Annahme, dass eine durch Überlagerung zahlreicher unabhängiger Einzeleinflüsse entstehende Zufallsgröße (z.B. Messfehler beim wiederholten Messen) als normalverteilt aufgefasst werden kann → Sonderstellung der Normalverteilung.
- es gibt zahlreiche Verallgemeinerungen des zentralen GWS - z.B. kann die Annahme der identischen Verteilungen fallen gelassen werden, wenn sichergestellt ist, dass keine der Einzelgrößen X_1, \dots, X_n zu großen Einfluss auf die Summe hat. Auch die Unabhängigkeit kann leicht abgeschwächt werden.

Grenzverteilungssatz von Poisson

Satz:

Gegeben sei eine Folge $Y_n \sim B(n, p_n)$ mit $p_n \xrightarrow{n \rightarrow \infty} 0$ und $np_n \xrightarrow{n \rightarrow \infty} \lambda$. Dann gilt

$$\lim_{n \rightarrow \infty} P(Y_n = k) = \pi_\lambda(k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

→ Zufallsgrößen Y_n sind asymptotisch Poisson-verteilt mit Parameter λ .

Bemerkung: Faustregel → Grenzverteilungssatz von Poisson liefert gute Approximation falls $np \leq 10$ und $1500p \leq n$.

1.4 Mehrdimensionale Verteilungen

bisher: eindimensionale (reellwertige) Zufallsgrößen, also Modellierung eines Merkmals eines benachbarten Objektes

Häufig braucht man jedoch ein Modell mit mehreren Merkmalen → mehrdimensionale Zufallsgröße/Zufallsvektoren.

Definition:

Seien X_1, \dots, X_n Zufallsgrößen. Dann heißt $\vec{X} = (X_1, \dots, X_n)$ n -dimensionaler Zufallsvektor. Die Verteilungsfunktion von \vec{X} ist definiert durch

$$F_{\vec{X}}(x_1, x_2, \dots, x_n) = P(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n)$$

(d.h. $F_{\vec{X}}(x_1, \dots, x_n)$ ist die Wahrscheinlichkeit, dass $X_1 < x_1$ und $X_2 < x_2$... und $X_n < x_n$ gelten).

Im Folgenden werden wir den Fall $n = 2$ behandeln, d.h. $\vec{X} = (X_1, X_2)$ (oder $\vec{X} = (X, Y)$). Die meisten Aussagen gelten dann analog für allgemeines n .

Definition:

Für einen zufälligen Vektor $\vec{X} = (X, Y)$ heißen die Funktionen

$$F_X(x) = \lim_{y \rightarrow \infty} F_{\vec{X}}(x, y) \quad \text{und} \quad F_Y(y) = \lim_{x \rightarrow \infty} F_{\vec{X}}(x, y)$$

Randverteilungen von X bzw. Y . Die zugehörigen Wahrscheinlichkeits- bzw. Dichtefunktionen werden ebenfalls mit Randverteilung bezeichnet.

allgemein:

$$F_{X_i}(x_i) \lim_{x_1 \rightarrow \infty} \cdot \lim_{x_2 \rightarrow \infty} \cdot \dots \cdot \lim_{x_{i-1} \rightarrow \infty} \cdot \lim_{x_{i+1} \rightarrow \infty} \cdot \dots \cdot \lim_{x_n \rightarrow \infty} F_{\vec{X}}(x_1, \dots, x_n)$$

Satz: (Eigenschaften der Verteilungsfunktion falls $n = 2$)

Sei $\vec{X} = (X, Y)$ ein Zufallsvektor mit Verteilungsfunktion $F_{\vec{X}}(x, y) = P(X < x, Y < y)$. Dann gilt

1. $0 \leq F(x, y) \leq 1 \quad \forall x, y \in \mathbb{R}$
2. $\lim_{x \rightarrow -\infty} F(x, y) = 0 \quad \forall y \in \mathbb{R}$ und $\lim_{y \rightarrow -\infty} F(x, y) = 0 \quad \forall x \in \mathbb{R}$
3. $F_{\vec{X}}(x) = \lim_{y \rightarrow \infty} F_{\vec{X}}(x, y) = P(X < x) \quad \forall x \in \mathbb{R}$ und $F_{\vec{Y}}(y) = \lim_{x \rightarrow \infty} F_{\vec{X}}(x, y) = P(Y < y) \quad \forall y \in \mathbb{R}$ d.h. Randverteilungen F_X und F_Y sind Verteilungsfunktionen der eindimensionalen Zufallsgröße X und Y .
4. $\lim_{\substack{x \rightarrow \infty \\ y \rightarrow \infty}} F_{\vec{X}}(x, y) = 1$
5. $F_{\vec{X}}(x, y)$ ist in beiden Komponenten monoton wachsend: $x_1 < x_2 \Rightarrow F(x_1, y) < F(x_2, y) \quad \forall y \in \mathbb{R}$, $y_1 < y_2 \Rightarrow F(x, y_1) < F(x, y_2) \quad \forall x \in \mathbb{R}$
6. $F_{\vec{X}}(x, y)$ ist in beiden Komponenten linksseitig stetig, d.h.

$$\lim_{h \rightarrow 0} F(x - h, y) = F(x, y) = \lim_{h \rightarrow 0} F(x, y - h) \quad \forall x, y \in \mathbb{R}$$

Definition:

Die Zufallsgrößen X und Y heißen stochastisch unabhängig, falls $\forall x, y \in \mathbb{R}$ gilt:

$$F_{\bar{X}}(x, y) = P(X < x, Y < y) = P(X < x) \cdot P(Y < y) = F_X(x) \cdot F_Y(y)$$

Definition:

Seien X und Y zwei Zufallsgrößen. Dann heißt

$$\text{cov}(X, Y) = E(X - EX)(Y - EY)$$

Kovarianz der Zufallsgröße X und Y . Wenn $\text{cov}(X, Y) = 0$ gilt, heißen Zufallsgrößen X und Y unkorreliert.

Satz:

Seien X und Y Zufallsgrößen mit $E|x| < \infty$, $E|Y| < \infty$ und $D^2X < \infty$, $D^2Y < \infty$. Dann gilt:

1. $\text{cov}(X, Y) = \text{cov}(Y, X) = E(XY) - EX \cdot EY$
2. $\text{cov}(aX + b, cY + d) = ac \cdot \text{cov}(X, Y)$ $a, b, c, d \in \mathbb{R}$ konstant
3. $\text{cov}(X, X) = D^2X$
4. $D^2(X + Y) \geq E(X + Y - E(X + Y))^2 = E((X - EX) + (Y - EY))^2 = E((X - EX)^2 + 2(X - EX)(Y - EY) + (Y - EY)^2) = E(X - EX)^2 + 2E(X - EX)(Y - EY) + E(Y - EY)^2 = D^2X + 2\text{cov}(XY) + D^2Y$

Definition:

Der Korrelationskoeffizient ϱ_{xy} der Zufallsgrößen X und Y ist definiert durch

$$\varrho_{xy} = \frac{\text{cov}(X, Y)}{\sqrt{D^2X} \sqrt{D^2Y}}$$

Satz:

Seien X und Y Zufallsgrößen. Dann gilt für den Korrelationskoeffizienten

- $\varrho_{xy} = \varrho_{yx}$
- $|\varrho_{xy}| \leq 1$

Bemerkung:

Der Korrelationskoeffizient misst, wie stark eine lineare Beziehung zwischen X und Y vorliegt. ($Y = aX + b$ ($a \neq 0$, $a, b \in \mathbb{R}$) \Rightarrow dann gilt $\text{cov}(X, Y) = \text{cov}(X, aX + b) = a \cdot \text{cov}(X, X) = aD^2X$, $D^2Y = D^2(aX + b) = a^2D^2X$

$$\Rightarrow \rho_{xy} = \frac{aD^2X}{\sqrt{D^2X}\sqrt{a^2D^2X}} = \frac{aD^2X}{|a|D^2X} = \begin{cases} +1 & a > 0 \\ -1 & a < 0 \end{cases}$$

Satz:

Sind X und Y stochastisch unabhängig, so gilt

1. $EXY = EX \cdot EY$
2. $\text{cov}(X, Y) = 0$
3. $D^2(X + Y) = D^2X + D^2Y$
4. $\rho_{xy} = 0$

Bemerkung:

X, Y unabhängig $\Rightarrow X, Y$ unkorreliert

aber aus Unkorreliertheit folgt i.A. nicht die Unabhängigkeit (das ist mehr!) da Korrelation nur linearen Zusammenhang misst (bei normalverteilten Zufallsgrößen gilt auch Umkehrung).

Fall $n > 2$: $\vec{X} = (X_1, \dots, X_n)^T \rightarrow$ Erwartungsvektor $\vec{\mu} = E\vec{X} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = \begin{pmatrix} EX_1 \\ \vdots \\ EX_n \end{pmatrix}$

\rightarrow Kovarianzmatrix $\vec{K} = \text{cov}\vec{X} = (\text{cov}(X_i, X_j))_{\substack{i=1\dots n \\ j=1\dots n}} = E(X - E\vec{X})(X - E\vec{X})^T$

\vec{K} ist symmetrisch $\vec{K}^T = \vec{K}$ und positiv definit ($\forall \vec{v} \in \mathbb{R}^n$ gilt: $\vec{v}^T \vec{K} \vec{v} = (\vec{K}\vec{v}, \vec{v}) \geq 0$)

Definition:

Ein zufälliger Vektor $\vec{X} = (X_1, \dots, X_n)^T$ heißt diskret verteilt, wenn alle eindimensionalen Randverteilungen diskrete Verteilungen sind.

Definition:

Ein zufälliges $\vec{X} = (X_1, \dots, X_n)$ heißt stetig verteilt und mit der Verteilungsfunktion F , wenn es eine integrierbare Dichtefunktion f mit

$$F(x_1, x_2, \dots, x_n) = P(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n) \\ = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f(t_1, t_2, \dots, t_n) dt_1 dt_2 \dots dt_n$$

gibt.

für $n = 2$: $\vec{X} = (X, Y)$

$$\text{Randdichte } f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy, f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

$$\Rightarrow EX = \int_{-\infty}^{\infty} x f_X(x) dx = \mu_x, EY = \int_{-\infty}^{\infty} y f_Y(y) dy = \mu_y$$

$$D^2X = \int_{-\infty}^{\infty} (x - \mu_x)^2 f_X(x) dx \quad D^2Y = \int_{-\infty}^{\infty} (y - \mu_y)^2 f_Y(y) dy$$

$$\text{cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x, y) dx dy$$

Satz:

Sei $\vec{X} = (X, Y)$ ein normalverteilter Zufallsvektor. Dann sind X und Y genau dann stochastisch unabhängig, wenn sie unkorreliert sind.

allgemein: $\vec{X} = (X_1, X_2, \dots, X_n)$ und X_i, X_j paarweise unkorreliert, d.h.

$$\text{cov}(X_i, X_j) = 0 \quad \forall i \neq j \quad \Rightarrow \vec{K} = \begin{pmatrix} \sigma_{x_1}^2 & & & \\ & \sigma_{x_2}^2 & & \\ & & \ddots & \\ & & & \sigma_{x_n}^2 \end{pmatrix} = \vec{K}^{-1} = \\ \begin{pmatrix} \frac{1}{\sigma_{x_1}^2} & & & \\ & \frac{1}{\sigma_{x_2}^2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_{x_n}^2} \end{pmatrix} \Rightarrow \det \vec{K}^{-1} = |\vec{K}^{-1}| = \frac{1}{\sigma_{x_1}^2} \cdot \frac{1}{\sigma_{x_2}^2} \cdot \dots \cdot \frac{1}{\sigma_{x_n}^2}$$

2 Einführung in die mathematische Statistik

beschreibende (deskriptive) Statistik	beurteilende (induktive/schließende) Statistik
Strukturierung und Zusammenfassung (Reduktion) von Daten Bestimmung typischer Kenngrößen	Auswertung einer erhobenen Stichprobe um Aufschluss über die Eigenschaften der größeren Grundgesamtheit zu erhalten

2.1 Grundbegriffe der beschreibenden Statistik

Datenbasis: eine statistische Untersuchung bezieht sich stets auf eine klar festgelegte Grundgesamtheit (GG) Ω (Population) die Menge aller möglichen (denkbaren) Beobachtungseinheiten. Die Elemente $\omega \in \Omega$ nennt man Merkmalsträger.

Beispiel: $\Omega = \{\omega : \omega \text{ ist Student an der TU Chemnitz im SS08}\}$

Gegenstand statistischer Erhebungen ist in der Regel nicht die GG Ω selbst, sondern Eigenschaften ihrer Elemente.

Definition:

Eine Abbildung $X : \Omega \rightarrow \mathbb{R}$, die jedem Merkmalsträger $\omega \in \Omega$ eine Zahl $X(\omega)$ zuordnet heißt Merkmal. Die Menge $X(\Omega)$ nennt man Menge der Merkmalsausprägungen.

Bemerkung:

Zufallsgrößen sind spezielle Merkmale (Messbarkeit). In der Literatur werden die Begriffe Merkmal, Zufallsgröße und Grundgesamtheit oft synonym verwendet.

Klassifizierung von Merkmalen:

Merkmal:

- diskretes Merkmal $X(\Omega) = \{x_1, x_2, \dots\}$
- stetiges Merkmal $X(\Omega) \subseteq \mathbb{R}$ (überabzählbar)

Merkmal:

- qualitativ: artmäßige Merkmale, z.B. Augenfarbe, Religionszugehörigkeit, Familienstand
- quantitativ: messbar (Erfassung durch Zahlen) z.B. Schuhgröße, Gewicht, Umsatz,... → Ausprägungen des Merkmals lassen sich in eine eindeutige Rangfolge bringen (ordnen)

Bemerkung:

Man kann qualitative Merkmale durch Zahlen kodieren. Trotzdem sind solche Merkmale nicht als quantitativ zu sehen, da sie in keine Reihenfolge gebracht werden können!

Skalierung von Merkmalen:

- **Norminalskala:** keine Ordnung der Ausprägungen (Bsp.: „männlich“, „weiblich“; Farben ...) ⇒ nur Untersuchungen auf Gleichheit sinnvoll
- **Ordinal- oder Rangskala:** Rangordnung der Merkmalsausprägung, Abstände zwischen Merkmalsausprägungen können nicht interpretiert werden (Bsp: Schadstoffklassen, Schulnoten)
- **metrische Skala:** Rangordnung und Abstände zwischen Merkmalsausprägungen sind messbar und interpretierbar.

Intervallskala: nur Differenzbildung möglich → nur Vergleich von Abständen (da keine natürliche Einheit, kein natürlicher Nullpunkt)

Verhältnisskala: Quotientenbildung zulässig und somit Verhältnis sinnvoll interpretierbar (keine natürliche Einheit, aber natürlicher Nullpunkt)

Absolutskala: zusätzlich natürliche Einheit (d.h. sie ergibt sich zwangsläufig) Bsp: Anzahl der Geschwister

Definition: (empirische Stichprobe)

Sei Ω eine Grundgesamtheit, X ein Merkmal und $\{\omega_1, \omega_2, \dots, \omega_n\} \subset \Omega$. Dann heißt (x_1, \dots, x_n) mit $x_i = X(\omega_i)$ $i = 1 \dots n$ (*empirische*) *Stichprobe* vom Umfang n (oder Messreihe oder Daten).

Eine Stichprobe heißt

1. zufällig, wenn jedem Element aus der Grundgesamtheit die gleiche Chance hat, in die Anwendung der Stichprobe bezüglich X zu gelangen
2. repräsentativ, wenn die der Stichprobe zugrundeliegenden ausgewählten Elemente alle Aspekte der Grundgesamtheit bezüglich des Merkmals X repräsentieren (z.B. Wahlprognose ganzes Parteienspektrum).

2.1 Grundbegriffe der beschreibenden Statistik

Bemerkung: insbesondere bilden die Realisierungen $x_1 \dots x_n$ von Zufallsgröße vom Typ i.i.d. eine zufällige und repräsentative Zufallsgröße vom Umfang n .

Ziel: Datenreduktion und Darstellung

Empirische (Häufigkeits-)Verteilung

- diskrete Merkmale (häufig nominale oder ordinale Merkmale): Merkmal X hat mögliche Ausprägungen $a_1 \dots a_n$

→ absolute Häufigkeiten $H_j = \sum_{k=1}^n 1_{\{x_k = a_j\}}$ $j = 1 \dots s$ (klar: $\sum_{j=1}^s H_j = n$,

$0 \leq H_j \leq n$)

→ relative Häufigkeiten: $h_j = \frac{1}{n} H_j$

- bei stetigen Merkmalen ist die Anzahl der beobachteten Merkmalsausprägungen sehr groß $\Rightarrow H_j \approx 1$, $h_j \approx \frac{1}{n}$ - deswegen Klasseneinteilung mit jeweiligen Repräsentanten (z.B. Intervallmitte)

$[a_1, a_2), [a_2, a_3), \dots, [a_s, a_{s+1})$

$$H_j = \sum_{k=1}^n 1_{\{x_k \in [a_j, a_{j+1})\}}, \quad h_j = \frac{1}{n} H_j$$

Regeln zur Klasseneinteilung:

1. $x_{\min} = \min_{k=1 \dots n} \{x_k\}$, $x_{\max} = \max_{k=1 \dots n} \{x_k\}$
bestimme Zahlen $a_1 < x_{\min} < a_2 < \dots < a_{s+1}$ ($> x_{\max}$)
2. Faustregel: $5 \leq s \leq 25$; $s \approx \sqrt{n}$ (im Zweifel ungerade)
3. i.A. Klassenbreiten $a_{j+1} - a_j$ konstant

Histogramm: Graph der empirischen Dichtefunktion wobei Histogrammfläche einer Klasse der relativen Häufigkeit h_j dieser Klasse entspricht

Höhe eines Balkens: $d_j = \frac{h_j}{a_{j+1} - a_j} \Rightarrow d_j(a_{j+1} - a_j) = h_j$

Lagemaße: bei der Berechnung von Lagemaßen von Stichproben muss sorgfältig auf die Art und Messbarkeit der entsprechenden Merkmale geachtet werden. Für nominale bzw. quantitative Merkmale macht die Berechnung von (arithmetischen) Mittelwerten keinen Sinn!

1. empirischer Modalwert

\bar{x}_{mod} = in Stichprobe am häufigsten vorkommender Wert (nicht eindeutig!)

→ Verwendung hauptsächlich für nominale Skalen, macht aber auch für andere Merkmalstypen Sinn. (auch wenn Aussagegehalt manchmal gering) (bei stetigen Merkmalen abhängig von Klasseneinteilung)

Im Folgenden betrachten wir nur noch ordinale oder metrische Merkmale und setzen geordnete Stichproben voraus.

$$x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq \dots \leq x_{(n)}$$

2. empirischer Median

$$\bar{x}_{0,5} = \begin{cases} x_{(\frac{n+1}{2})} & n \text{ ungerade} \\ \frac{1}{2} (x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}) & n \text{ gerade} \end{cases}$$

⇒ mindestens die Hälfte aller Stichproben-Elemente sind kleiner oder gleich $\bar{x}_{0,5}$ und mindestens die Hälfte aller Stichproben-Elemente sind größer oder gleich $\bar{x}_{0,5}$

Beispiele:

- Stichprobe: 3, 5, **7**, 9, 10 ⇒ $\bar{x}_{0,5} = 7$
- Stichprobe: 3, 5, **7, 9**, 10, 25 ⇒ $\bar{x}_{0,5} = \frac{1}{2}(7 + 9) = 8$

3. arithmetisches Mittel

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$$

⇒ \bar{x} sehr anfällig gegenüber sogenannten Datenausreißern. $\bar{x}_{0,5}$ ist hingegen robust.

Streuungsmaße → nicht für nominale Merkmale, in der Regel nur für metrische Merkmale

1. Spannweite: $x_{(1)} \leq \dots \leq x_{(n)}$ geordnete Stichprobe ⇒ $R = x_{(n)} - x_{(1)}$ = Differenz zwischen größtem und kleinstem Beobachtungswert

2. empirische Varianz (mittlere quadratische Abweichung)

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

3. mittlere absolute Abweichung (mittlere lineare Abweichung)

$$d = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

nicht so anfällig gegen Ausreißer, wird in der Praxis aber nicht so häufig verwendet.

4. Variationskoeffizient

$$v = \frac{s}{\bar{x}}$$

⇒ dient dem Vergleich der Streuung 2er verschiedener Stichproben

2.2 Grundbegriffe der beurteilenden Statistik

Wahrscheinlichkeitsrechnung: mathematische Modelle zufallsbedingter Vorgänge
(Ω, \mathfrak{a}, P), X, P

↓

mathematische Statistik: Auswertung von Ergebnissen, um Rückschlüsse auf X, F_X zu ziehen

↓

Stichprobe x_1, \dots, x_n

Gegenstand der beurteilenden Statistik

Wir wollen anhand von Beobachtungen der Realisierungen einer Zufallsgröße Rückschlüsse auf die Verteilung der Zufallsgröße ziehen.

- Schätzverfahren: Schluss auf Parameter
- Testverfahren: Beurteilung aufgestellter Hypothesen anhand empirisch gewonnener Daten

Definition:

Ein Zufallsvektor $\vec{X}_k = (X_1, \dots, X_n)$, dessen Komponenten unabhängige, identisch verteilte (i.i.d.) sind, heißt (mathematische) Stichprobe vom Umfang n . Eine Realisierung $\vec{x}_1 = (x_1, \dots, x_n)$ heißt konkrete Stichprobe vom Umfang n . Der Wertebereich von \vec{X}_n heißt Stichprobenraum und wird mit \mathfrak{X}_n bezeichnet.

Bemerkung:

- Die Annahme, X_1, \dots, X_n sind vom Typ i.i.d. spielt eine zentrale Rolle.
- Unabhängigkeit: Beobachtungen werden unabhängig durchgeführt
- identische Verteilung: alle potenzielle Merkmalsträger haben die gleiche Chance, in die Stichprobe aufgenommen zu werden

Beispiel: Qualitätskontrolle: Posten 100.000 Teilen, konkrete Stichprobe vom Umfang $n = 100$ ergibt: $K = 7$ fehlerhafte Teile

Frage 1: Ausschussanteil in der Ausgangsmenge?

mathematische Formulierung: $X = \begin{cases} 0 & \text{Teil in Ordnung} \\ 1 & \text{Teil defekt} \end{cases}$

SP $\vec{X}_{100} = (X_1, \dots, X_{100})$, konkrete Stichprobe $\vec{x} = (0, 0, 1, 0, \dots, 0, 0, 1, 0, 0)$

→ gesucht: Schätzung für $\theta \Rightarrow$ Parameterschätzung

Frage 2: Behauptung/Hypothese des Herstellers: Ausschussanteil $\leq 5\%$. Ist diese Behauptung zu halten? Es könnte ja Zufall sein, dass in der konkreten Stichprobe sich 7% ergeben haben \Rightarrow Testverfahren.

2.2.1 Parameterschätzung

Ziel: Aussagen über die unbekannte Verteilungsstruktur einer Zufallsgröße X aus der Beobachtung einer konkreten Stichprobe \vec{x} zu gewinnen. Wir bezeichnen mit $F_X(x, \theta) = P(X < x)$ die Verteilungsfunktion der Zufallsgröße X , wobei $\Theta = (\Theta_1, \dots, \Theta_m)$ die unbekanntes zu schätzenden Parameter bezeichnet.

Definition:

Die Menge aller Werte, die der Parameter θ annehmen kann, heißt Parameterraum (Parametermenge) und wird mit Θ bezeichnet.

Definition:

Eine Funktion $T_n^* \mathfrak{X}_n \rightarrow \theta$, die jeder Stichprobe \vec{X}_n einen Wert für den Parameter θ zuordnet, heißt Schätzfunktion für den Parameter θ (Zufallsgröße). Ist $\vec{x}_n = (x_1, \dots, x_n)$ eine konkrete Stichprobe, dann heißt $T_n^*(\vec{x}_n)$ Schätzwert für θ und wird mit t_n^* bezeichnet (Zahl).

Definition:

T_n^* heißt erwartungstreu für θ , wenn $ET_n^*(\vec{X}_n) = \theta$ gilt. Falls $ET_n^* = \theta + b(\theta)$ mit einer gewissen Funktion b , dann heißt $b(\theta)$ systematischer Fehler oder Bias.

Satz:

Sei X eine Grundgesamtheit mit Verteilungsfunktion $F_X(x, \theta)$ wobei $\theta = EX$. Dann ist die Schätzfunktion T_n^*

$$T_n^*(\vec{X}_n) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n$$

Dann ist die Schätzfunktion T_n^* erwartungstreu für θ .

Satz: (Schätzung der Varianz)

Sei X eine Grundgesamtheit mit Verteilungsfunktion $F(x, \theta)$ wobei $\theta = \sigma^2 = D^2X < \infty$. Ist $\mu = EX$ bekannt, so ist T_n^* mit

$$T_n^*(\vec{X}_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

eine erwartungstreue Schätzfunktion für die Varianz σ^2 . Wenn EX nicht bekannt ist, dann ist

$$T_n^*(\vec{X}_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

eine erwartungstreue Schätzfunktion für σ^2 .

Definition:

Sei X eine Grundgesamtheit mit Verteilungsfunktion $F(x, \theta)$. T_n^* und T_n^{**} seien zwei Schätzfunktionen für θ mit dem gleichen systematischen Fehler (bias) $b(\theta)$. Dann heißt T_n^* wirksamer als T_n^{**} , falls $D^2T_n^* < D^2T_n^{**}$ gilt.

Gütekriterien für große Stichprobenumfänge

Definition:

Eine Schätzfunktion für θ heißt asymptotisch erwartungstreu für θ , falls $\lim_{n \rightarrow \infty} ET_n^* = \theta$ gilt.

Bemerkung:

T_n^* erwartungstreu $\Rightarrow T_n^*$ asymptotisch erwartungstreu, der Umkehrschluss gilt jedoch nicht

Definition: Die Schätzfunktion T_n^* für θ heißt konsistent für θ , wenn $\forall \varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|T_n^* - \theta| > \varepsilon) = 0$$

gilt.

Bemerkung: $\hat{s}_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$, $s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$, $\tilde{s}_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$

konsistent

Satz:

T_n^* sei eine asymptotisch erwartungstreue Schätzfunktion für θ . Es gelte: $\lim_{n \rightarrow \infty} D^2 T_n^* = 0$. Dann ist T_n^* konsistent für θ .

Definition:

Sei X eine Grundgesamtheit mit Verteilungsfunktion $F(x, \theta)$ wobei $EX = \theta$ und $D^2 X < \infty$. Dann ist die Schätzfunktion T_n^* mit $T_n^*(\vec{X}_n) = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ konsistent für θ .

Zusammenfassung

Wenn X eine Grundgesamtheit und Verteilungsfunktion $F(x, \theta)$ mit $EX = \theta$ ist, dann ist $T_n^*(\vec{X}_n) = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ eine erwartungstreue und konsistente Schätzfunktion. Was kann man in anderen Situationen machen?

\Rightarrow verschiedene Methoden zur Konstruktion von Schätzfunktionen:

- Maximum-Likelihood-Methode
- Momentenmethode
- kleinste Quadrate Methode
- Bayesschätzungen

Maximum-Likelihood-Methode

X Grundgesamtheit mit Verteilungsfunktion $F(x, \theta)$, θ unbekannter Parameter, Stichprobe $\vec{X}_1 = (X_1, \dots, X_n)$, konkrete Stichprobe $\vec{x}_n = (x_1, \dots, x_n)$

Idee: wählen $\theta \in \Theta$ so, dass für diesen Parameter θ die konkrete Stichprobe \vec{x}_n am wahrscheinlichsten ist.

X diskrete Zufallsgröße: definieren die Likelihood-Funktion

$$L_x(\theta) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i) = \prod_{i=1}^n p_{x_i}(\theta)$$

lösen Optimierungsaufgabe $L_{\vec{x}_n}(\theta) \rightarrow \max_{\theta \in \Theta}$ und erhalten $\hat{t}_{ML} = \operatorname{argmax} L_{\vec{x}_n}(\theta)$ = Maximum-Likelihood Schätzung

X stetige Zufallsgröße: definieren Likelihood-Funktion

$$L_{\vec{x}_n}(\theta) = \prod_{k=1}^n f(x_k, \theta)$$

wobei $f(x, \theta)$ Dichtefunktion von X

Optimierungsaufgabe: $L_{\vec{x}_n}(\theta) \rightarrow \max_{\theta \in \Theta} \rightarrow$ erhalten $\hat{t}_{ML} = \operatorname{argmax} L_{\vec{x}_n}(\theta)$ bzw.

$$\ln L_{\vec{x}_n}(\theta) = \sum_{k=1}^n \ln f(x_k, \theta) \rightarrow \max_{\theta \in \Theta}$$

Bemerkung: (siehe Literatur)

- ML-Schätzer muss nicht immer existieren, außerdem hat $L_{\vec{x}_n}(\theta)$ gelegentlich mehrere Maxima
- alle ML-Schätzer sind konsistent
- existiert eine wirksamste Schätzfunktion so erhält man diese mit ML-Schätzung
- ML-Schätzer sind asymptotisch normalverteilt

Momentenmethode

X ist Grundgesamtheit, Verteilungsfunktion $F(x, \theta)$, θ unbekannter Parameter

$$\begin{array}{ll} m_k = EX^k & k\text{-te Momente} \\ M_K = E(X - EX)^j & k\text{-te zentrale Momente} \end{array}$$

→ Stichprobenmomente/empirische Momente

$$M_k = \frac{1}{n} \sum_{i=1}^n x_i^k \quad \text{bzw.} \quad \tilde{\mu}_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k$$

⇒ schätzen Parameter θ , indem wir geeignete Momente und die entsprechenden Stichprobenmomente gleichsetzen: $m_j = M_j \quad j \in J$

Bei Bedarf können auch die zentralen Momente verwendet werden: $\mu_j = \tilde{\mu}_j \quad j \in J$

Bemerkung: im Allgemeinen entstehen nichtlineare Gleichungssysteme

2.2.2 Intervallschätzungen (Bereichsschätzungen)

(bisher Punktschätzungen)

- X Grundgesamtheit mit Verteilungsfunktion $F(x, \theta)$
- Stichprobe $\vec{X}_n = (X_1, \dots, X_n)$ vom Umfang n
- Schätzfunktion $T_n^* : \mathfrak{X}_n \rightarrow \Theta$

Selbst wenn diese Schätzfunktion verschiedene Gütekriterien erfüllt, bleibt das Problem bestehen, dass der Parameter θ im Allgemeinen durch T_n^* nur näherungsweise bestimmt werden kann. Wenn T_n^* eine stetige Zufallsgröße ist, gilt sogar $P(T_n^*(\vec{X}_n) = \theta) = 0$. Deshalb ist oft ein Intervall von Interesse, in dem der unbekannte Parameter θ mit einer vorgegebenen Wahrscheinlichkeit liegt.

Definition:

Gegeben seien zwei Funktionen \underline{T}_n^* und \overline{T}_n^* , die jedem Element des Stichprobenraumes einen Wert aus dem Parameterraum zuordnen, .h. $\underline{T}_n^* : \mathfrak{X}_n \rightarrow \Theta$.

Sei $\alpha \in (0, 1)$ gegeben und es gelte $P(\underline{T}_n^*(\vec{X}_n) \leq \theta \leq \overline{T}_n^*(\vec{X}_n)) \geq 1 - \alpha$. Dann heißt das zufällige Intervall $[\underline{T}_n^*(\vec{X}_n), \overline{T}_n^*(\vec{X}_n)]$ Konfidenzintervall (Vertrauensintervall) für θ zum Konfidenzniveau $1 - \alpha$.

Ist $\vec{x}_n = (x_1, \dots, x_n)$ eine konkrete Stichprobe vom Umfang n , dann heißt $[\underline{T}_n^*(\vec{x}_n), \overline{T}_n^*(\vec{x}_n)]$ konkretes Konfidenzintervall.

Bemerkung:

Ein Konfidenzintervall (KI) überdeckt also den wahren Parameter θ mit einer Wahrscheinlichkeit von mind. $(1 - \alpha) \cdot 100\%$.

2.2 Grundbegriffe der beurteilenden Statistik

Vorgehen zur Konstruktion von KI.

- betrachten Schätzfunktion $T_n^*(\vec{X}_n)$ oder Punktschätzung für θ
- bestimme die Verteilung von T_n^* (oder falls es einfacher ist, von gewissen Funktionen von T_n^*)
- standardisiere diese Zufallsgröße, so dass die neue Zufallsgröße nicht mehr von θ abhängig ist („Entparametrisierung“)
- bestimme das KI

Beispiel: Konfidenzintervall für den Parameter $\mu = \theta$ einer Normalverteilung bei bekannter Varianz.

$X \sim N(\mu, \sigma^2)$, σ^2 bekannt, Stichprobe $\vec{X}_n = (X_1, \dots, X_n)$

Schätzfunktion für μ : $T_n^*(\vec{X}_n) = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

vgl. $X_i \sim N(\mu, \sigma^2)$ i.i.d.

$$\Rightarrow \sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2) \Rightarrow \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Standardisieren/Entparametrisieren: $Z = \frac{T_n^* - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\bar{X}_n - \mu}{\sigma} \cdot \sqrt{n} \sim N(0, 1)$

Verteilungsdichte von Z : $\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2}{2}\right)$

→ bestimme KI: wählen $Z_{1-\frac{\alpha}{2}}$ so, dass $P(-Z_{1-\frac{\alpha}{2}} \leq Z \leq Z_{1-\frac{\alpha}{2}})$

$$\begin{aligned} \text{haben } -Z_{1-\frac{\alpha}{2}} \leq Z \leq Z_{1-\frac{\alpha}{2}} &\Leftrightarrow \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n} \leq Z_{1-\frac{\alpha}{2}} \Leftrightarrow \frac{-Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \sigma \leq \bar{X}_n - \mu \leq \\ &\frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \sigma \Leftrightarrow -\bar{X}_n - \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \sigma \leq -\mu \leq -\bar{X}_n + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \sigma \Leftrightarrow \bar{X}_n + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \sigma \geq \mu \geq \bar{X}_n - \\ &\frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \sigma \Leftrightarrow \bar{X}_n - \frac{Z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \frac{Z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}} \end{aligned}$$

$$\Rightarrow P\left(\bar{X}_n - \frac{Z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \frac{Z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$\Rightarrow \left[\bar{X}_n - \frac{Z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}}, \bar{X}_n + \frac{Z_{1-\frac{\alpha}{2}} \sigma}{\sqrt{n}}\right]$$

Bemerkung:

- $Z_{1-\frac{\alpha}{2}}$ ist das Quantil der Ordnung $1 - \frac{\alpha}{2}$ der Standardnormalverteilung
- ohne Annahme, dass X einer normalverteilten Grundgesamtheit entstammt, könnten wir kein Konfidenzintervall berechnen
- falls σ^2 unbekannt ist, gilt nicht mehr, dass $Z \sim N(0, 1)$ verteilt ist
 \Rightarrow Verteilung typischer Stichprobenfunktionen

Verteilungen wichtiger Stichprobenfunktionen bei normalverteilter Grundgesamtheit

$X \sim N(\mu, \sigma^2)$, Stichprobe $\vec{X}_n = (X_1, \dots, X_n)$

wissen: $\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$ arithmetisches Mittel

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

χ^2 -Verteilung:

Eine stetige Zufallsgröße X mit der Dichtefunktion

$$f_n(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & x > 0 \end{cases}$$

heißt χ^2 -verteilt mit n Freiheitsgraden.

Schreibweise: $X \sim \chi_n^2$

Bemerkung:

- Gammafunktion $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ für $n \in \mathbb{N}$: $\Gamma(n+1) = n!$
- Verteilungsfunktion der χ^2 -Verteilung tabelliert in Abhängigkeit von n (Freiheitsgrade)
- Quantile der 2. Ordnung α : $\chi_{n,\alpha}^2$
- $X_1, \dots, X_n \sim N(0, 1)$ i.i.d. $\Rightarrow \sum_{i=1}^n X_i^2 \sim \chi_n^2$
- $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ i.i.d. $\Rightarrow \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi_n^2$
- ferner gilt: $\frac{n\hat{s}_n^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \sim \chi_{n-1}^2$

t-Verteilung (Student-Verteilung)

Definition:

Eine stetige Zufallsgröße X mit Dichtefunktion

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

heißt t -verteilt mit n Freiheitsgraden: Schreibweise: $X \sim t_n$

Bemerkung:

- Quantile $t_{n,\alpha}$ (Niveau α) sind in Abhängigkeit der Freiheitsgrade tabelliert
- $U \sim N(0, 1)$; $Q \sim \chi_n^2$ stochastisch unabhängig $\Rightarrow W = \frac{U}{\sqrt{Q}}\sqrt{n} \sim t_n$
- große Ähnlichkeit zur Normalverteilung, aber Wahrscheinlichkeitsmaße an den Flanken (tails) und dafür flacherer Verlauf in der Mitte
- t_n -Verteilung konvergiert für $n \rightarrow \infty$ gegen Standardnormalverteilung (Faustregel: $n \geq 40$)
- Stichprobe $\vec{X}_n = (X_1, \dots, X_n)$ wobei $X_i \sim N(\mu, \sigma^2) \Rightarrow U = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$,
 $Q = \frac{n \cdot \hat{s}_n^2}{\sigma^2} \sim \chi_{n-1}^2$ und U, Q stochastisch unabhängig

$$\Rightarrow \frac{U}{\sqrt{\frac{Q}{n-1}}} = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}} \sqrt{\frac{n \cdot \hat{s}_n^2}{\sigma^2(n-1)}}} = \frac{\bar{X}_n - \mu}{\hat{s}_n} \sqrt{n-1} \sim t_{n-1}$$

F-Verteilung

Definition:

Eine stetige Zufallsgröße X mit der Dichtefunktion

$$f_{m,n}(x) = \begin{cases} 0 & x \leq 0 \\ \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \cdot \left(\frac{m}{n}\right)^{\frac{m}{2}} \cdot x^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}} & x > 0 \end{cases}$$

heißt F -verteilt mit den Freiheitsgraden m und n . Schreibweise: $X \sim F_{m,n}$

Bemerkung:

- α -Quantile: $F_{m,n,\alpha}$ tabelliert

- es gilt: $F_{m,n,\alpha} = \frac{1}{F_{m,n,1-\alpha}}$
- $Q_n \sim \chi_n^2$, $Q_m \sim \chi_m^2$, Q_n , Q_m stochastisch unabhängig $\Rightarrow Z = \frac{Q_m}{Q_n} \sim F_{m,n}$
- Stichproben \vec{X}_n , X_m unabhängig

$$\Rightarrow \frac{1}{\sigma_x^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \sim \chi_{n-1}^2 \quad \frac{1}{\sigma_y^2} \sum_{i=1}^m (Y_i - \bar{Y}_m)^2 \sim \chi_{m-1}^2$$

$$\Rightarrow \frac{\frac{1}{(m-1)\sigma_y^2} \sum_{j=1}^m (Y_j - \bar{Y}_m)^2}{\frac{1}{(n-1)\sigma_x^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2} \sim F_{m-1,n-1}$$

Konfidenzintervalle bei normalverteilter Grundgesamtheit

$X \sim N(\mu, \sigma^2)$ Grundgesamtheit, Stichprobe $\vec{X}_n = (X_1, \dots, X_n)$

1. Fall: KI für μ , σ^2 bekannt (siehe Beispiel) $\rightarrow \left[\bar{X}_n - \frac{z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}, \bar{X}_n + \frac{z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}} \right]$
 ist ein KI zum Konfidenzniveau $1 - \alpha$ für den Parameter μ , d.h.
 $P\left(\bar{X}_n - \frac{z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \frac{z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}\right) = 1 - \alpha$

2. Fall: KI für μ , σ^2 unbekannt

analoges Vorgehen, aber ersetzen σ^2 durch s_n^2

$$\Rightarrow T = \frac{\bar{X}_n - \mu}{s_n} \sqrt{n} \text{ mit } \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

$\Rightarrow T \sim t_{n-1}$, d.h. T ist t -verteilt mit $n - 1$ Freiheitsgraden

$$\Rightarrow P(-t_{n-1,1-\frac{\alpha}{2}} \leq T \leq t_{n-1,1-\frac{\alpha}{2}}) = 1 - \alpha$$

$$\Rightarrow \text{KI} \left[\bar{X}_n - t_{n-1,1-\frac{\alpha}{2}} \frac{s_n}{\sqrt{n}}, \bar{X}_n + t_{n-1,1-\frac{\alpha}{2}} \frac{s_n}{\sqrt{n}} \right] \text{ zum Konfidenzniveau } 1 - \alpha$$

3. Fall: KI für σ^2 bei bekanntem μ

Betrachten: $T = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 = \frac{\hat{s}_n^2 n}{\sigma^2} \sim \chi_n^2$ und bestimmen Quantile $\chi_{n,\frac{\alpha}{2}}^2$

und $\chi_{n,1-\frac{\alpha}{2}}^2$, so dass $P\left(\chi_{n,\frac{\alpha}{2}}^2 \leq T \leq \chi_{n,1-\frac{\alpha}{2}}^2\right) = 1 - \alpha$

$$\begin{aligned} &\Rightarrow P\left(\chi_{n, \frac{\alpha}{2}} \leq \frac{\hat{s}_n^2}{\sigma^2} \leq \chi_{n, 1-\frac{\alpha}{2}}\right) = 1 - \alpha \Rightarrow \\ &P\left(\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{n, 1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{n, \frac{\alpha}{2}}}\right) = 1 - \alpha \\ &\Rightarrow \left[\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{n, 1-\frac{\alpha}{2}}^2}, \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{n, \frac{\alpha}{2}}^2} \right] \text{ KI für } \sigma^2 \text{ (} \mu \text{ bekannt) zum Niveau } 1 - \alpha \end{aligned}$$

4. Fall: KI für σ^2 bei unbekanntem μ

analoges Vorgehen, ersetzen aber μ durch \bar{X}_n

$$\rightarrow \text{ betrachten } T = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n \cdot \hat{s}_n^2}{\sigma^2} = \frac{(n-1)s_n^2}{\sigma^2} \sim \chi_{n-1}^2$$

\rightarrow bestimmen wieder Quantile $\chi_{n-1, \frac{\alpha}{2}}^2$ und $\chi_{n-1, 1-\frac{\alpha}{2}}^2$ mit $P(\chi_{n-1, \frac{\alpha}{2}}^2 \leq T \leq \chi_{n-1, 1-\frac{\alpha}{2}}^2) = 1 - \alpha$

$$\Rightarrow \left[\frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2}, \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{\chi_{n-1, \frac{\alpha}{2}}^2} \right] \text{ KI für Parameter } \sigma^2 \text{ (} \mu \text{ unbekannt) zum Niveau } 1 - \alpha$$

Bemerkung:

Wir haben jetzt für alle 4 Fälle (μ/σ^2 bekannt/unbekannt) der normalverteilten Grundgesamtheit Konfidenzintervalle hergeleitet. Wesentlich ist dabei die Annahme, dass $X \sim N(\mu, \sigma^2)$. Falls diese Voraussetzung nicht erfüllt ist, können mithilfe der Grenzwertsätze zumindest für große n auch für andere Verteilungsannahmen Konfidenzintervalle hergeleitet werden.

Asymptotisches Konfidenzintervall für binomialverteilte GG

Seien $X \sim B(1, p)$ (Zweipunkt-verteilt) $X = \begin{cases} 0 & \text{mit Wahrscheinlichkeit } 1 - p \\ 1 & \text{mit Wahrscheinlichkeit } p \end{cases}$

dann ist $\sum_{i=1}^n X_i \sim B(n, p) \approx N(np, np(1-p))$ für hinreichend große n (vgl. GWS von Moivre/Laplace)

$$\Rightarrow X_n = \frac{1}{n} \sum_{i=1}^n X_i \approx N\left(p, \frac{p(1-p)}{n}\right)$$

2.2.3 Hauptsatz der mathematischen Statistik

X Grundgesamtheit mit Verteilungsfunktion $F(x)$. Bis jetzt war $F(x)$ stets in der Form $F(x, \theta)$ gegeben, d.h. die prinzipielle Verteilungsform war bekannt, „nur“ der Parameter θ war zu bestimmen. Was ist zu tun, wenn keine Informationen über die Struktur von F vorhanden sind?

geg.: $\vec{X}_n = (X_1, \dots, X_n)$ Stichprobe vom Umfang n

ges.: Schätzung für $F(x)$

Definition: (empirische Verteilungsfunktion)

Die Funktion $F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i < x\}}$, $x \in \mathbb{R}$ heißt empirische Verteilungsfunktion von X .

Eigenschaften von $F_n(x)$:

- $F_n(x)$ ist eine zufällige Funktion
- für eine konkrete Stichprobe $\vec{x}_n = (x_1, \dots, x_n)$ ist $F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i < x\}}$, $x \in \mathbb{R}$ eine (deterministische) Treppenfunktion und wird als konkrete empirische Verteilungsfunktion bezeichnet.
- für x fixiert ist $F_n(x)$ die relative Häufigkeit des Ereignisses $A = \{\omega : X(\omega) < x\}$ in der (math.) Stichprobe \vec{X}_n
 \Rightarrow damit ist $F_n(x)$ eine erwartungstreue und konsistente Schätzfunktion für $\theta = P(A) = P(X < x) = F(x)$ (Voraussetzung: x fixiert)

Satz 2.5: (Hauptsatz der mathematischen Stochastik)

Sei X eine Grundgesamtheit mit Verteilungsfunktion $F(x)$. Sei $\vec{X}_n = (X_1, \dots, X_n)$ eine Stichprobe und $F_n(x)$ die empirische Verteilungsfunktion. Weiterhin sein $\Delta_n = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$.

Dann gilt $P(\lim_{n \rightarrow \infty} \Delta_n = 0) = 1$.

Bemerkung:

Satz 2.5 besagt, dass eine hinreichend große Stichprobe beliebig genaue Informationen über die Verteilung der Grundgesamtheit liefert. Das ist erstaunlich, da abzählbar viele Beobachtungen zu Aussagen über $F(x)$ auf ganz \mathbb{R} (überabzählbar) liefern. Die Frage, wie groß die Stichprobe sein muss, um eine gewisse Genauigkeit der Verteilungsfunktion zu erhalten, wird im Satz 2.5 aber nicht geklärt.

2.2.4 Prüfen statistischer Hypothesen (Tests)

Ziel eines Tests ist es, eine Hypothese (Annahme) über die tatsächliche Verteilung der Grundgesamtheit auf Grund der Daten einer Stichprobe zu verwerfen oder als möglich anzusehen. Bei sogenannten *Parametertests* werden Aussagen über den unbekannt Parameter $\theta \in \Theta$ überprüft. Wenn Aussagen über die Art der Verteilung der Grundgesamtheit getroffen werden, spricht man von *parameterfreien Tests*.

→ Signifikanztests (gibt es signifikante Abweichungen von der Hypothese)

Grundidee:

Untersuchen unter der Voraussetzung, dass die Herstellerbehauptung (Hypothese) wahr ist, die Wahrscheinlichkeit, in einer Stichprobe vom Umfang $n = 100$ mindestens $k = 10$ fehlerhafte Teile zu finden. 2 Möglichkeiten:

- diese Wahrscheinlichkeit ist „groß“: falls die Herstellerbehauptung richtig ist, haben wir ein „völlig normales“ Ergebnis beobachtet und können nichts gegen die Herstellerbehauptung sagen
- diese Wahrscheinlichkeit ist „klein“: falls die Herstellerbehauptung richtig ist, haben wir ein sehr seltenes Ereignis beobachtet. Da seltene Ereignisse naturgemäß selten eintreten, führen wir diese Beobachtung nicht auf ein solches seltenes Ereignis zurück, sondern darauf, dass die Herstellerangaben falsch sind. Dabei besteht das Risiko, die Herstellerbehauptung abzulehnen, obwohl sie eigentlich richtig sind.

konkret: $X \sim B(1, p)$, $X_i = \begin{cases} 1 & i\text{-tes Teil defekt} \\ 0 & \text{sonst} \end{cases}$

$$\Rightarrow Z = \sum_{i=1}^{100} X_i \sim B(n, p) \quad n = 100$$

Herstellerbehauptung: $p = 0,05 \Rightarrow$ falls Hypothese richtig ist gilt $Z \sim B(100, 0.05)$

Die Wahrscheinlichkeit, dass in der Stichprobe vom Umfang $n = 100$ mindestens $k = 10$ defekte Zeile sind, beträgt:

$$P(Z \geq 10) = 1 - P(Z < 10) = 1 - \sum_{k=0}^9 \binom{100}{k} 0.05^k \cdot 0.95^{100-k} = 0.0282$$

→ diese Wahrscheinlichkeit ist „klein“ und wir gehen davon aus, dass die Herstellerangaben falsch sind. Die Wahrscheinlichkeit, dem Hersteller Unrecht zu tun (also die Hypothese abzulehnen, obwohl sie richtig ist) beträgt nur 2,82%

Allgemeines Vorgehen:

geg.:

- GG X mit Verteilungsfunktion $F(x)$
 - Stichprobe $\vec{X}_n = (X_1, \dots, X_n)$ vom Umfang n
 - konkrete Stichprobe $\vec{x}_1 = (x_1, \dots, x_n)$
1. Aufstellen einer Hypothese H_0 (Nullhypothese) $H_0 : F(x) = F_0(x)$
 2. Konstruktion einer Testgröße $T: T = T(X_1, \dots, X_n)$
 Die Testgröße soll Unterschiede zwischen hypothetischer Verteilung F_0 und der tatsächlichen Verteilung F widerspiegeln und deren Verteilung bekannt ist, falls H_0 richtig ist.
 3. Konstruktion eines kritischen Bereiches K^* derart, dass zu geg. $\alpha \in (0, 1)$ gilt:
 $P(T \in K^* | H_0 \text{ richtig}) = \alpha$
 4. Für die konkrete Stichprobe $x_n = (x_1, \dots, x_n)$ ist der Testwert $t = T(x_1, \dots, x_n)$ zu bilden und
 - $t \in K^* \Rightarrow$ Ablehnen von H_0 (Test signifikant)
 - $t \notin K^* \Rightarrow$ auf der Basis des durchgeführten Tests ist nichts gegen die aufgestellte Hypothese einzuwenden (Test ist nicht signifikant)

Bemerkungen:

- zu 1)
 bei Parametertests wird ein gewisser Verteilungstyp vorausgesetzt und H_0 beschränkt sich auf die Wahl der Parameter $\theta = (\theta_1, \dots, \theta_s) \in \Theta$. Falls nur ein unbekannter Parameter θ kommen beispielsweise die Hypothesen
 - $H_0 : \theta = \theta_0, H_1 : \theta \neq \theta_0$
 - $H_0 : \theta \leq \theta_0, H_1 : \theta > \theta_0$
 - $H_0 : \theta \geq \theta_0, H_1 : \theta < \theta_0$
 in Frage.
- Die hypothetische Verteilung kann eine aus gewissen Erfahrungen vermutete Verteilung sein.
- zu 2)
 Testgrößen zu konstruieren ist im Allgemeinen schwierig (Verteilung muss bekannt sein, falls H_0 richtig ist)
- es existieren „vorgefertigte“ Testgrößen für verschiedene Aufgabenstellungen
- zu 3)
 da die Verteilung von T bekannt ist (falls H_0 richtig ist), stellt die Konstruktion des kritischen Bereichs K^* kein Problem dar

2.2 Grundbegriffe der beurteilenden Statistik

- α heißt Irrtumswahrscheinlichkeit oder Signifikanzniveau und ist „klein“ zu wählen \rightarrow typische Werte: $\alpha = 0.05$, $\alpha = 0.01$, $\alpha = 0.005$
- zu 4)
Die Wahrscheinlichkeit, dass T im kritischen Bereich liegt, obwohl H_0 richtig ist, beträgt α .
- eine entsprechende Fehlentscheidung heißt Fehler 1. Art. Bei Wiederholungen des Tests trifft eine solche Fehlentscheidung in $\alpha \cdot 100\%$ aller Fälle auf.
- aus dem Nichtablehnen von H_0 ($t \notin K^*$) folgt im Allgemeinen nicht, dass H_0 richtig ist (keine signifikante Aussage) \rightarrow Fehler 2. Art: H_0 ist falsch, wird aber nicht abgelehnt ($t \notin K^*$). Die Wahrscheinlichkeit für das Auftreten dieses Fehler ist im Allgemeinen nicht bekannt.

Parametertests bei normalverteilter Grundgesamtheit

GG $X \sim N(\mu, \sigma^2)$, Stichprobe $\vec{X}_n = (X_1, \dots, X_n)$, konkrete Stichprobe $\vec{x}_n = (x_1, \dots, x_n)$

Fall 1: Hypothesen über μ bei bekanntem σ^2

1. Nullhypothese $H_0: \mu = \mu_0$

Alternativhypothese $H_1: \mu \neq \mu_0$

2. Testgröße: $T = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$ falls H_0 richtig ist wobei $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_u \sim$

$N\left(\mu_0, \frac{\sigma^2}{n}\right)$ falls H_0 richtig ist

\rightarrow Abweichungen zwischen hypothetischer und tatsächlicher Verteilung sind erkennbar

3. kritischer Bereich: haben $H_0: \mu = \mu_0$, d.h. Abweichungen nach unten und nach oben sind kritisch \rightarrow zweiseitige Fragestellung
4. Entscheidungsregel
 - $t \in K^* \rightarrow$ Ablehnen von H_0 (signifikant)
 - $t \notin K^* \rightarrow$ Test hat gegen H_0 keinen Einwand

Beispiel:

X = Abfüllgewicht von Schokoladentafeln in g

konkrete Stichprobe $\vec{x}_{10} = (100, 97, 101, 96, 98, 102, 96, 100, 101, 98) \rightarrow \bar{x}_{10} = 98.9$,
 $s_{10} = 2.183$

Annahme: $X \sim N(\mu, \sigma^2)$, σ bekannt $\rightarrow \sigma = 2$, $\alpha = 0.05$

$H_0: \mu \geq 100 = \mu_0, H_1: \mu < 100 = \mu_0$

$$\text{Testgröße } Z = \frac{\bar{x}_{10} - \mu_0}{\sigma} \sqrt{n} = \frac{98.9 - 100}{2} \sqrt{10} = -1.739$$

kritischer Bereich: $z_\alpha = -z_{1-\alpha} \rightarrow z_{0.05} = -z_{0.95} = -1.6449$

$\rightarrow K^* = (-\infty, -1.6449) \Rightarrow z \in K^* \Rightarrow H_0$ ablehnen und stattdessen H_1 annehmen.

\rightarrow Das Abfüllgewicht der Schokoladentafeln ist signifikant kleiner als 100g bei einer Irrtumswahrscheinlichkeit von 5%.

Bemerkung: falls $\alpha = .01 \rightarrow z_{1-\alpha} = z_{0.99} = 2.326 \rightarrow K^* = (-\infty, -2.326) \rightarrow z \notin K^*$, d.h. der Test hat bei einer Irrtumswahrscheinlichkeit von 1% nichts gegen die Hypothese $H_0: \mu \geq 100$ anzuwenden.

Merke: bei einseitigen Fragestellungen immer die kritische Fragestellung in die Alternativhypothese, damit eine signifikante Aussage entsteht, falls H_0 abgelehnt wird und stattdessen H_1 angenommen wird.

Fall 2: Hypothesen über μ bei unbekanntem σ^2

$X \sim N(\mu, \sigma^2, \sigma^2 \text{ unbekannt})$

1. Hypothesen:

- $H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$
- $H_0: \mu \leq \mu_0, H_1: \mu > \mu_0$
- $H_0: \mu \geq \mu_0, H_1: \mu < \mu_0$

2. Testgröße:

$$T = \frac{\bar{X}_n - \mu_0}{s_1} \sqrt{n} \text{ mit } \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

falls H_0 richtig ist, gilt $T \sim t_{n-1}$, d.h. t -verteilt mit $n-1$ Freiheitsgraden

3. kritischer Bereich k^*

$$H_0: \mu = \mu_0 \Rightarrow k^* = \{t : |t| > t_{n-1, 1-\frac{\alpha}{2}}\} = (-\infty, -t_{n-1, 1-\frac{\alpha}{2}}) \cup (t_{n-1, 1-\frac{\alpha}{2}}, \infty)$$

$$H_0: \mu \leq \mu_0 \Rightarrow k^* = \{t : t > t_{n-1, 1-\alpha}\} = (t_{n-1, 1-\alpha}, \infty)$$

$$H_0: \mu \geq \mu_0 \Rightarrow k^* = \{t : t < t_{n-1, 1-\alpha}\} = (-\infty, -t_{n-1, 1-\alpha})$$