Induktive Statistik mit möglichst wenig Formeln

Friedrich-Alexander Universität
Erlangen-Nürnberg

WS 2012/13

http://www.statistik.wiso.uni-erlangen.de/lehre/sonstige-veranstaltungen/induktive-statistik-fast-ohne-formeln.shtml

Gliederung

- Grundlagen der Wahrscheinlichkeitsrechnung
- Zufallsvariablen und ihre Verteilungen
- Parametrische Verteilungsfamilien
- Stichproben
- Parameterschätzung
- Intervallschätzung
- Hypothesentests
- Lineares Modell

Beispielhafte Fragestellungen

- Beim Glücksspiel haben Sie das Gefühl, dass die "Gegenseite" mit manipulierten Würfeln spielt, d.h. die Zahl 6 fällt häufiger als jede andere Zahl. Wie können Sie Ihre Vermutung überprüfen?
- Kurz vor der Landtagswahl wird von einem Wahlforschungsinstitut prognostiziert, dass 42% der abgegebenen Stimmen auf die CSU entfallen.

  Wie kann die damit verbundene Unsicherheit minimiert werden?
- Die Stadtverwaltung möchte Kosten sparen und überlegt die U-Bahnen ab 21.30 Uhr nur noch im 20-Minuten Takt fahren zu lassen. Im Stadtrat einigt man sich darauf, dass dies sinnvoll ist, wenn im Durchschnitt pro Station nur noch 5 Personen einsteigen. Wie ist dies zu überprüfen?
- Bei der Erforschung möglicher Nebenwirkungen eines neuen Medikaments möchte ein Pharmaunternehmen herausfinden, in welchem Umfang die Einnahme zu einer erhöhten Reaktionszeit führt.

Ziel: Schlussfolgerungen von Stichprobe auf Grundgesamtheit (immer nur unter Unsicherheit möglich!)

Grundlagen der Wahrscheinlichkeitsrechnung

- Grundbegriffe: Ereignisse etc.
- Wahrscheinlichkeit
- Bemessung der Wahrscheinlichkeit
- Rechnen mit Wahrscheinlichkeiten
- Bedingte Wahrscheinlichkeit
- Stochastische Unabhängigkeit

Grundbegriffe

- Zufallsexperiment:
  Experimente, die tatsächlich oder nur gedanklich durchgeführt werden und deren Ausgang ungewiss ist.

  Beispiel Werfen zweier unterscheidbarer Würfel.

- Ergebnismenge Ω:
  Menge aller möglichen, sich gegenseitig ausschließenden Ausgänge, d. h. aller **Ergebnisse** eines Zufallsexperiments.

  Beispiel:

$$\boldsymbol{\Omega} = \left\{ \begin{array}{cccccc} (1;1) & (1;2) & (1;3) & (1;4) & (1;5) & (1;6) \\ (2;1) & (2;2) & (2;3) & (2;4) & (2;5) & (2;6) \\ (3;1) & (3;2) & (3;3) & (3;4) & (3;5) & (3;6) \\ (4;1) & (4;2) & (4;3) & (4;4) & (4;5) & (4;6) \\ (5;1) & (5;2) & (5;3) & (5;4) & (5;5) & (5;6) \\ (6;1) & (6;2) & (6;3) & (6;4) & (6;5) & (6;6) \end{array} \right\}$$

Grundbegriffe

- Elementarereignisse ω: Elemente der Ergebnismenge.

  Beispiel: $(i;j)$ für $i = 1, \ldots, 6$ und $j = 1, \ldots, 6$.

- Ereignisse:
  Teilmengen der Ergebnismenge - Bezeichnung: $(A, B, C, \ldots)$.

A	„Das Produkt der Augenzahlen beträgt 6“
B	„Der erste Würfel zeigt eine 1 oder eine 2“
C	„Die Summe der Augenzahlen ist größer als 7“
E	„Das Produkt der Augenzahlen ist negativ“
F	„Die Augenzahl des zweiten Würfels ist kleiner als 7“.

- Eintreten eines Ereignisses: Ereignis A tritt ein, wenn ein $\omega \in A$ realisiert wird.
- Besondere Ereignisse

  Unmögliches Ereignis: $\emptyset$ und Sicheres Ereignis: Ω.

Wahrscheinlichkeit

- Grundbegriff Wahrscheinlichkeit ist intuitiv klar, aber intuitive Vorstellung liefert nicht immer mathematisch verwertbare Ergebnisse.
- Wahrscheinlichkeit ordnet jedem möglichen Ereignis eine reelle Zahl zu $\Rightarrow P(A)$ (Rechenregeln später)
- Bestimmen von Wahrscheinlichkeiten
    - Laplace-Wahrscheinlichkeit
    - Empirische Wahrscheinlichkeit
    - Subjektive Wahrscheinlichkeit

Laplace–Wahrscheinlichkeit (1812)

- Gleichwahrscheinlichkeitsmodell: Jedes der n Elemetarereignisse ω_i, $i = 1, \ldots, n$, ist gleichwahrscheinlich, d.h.

$$P(\{\omega_i\}) = 1/n \;\; i = 1, 2, \ldots, n.$$

- Folgerung: Wenn A ein Ereignis mit k (günstigen) Elementarereignissen ist, dann ist

$$P(A) = \frac{\text{Anzahl der günstigen Elem.ereign.}}{\text{Anzahl aller möglichen Elem.ereign.}} = \frac{k}{n}.$$

- Beispiele: Würfelwurf, Ziehen aus einer Urne oder Kartenspiel.
- Gegenbeispiel: Ausschuss bei Fertigung von Schrauben
- Rechtfertigung für die Annahme der Gleichwahrscheinlichkeit: Prinzip des mangelnden Grundes nach Laplace.

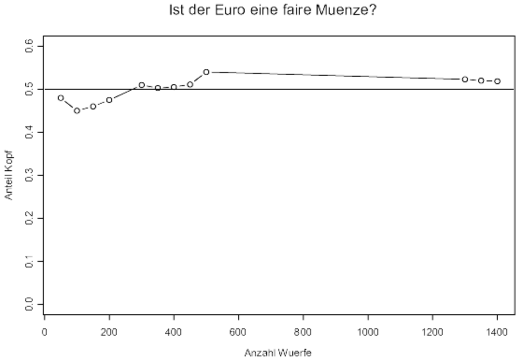
Empirische Wahrscheinlichkeit

- Ausgangspunkt: n-malige unabhängige Wiederholung eines Zufallsexperiments.
- $P(\{\omega\})$ wird durch die relative Häufigkeit "geschätzt", mit der das Elementarereignis ω bei der n-maligen Wiederholung auftritt:

$$P(\{\omega\}) \approx \frac{\text{Häufigkeit, mit der } \omega \text{ auftritt}}{n}.$$

- Beispiel 1: Ausschuss bei Fertigung von Schrauben mit den Ergebnissen K: „Schraube korrekt“ und N: „Schraube fehlerhaft“.

- Beispiel 2: Werfen des EURO zur empirischen Bestimmung der Wahrscheinlichkeit $P(K)$ mit K: "Kopf"!



- Rechtfertigung: Gesetze der großen Zahlen, d.h. die rel. Häufigkeit konvergiert für $n \to \infty$ gegen $P(A)$.

Rechnen mit Wahrscheinlichkeiten

- Beispiel: Werfen zweier Würfel
  - Ergebnismenge: $\Omega = \{(i;j)|i,j = 1,2,\ldots,6\}$.
  - Jedes Elementarereignis ist gleichwahrscheinlich also gilt:

$$P(\ \{(i;j)\}\ ) = 1/36 \text{ für } (i;j) \in \Omega.$$

- Axiome von Kolmogorov (1933): Die Wahrscheinlichkeit ordnet jedem möglichen Ereignis eine reelle Zahl zu, so dass gilt:
  - K1: Jedem Ereignis A wird eine nichtnegative Wahrscheinlichkeit $P(A)$ zugeordnet.
  - K2: Das sichere Ereignis bekommt die Wahrscheinlichkeit 1 (=100%) zugewiesen.
  - K3: Die Wahrscheinlichkeit, dass eines von zwei sich gegenseitig ausschließender Ereignisse eintritt, entspricht der Summe der Wahrscheinlichkeiten der einzelnen Ereignisse.
    Beispiel: Die Wahrscheinlichkeit, dass die Sonne scheint oder es regnet, ist gleich der Wahrscheinlichkeit, dass die Sonne scheint plus die Wahrscheinlichkeit, dass es regnet.

Ausgangspunkt: Empirische WS bzw. relative Häufigkeit

Folgerungen aus den Kolmogorovschen Axiomen

- Gegenwahrscheinlichkeit: Die Wahrscheinlichkeit, dass das Ereignis A nicht eintritt, ist 1 minus der Wahrscheinlichkeit, dass das Ereignis A eintritt.

  Beispiel: Wenn es mit einer Wahrscheinlichkeit von 0.2 regnet, regnet es mit einer Wahrscheinlichkeit von $1 - 0.2 = 0.8$ nicht.

- Monotonie: Zieht das Ereignis A das Ereignis B nach sich (d.h. $A \subseteq B$), dann kann A nicht mit größerer Wahrscheinlichkeit als B eintreten.

  Beispiel: Ereignis, dass die erste Augenzahl eine 1 ist (=Ereignis A_1) zieht das Ereignis nach sich, dass die Augensumme kleiner oder gleich 7 ist (=Ereignis B):

$$P(A_1) = \frac{6}{36} \leq \frac{21}{36} = P(B).$$

Folgerungen aus den Kolmogorovschen Axiomen

- Zerlegungssatz: Seien $A_1, A_2, \ldots, A_k$ sich gegenseitig ausschließende Ereignisse ($A_i \cap A_j = \emptyset$ mit $i \neq j$), so dass mindestens ein Ereignis A_i mit Sicherheit eintritt ($\Omega = A_1 \cup A_2 \cup \ldots \cup A_k$), dann gilt für jedes beliebige Ereignis B:

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \ldots + P(A_k \cap B)$$

Beispiel: Wenn A_i das Ereignis ist, dass der erste Würfel die Augenzahl i zeigt für $i = 1, 2, \ldots, 6$, ist die Wahrscheinlichkeit, dass die Augensumme gerade (=Ereignis B) ist, gleich der Wahrscheinlichkeit, dass der erste Würfel eine 1 zeigt und die Augensumme gerade ist plus der Wahrscheinlichkeit, dass der erste Würfel eine 2 zeigt und die Augensumme gerade ist etc.

$$P(B) = \frac{3}{36} + \frac{3}{36} + \ldots + \frac{3}{36} = \frac{18}{36}.$$

Folgerungen aus den Kolmogorovschen Axiomen

- Allgemeiner Additionssatz:

  Die Wahrscheinlichkeit, dass mindestens eines von zwei sich nicht gegenseitig ausschließenden Ereignissen eintritt, ist gleich der Summe der Wahrscheinlichkeiten der einzelnen Ereignisse minus der Wahrscheinlichkeit, dass beide Ereignisse gemeinsam eintreten.

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

- Beispiel: Wahrscheinlichkeit, dass der erste Würfel eine 1 zeigt (=Ereignis A_1) oder die Summe der Augenzahlen gerade (=Ereignis B) ist:

$$P(A_1 \cup B) = P(A_1) + P(B) - P(A_1 \cap B) = \frac{6}{36} + \frac{18}{36} - \frac{3}{36} = \frac{21}{36}.$$

Bedingte Wahrscheinlichkeit

- Angenommen man weiß, dass die Augensumme gerade ist $(= B)$. Wie groß ist dann die Wahrscheinlichkeit, dass die erste Augenzahl eine 1 ist $(= A_1)$?

  Antwort: Es sind nur noch 18 Ergebnisse möglich, von denen 3 für den ersten Würfel eine 1 aufweisen, d.h. $3/18$ oder

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{3/36}{18/36} = \frac{3}{18}.$$

- "$P(A_1|B)$" heißt bedingte Wahrscheinlichkeit, dass das Ereignis A_1 eintritt, unter der Bedingung, dass das Ereignis B bereits eingetreten ist.
- Was besagt $P(B|A_1)$?

Stochastische Unabhängigkeit

- Die Wahrscheinlichkeit, dass die Augensumme gerade ist ($= B$) unter der Bedingung, dass der erste Würfel eine 1 ($= A_1$) trägt ist

$$P(B|A_1) = \frac{P(A_1 \cap B)}{P(A_1)} = \frac{3/36}{6/36} = \frac{1}{2} = P(B).$$

- D.h. die Kenntnis, dass das Ereignis A_1 eingetreten ist, hat keinen Einfluss auf die Wahrscheinlichkeit, dass das Ereignis B eintritt.
- In diesem Falle heißen A_1 und B stochastisch unabhängig.
- Beachte: Auch A_2 und B, A_3 und B etc. sind stochastisch unabhängig.
- Sind A_1 und A_2 stochastisch unabhängig?

Stochastische Unabhängigkeit

- Wenn A und B stochastisch unabhängig sind, gilt

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

- Multipliziert man mit $P(B)$ und teilt man durch $P(A)$, so ergibt sich

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

  D.h. bei stochastischer Unabhängigkeit von A und B gilt auch

$$P(B|A) = P(B).$$

- Betrachte wieder die Ereignisse A_1 (erster Würfel trägt "1")und B (gerade Augensumme). Dann sind

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{3/36}{18/36} = \frac{6}{36} = P(A_1)$$

Stochastische Unabhängigkeit

- Schließlich ist bei stochastischer Unabhängigkeit von A und B auch

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

  weil $P(A|B) = P(A)$ bei stochastischer Unabhängigkeit gilt.

- D.h.: Bei stochastischer Unabhängigkeit ist die Wahrscheinlichkeit, dass zwei Ereignisse gemeinsam auftreten gleich dem Produkt der Wahrscheinlichkeiten, dass die Ereignisse einzeln auftreten.
- Betrachte wieder die Ereignisse A_1 und B. Dann ist

$$P(A_1 \cap B) = \frac{3}{36} = \frac{6}{36} \cdot \frac{18}{36} = P(A_1)P(B).$$

Vollständige stochastische Unabhängigkeit

- Bislang wurde nur die Unabhängigkeit von zwei Ereignissen betrachtet. Bei mehr als zwei Ereignissen verkompliziert sich die Bedingung für die stochastische Unabhängigkeit.
- Drei Ereignisse A, B, C heißen vollständig stochastisch unabhängig, wenn
  - sie paarweise stochastisch unabhängig sind (d.h. $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, $P(B \cap C) = P(B)P(C)$) und
  - $P(A \cap B \cap C) = P(A)P(B)P(C)$

  gelten.
- D.h.: Aus der paarweisen stochastischen Unabhängigkeit folgt nicht notwendig die vollständige stochastische Unabhängigkeit.

Vollständige stochastische Unabhängigkeit

- Beispiel: In einer Urne sind 4 Kugeln, von denen 3 jeweils nur die Zahl 1, 2 oder 3 tragen. Die vierte Kugel trägt alle drei Zahlen.

  Das Ereignis A_i tritt ein, wenn eine Kugel gezogen wird, die die Zahl i trägt für $i = 1, 2, 3$.

  Paarweise Unabhängigkeit, weil z.B.

  $$P(A_1 \cap A_2) = 0.25 = 0.5 \cdot 0.5 = P(A_1)P(A_2).$$

  Aber: Keine vollständige stochastische Unabhängigkeit

  $$P(A_1 \cap A_2 \cap A_3) = 0.25 \neq 0.5 \cdot 0.5 \cdot 0.5 = P(A_1)P(A_2)P(A_3).$$

Rechenregeln für bedingte Wahrscheinlichkeiten

- Beispiel: Angenommen
  - die Wahrscheinlichkeit, dass es regnet ($=A$), wenn man einen Regenschirm mitnimmt ($= B_1$) ist 0.2 ($= P(A|B_1)$),
  - die Wahrscheinlichkeit, dass es regnet, wenn man keinen Regenschirm mitnimmt ($= B_2 = \overline{B}_1$) ist 0.4 ($= P(A|B_2)$)
  - und die Wahrscheinlichkeit, dass ein Regenschirm mitgenommen wird, ist 0.4 ($=P(B_1)$).
- Wie groß ist die Wahrscheinlichkeit, dass es regnet und man einen Regenschirm dabei hat?

$$P(A \cap B_1) = P(B_1) \cdot P(A|B_1) = 0.2 \cdot 0.4 = 0.08.$$

(Multiplikationssatz)

Rechenregeln für bedingte Wahrscheinlichkeiten

- Wie groß ist die Wahrscheinlichkeit, dass es regnet?

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) = 0.2 \cdot 0.4 + 0.4 \cdot 0.6 = 0.32.$$

(Formel der totalen Wahrscheinlichkeit)

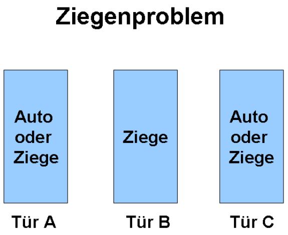
- Wie groß ist die Wahrscheinlichkeit, dass man einen Regenschirm dabei hat, wenn es regnet?

$$P(B_1|A) = P(A|B_1)P(B_1)/P(A) = 0.2 \cdot 0.4/0.32 = 0.25.$$

(Bayes-Formel)

Ziegenproblem

- Ein Teilnehmer einer Fernsehshow wird vor drei gleich aussehende Türen gestellt.
- Hinter zwei der Türen befinden sich Ziegen, die als Nieten fungieren sollen.
- Hinter der dritten steht der Hauptgewinn, ein Auto.



- Der Kandidat soll nun eine Tür auswählen (z.B. die erste Tür).
- Dann öffnet der Showmaster eine der anderen Türen, hinter der kein Auto steht (z.B. die zweite Tür).
- Frage: Lohnt es sich für den Kandidaten zur dritten Tür zu wechseln?

Ziegenproblem

- Lösungsvorschläge:
  1. Da eine Tür (ohne Auto) geöffnet wurde, bleiben zwei Türen übrig, hinter denen das Auto mit Wahrscheinlichkeit $1/2$ steht, so dass sich ein Wechsel nicht lohnt.
  2. Die Wahrscheinlichkeit, dass das Auto hinter der ersten Tür steht, ist zunächst $1/3$, so dass das Auto mit einer Wahrscheinlichkeit von $2/3$ hinter der zweiten oder dritten Tür steht. Wenn die zweite Türe wegen Öffnung wegfällt, muss das Auto mit Wahrscheinlichkeit $2/3$ hinter der dritten Tür stehen.
- Kommentare:
  1. Der erste Vorschlag ist falsch, da sich die Wahrscheinlichkeit, dass das Auto hinter der ersten Tür steht, nicht durch das Öffnen der zweiten Tür erhöht.
  2. Die Wahrscheinlichkeit, dass das Auto hinter der dritten Tür steht, wird jedoch durch das Öffnen der zweiten Tür erhöht, da eine gleichwahrscheinliche Alternative ausscheidet.

Ziegenproblem

- Formale Behandlung: Ereignisse:
  - A_i: Auto steht hinter der Tür i, $i = 1, 2, 3$ $(P(A_i) = 1/3)$.
  - T_2: Showmaster öffnet Tür 2.

Gesucht: $P(A_3|T_2)$.

Lösung:

- Bekannt:

$$P(T_2|A_1) = 1/2, \quad P(T_2|A_2) = 0, \quad P(T_2|A_3) = 1.$$

- Totale Wahrscheinlichkeit:

$$\begin{aligned} P(T_2) &= P(T_2|A_1)P(A_1) + P(T_2|A_2)P(A_2) \\ &\quad +P(T_3|A_3)P(A_3) = 1/2 \cdot 1/3 + 1 \cdot 1/3 = 1/2. \end{aligned}$$

- Bayes-Formel:

$$\begin{aligned} P(A_3|T_2) &= P(T_2|A_3)P(A_3)/P(T_2) \\ &= 1 \cdot (1/3)/(1/2) = 2/3 > P(A_1) = 1/3. \end{aligned}$$

Anwendung: Zweistufiges Zufallsexperiment

- Ausgangssituation: Drei Urnen U_1, U_2 und U_3.
  - Die erste Urne enthält 2 schwarze und 1 weiße Kugel.
  - Die zweite Urne enthält 1 schwarze und 2 weiße Kugeln.
  - Die dritte Urne enthält 3 schwarze und 3 weiße Kugeln.
- Erste Stufe des Zufallsexperiments: Wähle eine der drei Urnen zufällig aus per Werfen eines Würfels:
  - Bei einer Augenzahl kleiner als 4 wird die erste Urne gewählt.
  - Bei einer 4 wird die zweite Urne gewählt.
  - Bei einer Augenzahl größer als 4 wird die dritte Urne gewählt.
- Zweite Stufe des Zufallsexperiments: Aus der gewählten Urne wird zufällig eine Kugel entnommen.

  Wenn die i-te Urne gewählt wird, tritt das Ereignis U_i ein, für $i = 1, 2, 3$.

  Wenn eine weiße Kugel gezogen wird, tritt das Ereignis W ein.

  Wenn eine schwarze Kugel gezogen wird, tritt das Ereignis S ein.
- Fragen: Wie groß ist die Wahrscheinlichkeit,
  1. dass eine weiße Kugel aus der zweiten Urne stammt?
  2. eine weiße Kugel zu ziehen?

Verteilung einer Zufallsvariablen

- Zufallsvariable
- Wahrscheinlichkeits- und Dichtefunktion
- Verteilungsfunktion
- Erwartungswerte

Übergang von Ereignis zu Zufallsvariable

- In der Wahrscheinlichkeitsrechnung wird berechnet, wie groß die Wahrscheinlichkeit eines Ereignisses (z.B. eine 6 zu würfeln) ist, was zumindest bei der Laplace-Wahrscheinlichkeit über das Auszählen von Elementarereignissen geschieht.
- Für die Statistik sind aber Ereignisse nur im Vergleich interessant.

  D.h. die Wahrscheinlichkeit, dass die Augensumme beim zweimaligen Würfelwurf 2 ist, soll mit den Wahrscheinlichkeiten verglichen werden, dass die Augensumme die anderen möglichen Werte $3, 4, \ldots, 12$ annimmt.
- Die Augensumme wird deshalb als eine Funktion aufgefasst, die der Ergebnismenge, bestehend aus den Paaren $(1, 1), (1, 2), \ldots, (6, 6)$, die möglichen alternativen Werte ($=$Realisationen) der Augensumme zuordnet.

  Ihre Funktionswerte variieren für die Elementarereignisse, so dass man bei der Augensumme von einer Zufallsvariablen spricht.
- Kennt man die möglichen Realisationen einer Zufallsvariablen, dann ist es möglich, die gesamte Wahrscheinlichkeitsmasse 1 der Ergebnismenge auf Realisationen zu verteilen.
- Solche Verteilungen aufzustellen ist eine wesentliche Aufgabe der Statistik, die auch als Lehre von den Verteilungen bezeichnet wird.

Zufallsvariable

- Eine Zufallsvariable X bildet die Elementarereignisse eines Zufallsexperimentes so in die reellen Zahlen ab, dass dort für sämtliche interessierenden Ereignisse die (induzierten) Wahrscheinlichkeiten berechnet werden können.
- Was sind z.B. interessierende Ereignisse für reelle Zahlen?
  - $X = 5$,
  - $-2 < X \leq 5$,
  - $X > 4$
  - $X \leq 2$ und $X \geq 7$.
- BEISPIEL: "Münzwurf"

  Ergebnismenge: $\Omega = \{\text{Kopf,Zahl}\}$.

  Zufallsvariable:

$$X : \begin{cases} \text{Kopf} & \mapsto 1 \\ \text{Zahl} & \mapsto 0. \end{cases}$$

  (Induzierte) Wahrscheinlichkeiten für spezielle interessierende Ereignisse:
  - $P_X(X = 1) = P(\{\text{Kopf}\} = 1/2$,
  - $P_X(X \leq -1) = P(\emptyset) = 0$,
  - $P_X(X \leq 2) = P(\{\text{Kopf,Zahl}\}) = P(\Omega) = 1$.

Zufallsvariable: Augensumme und Augenprodukt

- Ergebnismenge

$$\Omega = \{(i,j) | i = 1,2,\ldots,6 \text{ und } j = 1,2,\ldots,6\}.$$

- Zufallsvariablen: Augensumme X und Augenprodukt Y

$$X((i,j)) = i + j \text{ und } Y((i,j)) = i \cdot j.$$

- (Induzierte) Wahrscheinlichkeit für das Ereignis "Augensumme kleiner oder gleich 8":

$$\begin{aligned} P_X(X \leq 8) &= P(\{(i,j)\} | X((i,j)) \leq 8\}) \\ &= P(\{(1,1),\ldots,(1,6),(2,1),\ldots,(2,6),\ldots,(6,2)\}) = 26/36. \end{aligned}$$

  (Induzierte) Wahrscheinlichkeit für das Ereignis "Augensumme kleiner oder gleich x":

$$P_X(X \leq -1) = P(\emptyset) = 0 \text{ und } P(X \leq 400) = P(\Omega) = 1.$$

Arten von Zufallsvariablen

- Die Zufallsvariablen Augensumme kann nur endlich viele diskrete Werte als Realisationen (nämlich $2, 3, \ldots, 12$) mit einer positiven Wahrscheinlichkeit annehmen.

  Die Wahrscheinlichkeit für alle anderen Werte ist immer 0.
- Man spricht in einem solchen Falle von einer diskreten Zufallsvariablen.
- Eine diskrete Zufallsvariable liegt aber auch vor, wenn man zählt, wie häufig beim Münzwurf ,,Kopf" oben liegt, wenn man den Münzwurf unendlich oft wiederholt. Die möglichen (ebenfalls diskreten) Realisationen sind $0, 1, 2, \ldots$.

  Da man diese unendliche Menge abzählen kann, spricht man von einer abzählbar-unendlichen Menge.
- Keine diskrete Zufallsvariable liegt vor, wenn man die Bogenlänge betrachtet, die beim Drehen eines Glücksrades vom Nullpunkt aus zurückgelegt wird. Wenn r der Radius des Glücksrades ist, kann die Zufallsvariable Bogenlänge jede reelle Zahl zwischen 0 und $2\pi r$ (=Kreisumfang) als Realisation annehmen.

  Beachte: Diese Menge ist so groß, dass sie nicht mehr abgezählt werden kann. Sie heißt überabzählbar.
- Ist die Menge der Realisationen einer Zufallsvariablen überabzählbar, spricht man von einer stetigen Zufallsvariablen.

Arten von Zufallsvariablen

- Die Zahl $\pi = 3.14\ldots$ besitzt unendlich viele Nachkommastellen.
- Will man also eine konkrete Messung für die Bogenlänge angeben, wird diese Messung nur genau bis auf eine bestimmte Anzahl von Nachkommastellen sein.
- Somit sind die meisten stetigen Zufallsvariablen bei konkreter Messung der Realisationen eigentlich auch diskret. Man bezeichnet sie auch als quasi-stetig.
- Zur Vereinfachung wird aber angenommen, dass man die Realisationen ,,unendlich genau" messen kann, so dass die Menge der Realisationen überabzählbar ist.

  In diesem Falle ist aber Wahrscheinlichkeit einer einzelnen Realisation immer gleich 0.

Beispiele diskreter Zufallsvariablen

Experiment	ZV	mögliche Werte
1 mal Würfeln	Augenzahl	1, 2, . . . , 6
Qualitätskontrolle	Anz. defekt unter n	0, 1, 2, . . . , n
Kundenumfrage	Anz. zufrieden unter n	0, 1, 2, . . . , n
Bankschalter	Anz. Kunden pro Std.	0, 1, 2, . . .
20 Antworten in Test	Anz. richtige	0, 1, 2,. . . , 20
Passagiere in U-Bahn	Anz. Passagiere	0, 1, 2, . . .
100 Verkaufsgespräche	Anz. Verkäufe	0, 1, 2, . . . , 100

Kennzeichen sind, dass

- meist nur ganze Zahlen als Ergebnisse möglich sind,
- sie meist ein Ergebnis von Zählexperimenten sind und
- sie meist nur endliche Anzahl von Werten besitzen.

Beispiele stetiger Zufallsvariablen

- Alle Größen, die eine Dimensionseinheit besitzen und die man prinzipiell beliebig genau messen kann.
- Umsätze, Kosten, Gewinne von Unternehmen
- Körpergrößen, Körpergewicht von Personen
- Entfernungen
- Kennzeichen: Im Gegensatz zur Zufallsvariablen Bogenlänge, die durch das Drehen des Glücksrades als Zufallsexperiment entstanden ist, wird für Umsätze, Größen, Entfernungen etc. zumeist unterstellt, dass es sich um Zufallsvariablen handelt, obwohl im Allgemeinen kein Zufallsexperiment angegeben werden kann.

Wahrscheinlichkeitsfunktion

- Ausgangspunkt: Diskrete Zufallsvariable mit den Realisationen $x_1, x_2, \ldots, x_n$.
- Wahrscheinlichkeitsfunktion ist die Liste aller Paare $(x_i, f(x_i))$ mit

  x_i ... Realisation der Zufallsvariablen X

  $f(x_i)$ ... Wahrscheinlichkeit, dass Realisation x eintritt $f(x_i) = P(X = x_i)$
- Eigenschaften (Axiome von Kolmogorov) der Wahrscheinlichkeitsfunktion
  - $0 \leq f_X(x_i) \leq 1$
  - $\sum_{i=1}^{n} f_X(x_i) = 1$
- Beispiel 1: Z als Ergebnis beim "Münzwurf"

$$f_Z(z) = \begin{cases} 1/2 & \text{für } z \in \{0, 1\} \\ 0 & \text{sonst} \end{cases}$$

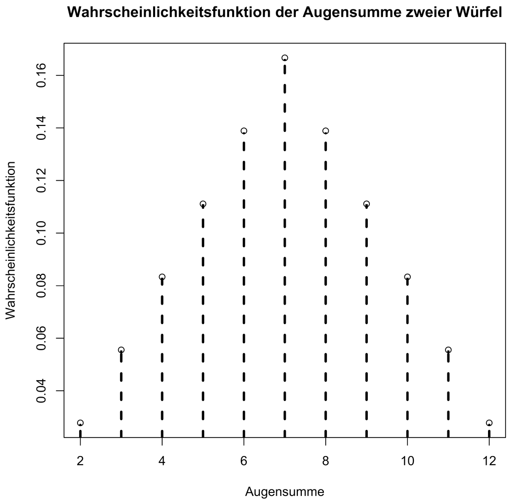
- Beispiel 2: Y als Anzahl der "Wappen" beim zweimaligen Münzwurf

Wahrscheinlichkeitsfunktion

- Beispiel 3: X als "Augensumme" beim zweimaligen Würfeln

$$f(x) = \begin{cases} 1/36 & \text{für } x = 2 \\ 2/36 & \text{für } x = 3 \\ 3/36 & \text{für } x = 4 \\ 4/36 & \text{für } x = 5 \\ 5/36 & \text{für } x = 6 \\ 6/36 & \text{für } x = 7 \\ 5/36 & \text{für } x = 8 \\ 4/36 & \text{für } x = 9 \\ 3/36 & \text{für } x = 10 \\ 2/36 & \text{für } x = 11 \\ 1/36 & \text{für } x = 12 \\ 0 & \text{sonst} \end{cases}$$

Grafische Darstellung der Wahrscheinlichkeitsfunktion



Verteilungs- und Wahrscheinlichkeitsfunktion

- Verteilungsfunktion $F_X(x)$ einer ZV X gibt an, mit welcher Wahrscheinlichkeit die ZV X den Wert x nicht überschreitet
- Für $D(X) = \{x_1, x_2, x_3, \ldots | x_1 \leq x_2 \leq x_3 \leq \ldots\}$ gilt:

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

d.h. Summation über alle x_i, die kleiner gleich x sind.

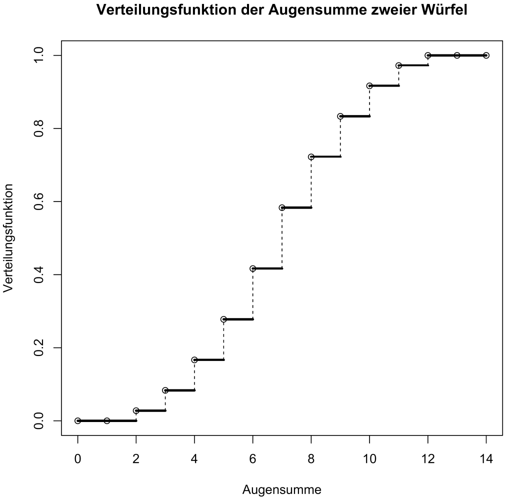
- Eigenschaften
  - F monoton steigend in $\mathbb{R}$
  - $P(X > a) = 1 - P(X \leq a) = 1 - F(a)$
  - $F(b) - F(a) = P(a < X \leq b)$ für alle $a, b \in \mathbb{R}$ mit $a \leq b$

Verteilungsfunktion

- Beispiel: X als "Augensumme" beim zweimaligen Würfeln

$$F(x) = P(X \leq x) = \begin{cases} 0 & x < 2 \\ 1/36 & 2 \leq x < 3 \\ 3/36 & 3 \leq x < 4 \\ 6/36 & 4 \leq x < 5 \\ 10/36 & 5 \leq x < 6 \\ 15/36 & 6 \leq x < 7 \\ 21/36 & 7 \leq x < 8 \\ 26/36 & 8 \leq x < 9 \\ 30/36 & 9 \leq x < 10 \\ 33/36 & 10 \leq x < 11 \\ 35/36 & 11 \leq x < 12 \\ 1 & 12 \leq x \end{cases}$$

Grafische Darstellung der Verteilungsfunktion



Beispiel: Zufallsvariable Y Augenprodukt

Realisation y	Elementar-ereignisse	Wahrschein-funktion $f_Y(y)$	Verteilungs-funktion $F_Y(y)$
1	$(1,1)$	$1/36$	$1/36$
2	$(1,2)$, $(2,1)$	$2/36$	$3/36$
3	$(1,3)$, $(3,1)$	$2/36$	$5/36$
4	$(2,2)$, $(1,4)$, $(4,1)$	$3/36$	$8/36$
5	$(1,5)$, $(5,1)$	$2/36$	$10/36$
6	$(2,3)$, $(3,2)$, $(1,6)$, $(6,1)$	$4/36$	$14/36$
8	$(2,4)$, $(4,2)$	$2/36$	$16/36$
9	$(3,3)$	$1/36$	$17/36$
10	$(2,5)$, $(5,2)$	$2/36$	$19/36$
12	$(2,6)$, $(6,2)$, $(3,4)$, $(4,3)$	$4/36$	$23/36$
15	$(3,5)$, $(5,3)$	$2/36$	$25/36$
16	$(4,4)$	$1/36$	$26/36$
18	$(3,6)$, $(6,3)$	$2/36$	$28/36$
20	$(4,5)$, $(5,4)$	$2/36$	$30/36$
24	$(4,6)$, $(6,4)$	$2/36$	$32/36$
25	$(5,5)$	$1/36$	$33/36$
30	$(5,6)$, $(6,5)$	$2/36$	$35/36$
36	$(6,6)$	$1/36$	$36/36$

Beispiel: Zufallsvariable Y Augenprodukt

- $P(Y = 12) = F_Y(12) - F_Y(10) = 14/36 - 10/36 = 4/36 = f_Y(12)$
- $P(Y \leq 18) = F_Y(18) = 28/36$
- $P(Y < 18) = P(Y \leq 16) = F_Y(16) = 26/36$
- $P(5 < Y \leq 18) = P(Y \leq 18) - P(Y \leq 5) = F_Y(18) - F_Y(5) = 28/36 - 10/36 = 18/36$
- $P(5 \leq Y \leq 18) = P(Y \leq 18) - P(Y < 5) = F_Y(18) - F_Y(4) = 28/36 - 8/36 = 20/36$
- $P(5 < Y < 18) = P(Y \leq 16) - P(Y \leq 4) = F_Y(16) - F_Y(4) = 26/36 - 8/36 = 18/36$
- $P(5 \leq Y < 18) = F_Y(16) - F_Y(5) = 26/36 - 10/36 = 16/36$
- $P(Y > 12) = 1 - P(Y \leq 12) = 1 - F_Y(12) = 1 - 23/36 = 13/36$
- $P(Y \geq 12) = 1 - P(Y < 12) = 1 - P(Y \leq 10) = 1 - 19/36 = 17/36$

Erwartungswert und Varianz

- Maßzahlen einer Wahrscheinlichkeitsverteilung
- Erwartungswert von X (vgl. arithm. Mittel)
  - Gewichtetes Mittel aller möglichen Realisationen $x_1, x_2, \ldots, x_n$.
  - Mittel der Wahrscheinlichkeitsverteilung

$$\mu = E(X) = \sum_{i=1}^{n} x_i \cdot f(x_i)$$

  - Wert von X der bei einer großen Zahl von Wiederholungen des Experiments im Durchschnitt zu erwarten ist.
- Varianz = Mittlere quadratische Abweichung vom Erwartungswert

$$\sigma^2 = Var(X) = E[(X - \mu)^2] = \sum_{i=1}^{n} (x_i - \mu)^2 \cdot f(x_i)$$

U.U. einfachere Berechnung mittels Varianzverschiebung:

$$\sigma^2 = \sum_{i=1}^{n} x_i^2 \cdot f(x_i) - \mu^2$$

Erwartungswert und Varianz - Beispiel

Zufallsvariable X: Augensumme beim zweimaligen Würfeln

$$\begin{aligned} E[X] &= 2f(2) + 3f(3) + \ldots + 12f(12) \\ &= 2 \cdot 1/36 + 3 \cdot 3/36 + \ldots + 12 \cdot 1/36 = 7. \end{aligned}$$

$$VAR[X] = (2-7)^2 f(2) + (3-7)^2 f(3) + \ldots (12-7)^2 f(12) = 5.83.$$

oder

$$VAR[X] = 2^2 \cdot f(2) + 3^2 \cdot f(3) + \ldots + 12^2 \cdot f(12) - 7^2 = 5.83$$

Erwartungswert und Varianz - Beispiel

Zufallsvariable Y: Augenprodukt beim zweimaligen Würfeln

$$\begin{aligned} E[Y] &= 1f(1) + 2f(2) + \ldots + 36f(36) \\ &= 1 \cdot 1/36 + 2 \cdot 2/36 + \ldots + 36 \cdot 1/36 = 12.25 \end{aligned}$$

$$VAR[Y] = 1^2 f(1) + 2^2 \cdot f(2) + \ldots + 36^2 \cdot f(36) - 12.25^2 = 79.965$$

Erwarteter Gewinn und Bernoulli-Entscheidungsprinzip

- Zufallsexperiment: Werfen eines Würfels
  Zufallsvariable: Augenzahl Y beim einfachen Münzwurf.
  Erwartete Augenzahl: $\mu_Y = E[Y] = 3.5$.
- Ihnen wird folgendes Glücksspiel angeboten:
  Nachdem Sie einen Einsatz von EUR 10,– geleistet haben, wird der Würfel geworfen und Sie erhalten das vierfache der Augenzahl als EURO-Betrag ausgezahlt.
  Würden Sie an diesem Spiel teilnehmen?
- Bernoulli-Entscheidungsprinzip: Das Glücksspiel lohnt sich, wenn der erwartete Gewinn nicht kleiner als der Einsatz ist.

$$\text{Erwarteter Gewinn} = 3.5 \cdot 4 = 14 > 10.$$

Rechenregeln für Erwartungswerte

- Erwartungswert einer linearen Transformation von X:

$$E[a + bX] = a + bE[X] \qquad \text{für} \quad a, b \in \mathbb{R}.$$

- Varianz einer linearen Transformation von X:

$$VAR[a + bX] = b^2 VAR[X] \qquad \text{für} \quad a, b \in \mathbb{R}.$$

Zentrierung und Standardisierung

- Standardisieren von Zufallsvariablen um Vergleichbarkeit zu schaffen.
- $Y = X - \mu_X$ heißt zentrierte Zufallsvariable. Mittels Zentrierung gelangt man immer zu einer Zufallsvariablen mit Mittelwert 0.

$$\mu_Y = E[Y] = E[X - \mu_X] = E[X] - \mu_X = 0$$

- $Y = (X - \mu_X)/\sigma_X$ heißt standardisierte Zufallsvariable. Mittels Standardisierug gelangt man immer zu einer Zufallsvariablen mit Mittelwert 0 und Varianz 1.

$$\begin{aligned}\mu_Y &= E[Y] = E\left[\frac{X - \mu_X}{\sigma_X}\right] = \frac{1}{\sigma_X} E[X - \mu_X] = 0 \\ \sigma_Y^2 &= Var[Y] = Var\left[\frac{X - \mu_X}{\sigma_X}\right] = \frac{1}{\sigma_X^2} Var[X] = 1\end{aligned}$$

Beispiel: Zentrierung und Standardisierung

- Zentrierte Zufallsvariable Augensumme: $X - 7$
- Standardisierte Zufallsvariable Augensumme: $(X - 7)/\sqrt{5.83}$
- Zentrierte Zufallsvariable Augenprodukt: $X - 12.25$
- Standardisierte Zufallsvariable Augensumme: $(X - 12.25)/\sqrt{79.965}$

Maßzahlen für zweidimensionale Zufallsvariablen

- Gegeben: Zufallsvariable X mit $E[X] = \mu_X$, $Var[X] = \sigma_X^2$ und Zufallsvariable Y mit $E[Y] = \mu_Y$ und $Var[Y] = \sigma_Y^2$
- Maß für die Stärke des linearen Zusammenhangs: Kovarianz

$$Cov[X,Y] = E\{[X - E[X]] \cdot [Y - E[Y]]\} = E[X \cdot Y] - E[X] \cdot E[Y]$$

- Korrelationskoeffizient:

$$-1 \leq \rho = \frac{Cov[X,Y]}{\sigma_X \cdot \sigma_Y} \leq 1$$

- Die Linearkombination $Z = a \cdot X + b \cdot Y$ ist ebenfalls wieder eine Zufallsvariable mit

$$E(Z) = aE(X) + bE(Y)$$

$$Var(Z) = a^2 Var(X) + b^2 Var(Y) + 2ab\,Cov(X,Y)$$

Anwendung: Korrelation von Augensumme und Augenprodukt

- X: Augensumme, Y: Augenprodukt
- Bekannt: $E[X] = 7$, $Var[X] = 5.83$, $E[Y] = 12.25$, $Var[Y] = 79.965$.
- Gesucht: $E[XY]$?

$$\begin{aligned} E[XY] &= 2 \cdot 1P(X = 2, Y = 1) + 3 \cdot 1P(X = 3, Y = 1) + \ldots \\ &\quad +12 \cdot 32P(X = 6, Y = 6) \\ &= 2 \cdot 1\frac{1}{36} + 3 \cdot 1 \cdot 0 + \ldots + 12 \cdot 32\frac{1}{36} = 106.17 \end{aligned}$$

- Kovarianz:

$$Cov[X, Y] = E[XY] - E[X]E[Y] = 106.17 - 7 \cdot 12.25 = 20.41.$$

- Korrelationskoefizient:

$$\frac{Cov[X, Y]}{\sqrt{Var[X]Var[Y]}} = \frac{20.41}{\sqrt{5.83 \cdot 79.965}} = 0.947.$$

- Fazit: Ein fast perfekter positiver linearer Zusammenhang zwischen Augensumme und Augenprodukt.

Anwendung: Anlageentscheidung

- Ein Kapitalanleger stehe vor der Wahl, einen festen Betrag X in Aktien anzulegen.
- Er habe zwei Aktien zur Auswahl.
- Die Renditen dieser Aktien, R_1 und R_2, gemessen in Prozent, sind nicht perfekt vorhersehbar und damit Zufallsvariablen.
- Der Investor hat jedoch feste Vorstellungen vom durchschnittlichen Ertrag der beiden Aktien (gemessen durch die jeweiligen Erwartungswerte) und deren Risiko (gemessen durch die Varianzen):

$$\mu_1 = 5\,,\ \mu_2 = 20\,,\ \sigma_1^2 = 4\,,\ \sigma_2^2 = 49$$

- Für welche Aktienanlage würden Sie sich entscheiden?

Anlageentscheidung: Durchschnittliche Rendite und Risiko eines Portfolios

- Ein Portfolio besteht aus einem Anteil z, $0 \leq z \leq 1$ den der Anleger in die erste Aktie investiert. Für die zweite Aktie verbleibt ein Anteil $1 - z$.
- Rendite des Portfolios:

$$R_p = zR_1 + (1-z)R_2$$

- Durchschnittliche Rendite des Portfolios:

$$E[R_p] = zE[R_1] + (1-z)E[R_2] = E[R_2] + z(E[R_1] - E[R_2])$$

- Risiko des Portfolios:

$$\begin{aligned} Var[R_p] &= z^2 Var[R_1] + (1-z)^2 Var[R_2] \\ &\quad +2z(1-z)Cov[R_1, R_2] \\ &= z^2 Var[R_1] + (1-z)^2 Var[R_2] \\ &\quad +2z(1-z)\rho_{R_1,R_2}\sqrt{Var[R_1]Var[R_2]} \end{aligned}$$

Anlageentscheidung: Diversifikationseffekt

- Wenn $z = 0.5$ und $\rho_{R_1,R_2} = \rho_{1,2} = -0.8$, dann sind die durchschnittliche Rendite des Portfolios

$$E[R_p] = 0.5 \cdot 5 + 0.5 \cdot 20 = 12.5$$

und das Risiko des Portfolios

$$VAR[R_p] = 0.25 \cdot 4 + 0.25 \cdot 49 + 2 \cdot 0.25 \cdot (-0.8) \cdot \sqrt{4} \cdot \sqrt{49} = 2.95$$

D.h. das Risiko des Portfolios ist geringer als das der beiden Aktien (Diversifikationseffekt).

- Sind die beiden Einzelrisiken gleich (d.h. $VAR[R_1] = VAR[R_2] = \sigma^2$) und herrscht eine perfekte negative Korrelation (d.h. $\rho_{1,2} = -1$), dann ist das Risiko des Portfolios

$$VAR[R_p] = (z^2 + (1-z)^2 - 2z(1-z))\sigma^2$$

minimal ($= 0$) für $z = 0.5$. In diesem Falle lässt sich das Risiko durch eine Diversifikation völlig ausschalten.

Einführungsbeispiel

- In einer Urne befinden sich 10 Kugeln, davon 2 schwarze. Es wird eine Stichprobe vom Umfang 4 mit Zurücklegen gezogen. Die ZV X bezeichnet die Anzahl der schwarzen Kugeln in der Stichprobe.
- Wie groß ist die WS, dass bei den **vier** gezogenen Kugeln
  - genau 1 schwarze Kugel enthalten ist?
  - genau 2 schwarze Kugeln enthalten sind?
  - keine schwarze Kugel enthalten ist?
  - höchstens 3 schwarze Kugeln enthalten sind?
  - mindestens 2 und höchstens 3 schwarze Kugeln enthalten sind?
- Berechnen Sie den Erwartungswert für X.

Was ist eine parametrische Verteilungsfamilie?

- Viele Fragestellungen lassen sich mithilfe weniger Annahmen auf die gleiche Art von Zufallsexperiment zurückführen.
- Einfache Beispiele: Würfel- und Münzwurf oder Lostrommel mit Gewinnen und Nieten
- Idee: Wahrscheinlichkeitsfunktion, die durch einige wenige Parameter identifiziert ist.
- Eine parametrische Verteilungsfamilie ist allgemein

$$\{f_X(x;\theta)|\theta = (\theta_1, \theta_2, \ldots, \theta_k) \in \Theta\}$$

- Vorteil:
  - Sobald Parameter der Verteilung bekannt sind, können Wahrscheinlichkeiten, Erwartungswert usw. einfach und schnell berechnet werden (WS-Tabellen).
  - Beim Übertragen von Ergebnissen der Stichprobe auf die Grundgesamtheit muss nicht komplette WS-funktion geschätzt werden, sondern nur wenige Parameter.

Diskrete Verteilungsmodelle

- Diskrete Gleichverteilung
- Bernoulli- oder Zweipunktverteilung
- Binomialverteilung
- Hypergeometrische Verteilung
- Geometrische Verteilung
- Poisson-Verteilung
- ...

Binomialverteilung

- Ausgangspunkt: Bernoulli-Experiment
  - Experiment bei dem nur zwei unterschiedliche Ergebnisse möglich sind.
  - Beispiele: Geburt mit "Mädchen" oder "Junge", Münzwurf mit "Kopf" oder "Zahl"
  - Allgemein: ZV Z mit den Ergebnissen "Erfolg" und "Misserfolg"
  - Bezeichnung: Erfolg mit $Z = 1$ und Misserfolg mit $Z = 0$
  - Wahrscheinlichkeit für Erfolg bzw. Misserfolg:

$$P(Z = 1) = p \qquad P(Z = 0) = 1 - p$$

  - Zweipunktverteilung
- Zufallsexperiment: Es wird ein Bernoulli-Experiment mit Erfolgswahrscheinlichkeit p n-mal unabhängig wiederholt.
- Sei die Zufallsvariable X die Anzahl der Erfolge.
- Notation: $X \sim Bin(n, p)$

Binomialverteilung

- Wenn $X \sim Bin(n, p)$, dann gilt:
  - Träger: $D(X) = \{0, 1, 2, \ldots, n\}$
  - Wahrscheinlichkeitsfunktion

$$P(X = x) = f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{für } x = 0, 1, \ldots, n \\ 0 & \text{sonst} \end{cases}$$

  - Momente: $E[X] = np, \;\; VAR[X] = np(1-p)$
- Entscheidend: Erfolgswahrscheinlichkeit in jedem Zug gleich!
- Berechnung der Momente noch obiger Formel und Tabellierung der Wahrscheinlichkeitsfunktion und Verteilungsfunktion für verschiedene n, p-Kombinationen.

Hypergeometrische Verteilung

- Beispiel: In einer Urne befinden sich 10 (10.000) Kugeln, wobei 2 (2.000) Kugeln schwarz sind. Es werden 4 Kugeln ohne Zurücklegen gezogen. Wie groß ist die Wahrscheinlichkeit, dass genau eine schwarze Kugel gezogen wird.
- Unterschied zur Binomialverteilung: Erfolgswahrscheinlichkeit ändert sich mit jedem Zug.
- Allgemein: In einer Urne befinden sich N Kugeln, davon sind M Kugeln schwarz. Es wird eine Stichprobe vom Umfang n ohne Zurücklegen gezogen.
- Approximation der Hypergeometrischen Verteilung durch die Binomialverteilung für $n/N \leq 0.05$ (Auswahlsatz sehr klein)
- In diesem Fall gilt ungefähr: $X \sim Bin(x; n, p = M/N)$

Diskrete Gleichverteilung

- Einfachste Form einer parametrischen Verteilung
- Ausgangspunkt: Laplace-Wahrscheinlichkeit
- Situation: Nimmt eine Zufallsvariable X die Massenpunkte $x = 1, 2, \ldots, N$ jeweils mit Wahrscheinlichkeit $1/N$ an, so folgt X einer diskreten Gleichverteilung.
- Notation: $X \sim Gl(N)$
- Beispiel: Augenzahl X beim Würfelwurf mit $X \sim Gl(6)$

Diskrete Gleichverteilung

- Träger: $D(X) = \{1, 2, \dots, N\}$
- Wahrscheinlichkeitsfunktion:

$$P(X = x) = f_{Gl}(x; N) = \begin{cases} 1/N & \text{für } x = 1, 2, \dots, N \\ 0 & \text{sonst} \end{cases}$$

- Verteilungsfunktion:

$$P(X \leq x) = F_{Gl}(x; N) = \begin{cases} 0 & \text{für } x < 1 \\ 1/N & \text{für } 1 \leq x < 2 \\ 2/N & \text{für } 2 \leq x < 3 \\ \dots & \dots \\ (N-1)/N & \text{für } (N-1) \leq x < N \\ 1 & \text{für } x \geq N \end{cases}$$

- Momente: $E[X] = \dfrac{N+1}{2}$, $VAR[X] = \dfrac{N^2 - 1}{12}$

Beispiel für stetige Zufallsvariable

- Ein Reisender erreicht den Bahnsteig einer S-Bahn. Er weiß, dass die Züge im 30-Minuten-Takt verkehren, hat aber ansonsten keine Fahrplaninformationen und weiß auch nicht, wann der letzte Zug den Bahnhof verlassen hat. Er entschließt sich, auf den nächsten Zug zu warten.
- Die Zufallsvariable X sei definiert als die „Wartezeit in Minuten“.
- Wie lautet der Träger $D(X)$?
- Wie groß ist die Wahrscheinlichkeit, dass der Reisende exakt 15 Minuten warten muss?
- Wie groß ist die Wahrscheinlichkeit, dass der Reisende höchstens eine Viertelstunde (20 min, 50 min) warten muss?
- Bestimmen Sie die Verteilungsfunktion von X und leiten Sie die zugehörige Wahrscheinlichkeitsdichte ab. Zeichnen Sie die Dichte- und die Verteilungsfunktion.
- Bestimmen Sie die Wahrscheinlichkeit, dass die Wartezeit zwischen 5 und 18 Minuten liegt (a) mit Hilfe von $F(x)$, (b) mit Hilfe von $f(x)$.

Stetige Zufallsvariable

- Stetige (kontinuierliche) Zufallsvariable liegt vor, wenn die Ergebnisse eines Zufallsexperiments nicht abgezählt werden können. Die Werte können in einem Intervall beliebig (unendlich) viele Ausprägungen aufweisen. Zu viele, um sie wie bei diskreten ZV aufschreiben zu können.
- Träger: $D(X)$ umfasst ein Intervall der reellen Zahlen.
- Es gibt keine Massenpunkte d.h. keine Realisationen mit $P(X = x) > 0$. $\Rightarrow$ Punktwahrscheinlichkeit gleich Null
- Daher gibt man nicht an, mit welcher Wahrscheinlichkeit die ZV X einen bestimmten Wert annimmt, sondern mit welcher Wahrscheinlichkeit die ZV X in ein bestimmtes Intervall fällt also $P(X \leq b)$ oder auch $P(a \leq X \leq b)$

Beispiele für stetige Zufallsvariablen

Experiment	ZV	mögliche Werte
Täglicher Wasserverbrauch	in Liter	100.2, 73.8, ...
Wiegen von Personen	Gewicht in kg	60.1, 80.5, ...
Fragen nach Ausgaben für Nahrungsmittel	Ausgaben in Euro	102.3, 332.4, ...
Messen der Zeit zwischen zwei Anrufen	Zeit (min)	2.4, 3.11, ...
Messen der Abweichung eingestellte Schraubenlänge	in mm	-0.23, -0.28, +0.34, ...

- Ganze Zahlen (gerundet), Bruchzahlen oder reelle Zahlen
- Erhält man bei Messungen
- Unendlich viele mögliche Werte in einem Intervall

(Wahrscheinlichkeits-)Dichtefunktion

- Da Punktwahrscheinlichkeit gleich Null, ist WSfunktion für stetige Zufallsvariablen nicht sinnvoll $\Rightarrow$ Darstellung als Dichtefunktion
- Dichtefunktion für stetige Zufallsvariablen
  - Jedem Wert von x wird ein Dichtefunktionswert $f(x)$ zugeordnet
  - Dichtefunktionswert ist KEINE Wahrscheinlichkeit

$$f_X(x) \neq P(X = x)$$

- WS, dass stetige ZV in das Intervall a bis b fällt, entspricht der Fläche unterhalb der Dichtefunktion von a bis b

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

- Eigenschaften der Dichtefunktion
  - Die Dichtefunktion nimmt nur positive Werte an: $f(x) \geq 0$
  - Die Fläche unter der Kurve ist 1

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Verteilungsfunktion

- Die Verteilungsfunktion gibt die Wahrscheinlichkeit an, dass die Zufallsvariable einen Wert annimmt, der höchstens so groß ist wie ein bestimmter Wert.
- Diese kumulierte Wahrscheinlichkeit wird bei stetigen ZV durch das Integral von $-\infty$ bis x der Wahrscheinlichkeitsdichtefunktion bestimmt
- Verteilungsfunktion einer stetigen Zufallsvariablen:

$$P(X \leq x) = F(x) = \int_{-\infty}^{x} f(s)ds$$

Erwartungswert und Varianz für stetige ZV

- Erwartungswert (vgl. arithm. Mittel)

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

- Varianz = Mittlere quadratische Abweichung vom E-Wert

$$\sigma^2 = Var(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)dx$$

Rechteckverteilung, Stetige Gleichverteilung

- Zufallsexperiment: Nimmt eine Zufallsvariable X Werte im Intervall $[a;b]$ an und besitzen alle Teilintervalle der gleichen Länge die selbe Wahrscheinlichkeit , so folgt die Zufallsvariable X einer stetigen Gleichverteilung.
- Einfachste Form der parametrischen Verteilung für stetige Zufallsvariablen
- Träger: $D(X) = [a,b]$
- Notation: $X \sim Re(a,b)$.
- Beispiel: S-Bahn Aufgabe $\Rightarrow X \sim Re(0,30)$

Momente und Dichte- bzw. Verteilungsfunktion

- Erwartungswert und Varianz

$$E[X] = \frac{a+b}{2}, \;\; VAR[X] = \frac{(b-a)^2}{12}$$

- Dichtefunktion

$$f_{Re}(x; a, b) = \begin{cases} \dfrac{1}{b-a} & \text{für } a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$$

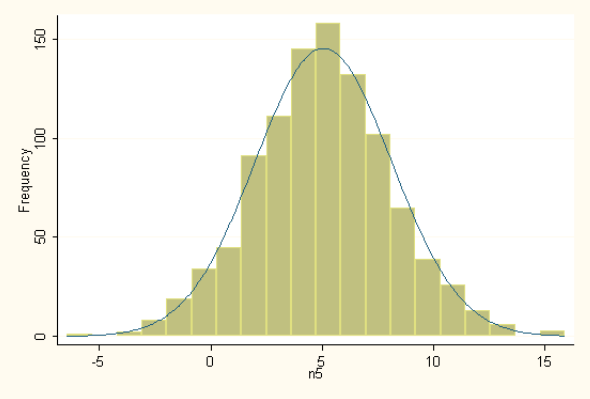
- Verteilungsfunktion:

$$F_{Re}(x; a, b) = \begin{cases} 0 & \text{für } x < a \\ \dfrac{x-a}{b-a} & \text{für } a \leq x \leq b \\ 1 & \text{für } x > b \end{cases}$$

Normalverteilung

- Normalverteilung oder Gauß Verteilung wurde von C.F. Gauß bei der Untersuchung von Messfehlern eingeführt.
- Idee: Verteilung, bei der die Werte im Zentrum der Verteilung relativ häufig auftreten. Je weiter man sich vom Zentrum entfernt, umso seltener werden die Werte.
- Die Dichtefunktion hat die Form einer symmetrischen Glockenkurve
- Die Verteilung wird durch ihren Erwartungswert μ und ihre Varianz σ^2 eindeutig bestimmt.
- Notation: $X \sim N(\mu_X, \sigma_X^2)$ mit Träger: $D(X) = \mathbb{R}$

Histogramm und Normalverteilung



Es gilt natürlich auch hier: $\int_{-\infty}^{+\infty} f(x)dx = 1$
Die Dichtefunktion nähert sich beliebig nahe der x-Achse an.

Normalverteilung als "wichtigste" Verteilung

- Viele naturwissenschaftliche Größen, aber auch Fehler haben die Eigenschaft, dass sie mit solch einer rel. Häufigkeitsverteilung auftreten, dass das zugehörige Histogramm durch eine Glockenkurve approximiert werden kann.
- Zentraler Grenzwertsatz: Die Summe einer großen Anzahl unabhängiger ZV (egal welcher Verteilung) ist annähernd normalverteilt. Diese Eigenschaft macht die Normalverteilung für Schätz- und Testverfahren so wichtig.
- Mit wachsender Anzahl von Wiederholungen nähert sich die Binomialverteilung (und unter bestimmten Voraussetzungen auch eine ganze Reihe weiterer Verteilungen) einer Normalverteilung

Momente, Dichte- und Verteilungsfunktion

- Erwartungswert und Varianz

$$E[X] = \mu \qquad VAR[X] = \sigma^2$$

- Dichtefunktion:

$$f_N(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

- Um zu zeigen, dass Fläche unterhalb der Dichtefunktion 1 ist, muss eine Grenzwertbildung vorgenommen werden.
- Verteilungsfunktion:

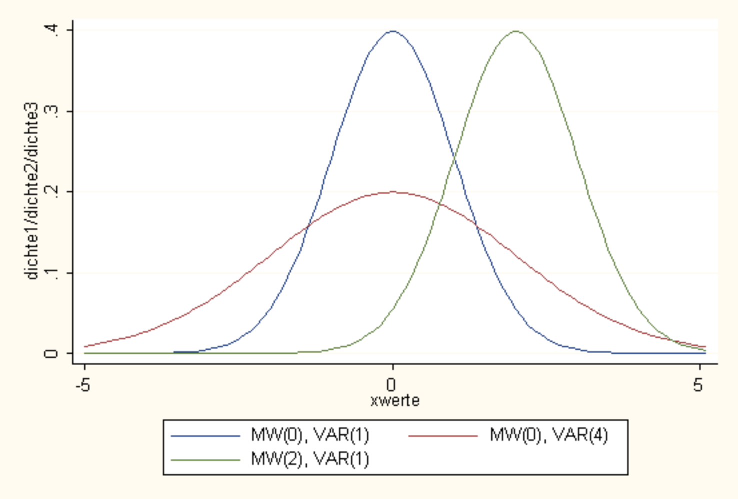
$$F_N(x; \mu, \sigma^2) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{\frac{-(z-\mu)^2}{2\sigma^2}} dz$$

Variation der Parameter μ und σ^2

- Die Normalverteilung ist symmetrisch bezüglich μ d.h. globales Maximum der Dichtefunktion entspricht dem Erwartungswert entspricht dem Median.
- Eine Variation von μ bewirkt eine Verschiebung der Glockenkurve entlang der x-Achse.
- Eine Variation von σ bewirkt eine Veränderung der Glockenöffnung (Stauchung oder Streckung)
- Für große Werte von σ ist der Kurvenverlauf flach, für kleine Werte von σ wird der Maximalwert größer und der Anstieg steiler.
- Je größer Streuung ist, umso geringer ist die Wahrscheinlichkeit für Werte, die weit entfernt von μ sind.
- Bei $x = \mu \pm \sigma$ sind die Wendepunkte der Dichtefunktion.

Abbildungen zur Normalverteilung

- Verändern von μ bzw. σ^2

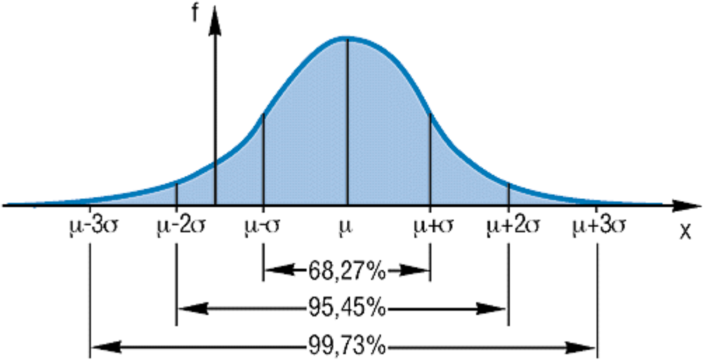


Beispiel zur Normalverteilung

- Sie sind verantwortlich für das Abfüllen von Zucker in 1kg-Tüten. Um so wenig Ausschuss wie möglich zu produzieren, muss das Füllgewicht möglichst exakt getroffen werden.
- Bei einer internen Qualitätskontrolle werden alle Tüten beanstandet, die weniger als 995g bzw. mehr als 1010g beinhalten.
- Aus langer Erfahrung wissen Sie, dass das Füllgewicht X ihrer Tüten normalverteilt ist mit Mittelwert 1000g und einer Standardabweichung von 5g.

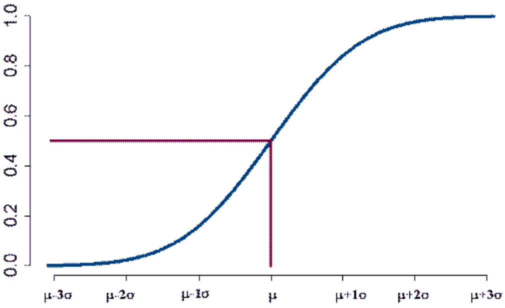
1. Mit welcher Wahrscheinlichkeit wiegt ein zufällig überprüftes Zuckerpaket (a) mehr als 1010g, (b) weniger als 995g, (c) zwischen 995 und 1010g?
2. Mit welcher Wahrscheinlichkeit wiegt ein zufällig überprüftes Zuckerpaket (a) weniger als 997g, (b) mehr als 997g, (c) weniger als 1003g

Dichte und Wahrscheinlichkeit



Gilt für alle μ-σ-Kombinationen!
Wie groß ist die WS, dass eine normalverteilte Zufallsvariable einen Wert kleiner als $\mu - 3\sigma$ oder kleiner als $\mu + 2\sigma$ annimmt?

Verteilungsfunktion



Zunächst zunehmende Steigung, dann abnehmende Steigung; Wendepunkt bei μ

Problem der Verteilungsfunktion

- Berechnung der Verteilungsfunktion äußerst komplex
  - Kein analytischer Ausdruck anzugeben
  - Näherungsweise mit numerischen Verfahren
- Theoretisch notwendig: Unendliche viele Wahrscheinlichkeitstabellen für alle μ-σ-Kombinationen $\Rightarrow$ Natürlich nicht praktikabel.
- Es gilt: Linear-affine Transformationen erhalten die Normalverteilung:

$$X \sim N(\mu_X, \sigma_X^2),\ Y = a + bX \Longrightarrow Y \sim N(a + b\mu_X, b^2\sigma_X^2).$$

  d.h. jede normalverteilte Zufallsvariable X mit μ_X und σ_X^2 kann durch geeignete Transformation in eine normalverteilte Zufallsvariable Y mit μ_Y und σ_Y^2 umgewandelt werden.
- Idee: Verteilungsfunktion für Referenzverteilung und Tabellierung der Wahrscheinlichkeiten nur für diese Verteilung

Standardnormalverteilung

- Wenn $Z \sim N(0,1)$ ist, dann folgt Z der so genannten Standardnormalverteilung.
- Bezeichnung:

$$\phi(z) = f(z;0,1) = \frac{1}{\sqrt{2\pi}} \exp\{-z^2/2\}$$

$$\Phi(z) = F(z;0,1) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\{-y^2/2\} dy$$

- Standardnormalverteilung wird als Referenzverteilung verwendet. Die Werte ihrer Verteilungsfunktion werden tabelliert (siehe Formelsammlung).
- Jede andere Normalverteilung kann mit Hilfe der so genannten Z-Transformation in eine Standardnormalverteilung umgewandelt werden:

$$P(X \leq x) = F_N(x;\mu,\sigma^2) = F_N(z;0,1) = \Phi(z) \qquad \text{mit} \quad z = \frac{x-\mu}{\sigma}$$

Verteilungsfunktion der Standardnormalverteilung

y	0	1	2	3	4	5	6	7	8	9
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767

Verteilungsfunktion der Standardnormalverteilung

y	0	1	2	3	4	5	6	7	8	9
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.7	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.8	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.9	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.

Quantile der Standardnormalverteilung

- Fortsetzung Beispiel: Welches Gewicht wird von einer zufällig der Produktion entnommenen Zuckerpackung mit einer Wahrscheinlichkeit von (a) 50% (b) 5% (c) 99% nicht überschritten?
- Häufig such man zu einer gegebenen Wahrscheinlichkeit α den dazu gehörigen x-Wert der Verteilungsfunktion

  $$F(x) = P(X \leq x) = \alpha$$

  Dieser wird als α-Quantil bezeichnet (Konzept aus deskriptiver Statistik bekannt!).
- Die α-Quantile der Standardnormalverteilung werden mit λ_α bezeichnet.
- Lösung für $P(X \leq x) = 0.99$
  - Zunächst: Für welches z ($Z \sim N(0,1)$) gilt 0.99 ⇒ FS: in etwa 2.33 ⇒ Tabelle mit Quantilen: $Z = 2.3263$
  - Aufgrund von $Z = (X - \mu_X)/\sigma_X$ gilt für X:

    $$x = z \cdot \sigma_X + \mu_X = 2.3263 \cdot 5 + 1000 = 1011.615$$

Quantile der Standardnormalverteilung

P	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
50	0.0000	.0025	.0050	.0075	.0100	.0125	.0150	.0175	.0201	.0226
51	.0251	.0276	.0301	.0326	.0351	.0376	.0401	.0426	.0451	.0476
52	.0502	.0527	.0552	.0577	.0602	.0627	.0652	.0677	.0702	.0728
53	.0753	.0778	.0803	.0828	.0853	.0878	.0904	.0929	.0954	.0979
54	.1004	.1030	.1055	.1080	.1105	.1130	.1156	.1181	.1206	.1231
55	.1257	.1282	.1307	.1332	.1358	.1383	.1408	.1434	.1459	.1484
56	.1510	.1535	.1560	.1586	.1611	.1637	.1662	.1687	.1713	.1738
57	.1764	.1789	.1815	.1840	.1866	.1891	.1917	.1942	.1968	.1993
58	.2019	.2045	.2070	.2096	.2121	.2147	.2173	.2198	.2224	.2250
59	.2275	.2301	.2327	.2353	.2378	.2404	.2430	.2456	.2482	.2508
60	.2533	.2559	.2585	.2611	.2637	.2663	.2689	.2715	.2741	.2767
61	.2793	.2819	.2845	.2871	.2898	.2924	.2950	.2976	.3002	.3029
62	.3055	.3081	.3107	.3134	.3160	.3186	.3213	.3239	.3266	.3292
63	.3319	.3345	.3372	.3398	.3425	.3451	.3478	.3505	.3531	.3558
64	.3585	.3611	.3638	.3665	.3692	.3719	.3745	.3772	.3799	.3826
65	.3853	.3880	.3907	.3934	.3961	.3989	.4016	.4043	.4070	.4097
66	.4125	.4152	.4179	.4207	.4234	.4261	.4289	.4316	.4344	.4372
67	.4399	.4427	.4454	.4482	.4510	.4538	.4565	.4593	.4621	.4649
68	.4677	.4705	.4733	.4761	.4789	.4817	.4845	.4874	.4902	.4930
69	.4958	.4987	.5015	.5044	.5072	.5101	.5129	.5158	.5187	.5215
70	.5244	.5273	.5302	.5330	.5359	.5388	.5417	.5446	.5476	.5505
71	.5534	.5563	.5592	.5622	.5651	.5681	.5710	.5740	.5769	.5799
72	.5828	.5858	.5888	.5918	.5948	.5978	.6008	.6038	.6068	.6098
73	.6128	.6158	.6189	.6219	.6250	.6280	.6311	.6341	.6372	.6403
74	.6433	.6464	.6495	.6526	.6557	.6588	.6620	.6651	.6682	.6713

Quantile der Standardnormalverteilung

P	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
75	.6745	.6776	.6808	.6840	.6871	.6903	.6935	.6967	.6999	.7031
76	.7063	.7095	.7127	.7160	.7192	.7225	.7257	.7290	.7323	.7356
77	.7388	.7421	.7454	.7488	.7521	.7554	.7588	.7621	.7655	.7688
78	.7722	.7756	.7790	.7824	.7858	.7892	.7926	.7961	.7995	.8030
79	.8064	.8099	.8134	.8169	.8204	.8239	.8274	.8310	.8345	.8381
80	.8416	.8452	.8488	.8524	.8560	.8596	.8632	.8669	.8705	.8742
81	.8779	.8816	.8853	.8890	.8927	.8965	.9002	.9040	.9078	.9116
82	.9154	.9192	.9230	.9269	.9307	.9346	.9385	.9424	.9463	.9502
83	.9542	.9581	.9621	.9661	.9701	.9741	.9781	.9822	.9863	.9904
84	.9945	.9986	1.0027	1.0069	1.0110	1.0152	1.0194	1.0237	1.0279	1.0322
85	1.0364	1.0407	1.0450	1.0494	1.0537	1.0581	1.0625	1.0669	1.0714	1.0758
86	1.0803	1.0848	1.0893	1.0939	1.0985	1.1031	1.1077	1.1123	1.1170	1.1217
87	1.1264	1.1311	1.1359	1.1407	1.1455	1.1503	1.1552	1.1601	1.1650	1.1700
88	1.1750	1.1800	1.1850	1.1901	1.1952	1.2004	1.2055	1.2107	1.2160	1.2212
89	1.2265	1.2319	1.2372	1.2426	1.2481	1.2536	1.2591	1.2646	1.2702	1.2759
90	1.2815	1.2873	1.2930	1.2988	1.3047	1.3106	1.3165	1.3225	1.3285	1.3346
91	1.3408	1.3469	1.3532	1.3595	1.3658	1.3722	1.3787	1.3852	1.3917	1.3984
92	1.4051	1.4118	1.4187	1.4255	1.4325	1.4395	1.4466	1.4538	1.4611	1.4684
93	1.4758	1.4833	1.4908	1.4985	1.5063	1.5141	1.5220	1.5301	1.5382	1.5464
94	1.5548	1.5632	1.5718	1.5805	1.5893	1.5982	1.6072	1.6164	1.6258	1.6352
95	1.6448	1.6546	1.6646	1.6747	1.6849	1.6954	1.7060	1.7169	1.7279	1.7392
96	1.7507	1.7624	1.7744	1.7866	1.7991	1.8119	1.8250	1.8384	1.8522	1.8663
97	1.8808	1.8957	1.9110	1.9268	1.9431	1.9600	1.9774	1.9954	2.0141	2.0335
98	2.0537	2.0748	2.0969	2.1201	2.1444	2.1701	2.1973	2.2262	2.2571	2.2904
99	2.3263	2.3656	2.4089	2.4572	2.5121	2.5758	2.6520	2.7477	2.8781	3.0901

Symmetrie der Standardnormalverteilung

- Allgemein: Eine Zufallsvariable X ist symmetrisch um den Punkt a verteilt, wenn $X - a$ und $-(X - a) = a - X$ dieselbe Verteilung besitzen
- Standardnormalverteilung ist symmetrisch um 0 (allgemein: μ)
- Symmetriebeziehung für
  - Dichtefunktionswert: $\phi(z) = \phi(-z)$
  - Verteilungsfunktion: $\Phi(z) = 1 - \Phi(-z)$
  - Quantile: $\lambda_\alpha = -\lambda_{1-\alpha}$
- Vorteile der Symmetrie: Eine Tabellierung von Dichte- und Verteilungsfunktionswerten ist nur für $z \geq 0$ und eine Tabellierung von Quantilen λ_α nur für $\alpha \geq 0.5$ nötig.

Approximaton der Binomialverteilung durch Normalverteilung

- Für $np(1-p) > 9$ gilt:

$$F_{Bin}(x; n, p) \approx F_N(x; \mu = np, \sigma^2 = np(1-p))$$

- Problem: Approximation einer diskreten (z.B. Binomialverteilung) durch eine stetige Verteilung (z.B. Normalverteilung) $\Rightarrow$ Stetigkeitskorrektur

$$\begin{aligned} F_{Bin}(x; n, p) &\approx F_N(x + 0.5; \mu = np, \sigma^2 = np(1-p)) \\ &= \Phi\left(\frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

Beispiel

- Der IQ in der deutschen Bevölkerung im Alter zwischen 16 und 60 sei normalverteilt mit Mittelwert 100 und Standardabweichung 15.
- In einer Studie wird untersucht ob es regionale Unterschiede im IQ gibt. Hierzu werden unter anderem in einer Großstadt XY 40 Personen zufällig aus dem Einwohnermelderegister gezogen. Dieser Personen müssen dann an einem IQ-Test teilnehmen.
- Der mittlere IQ für diese 40 Personen (Stichprobenmittelwert) liegt bei 94.61, die Stichprobenstandardabweichung beträgt 14.46.
- Bei einer weiteren einfachen Zufallsstichprobe von 40 Personen liegt der Stichprobenmittelwert bei 96.34 und die Stichprobenstandardabweichung bei 16.46.
- Besitzen die Einwohner die Stadt XY im Mittel einen geringeren IQ als die deutsche Bevölkerung insgesamt?

Stichproben und Stichprobenfunktionen

1. Stichproben
2. Stichprobenfunktionen
3. Verteilung von Stichprobenfunktionen
4. Beziehung zwischen den Momenten von Stichprobenmomenten und den Momenten der Grundgesamtheit
5. Asymptotische Eigenschaften
6. Stichproben aus normalverteilter Grundgesamtheit

Problemstellung

- Bislang wurde stets davon ausgegangen, dass die Verteilung (Art und Parameter) einer Zufallsvariable bekannt ist. Darauf aufbauend wurden Wahrscheinlichkeiten berechnet, dass die Zufallsvariable einen bestimmten Wert annimmt bzw. in ein bestimmtes Intervall fällt.
- Jetzt wird der realistischere Fall betrachtet, dass zwar die Art der Verteilung bekannt ist, nicht aber deren Parameter.
- Die Aufgabe liegt nun darin, Informationen über diese unbekannten Parameter der statistischen Grundgesamtheit zu beschaffen.
- Teilerhebung in Form einer zufälligen Stichprobe und Schlussfolgerungen von Stichprobe auf Grundgesamtheit
- Schlussfolgerung immer unter Unsicherheit $\Rightarrow$ Versuch diese Unsicherheit zu quantifizieren

Zufallsstichprobe

- Von einer Zufallsstichprobe spricht man, wenn jedes Element der Grundgesamtheit eine positive, mit den Regeln der WSrechnung berechenbare WS besitzt, in die Stichprobe zu gelangen.
- Ob es sich bei einer Stichprobe um eine Zufallsstichprobe handelt, wird also durch das Auswahlverfahren entschieden.
- Uneingeschränkte Zufallsauswahl: Jedes Element der Grundgesamtheit hat die gleiche Wahrscheinlichkeit in die Stichprobe zu gelangen.
- Komplexe Zufallsauswahl: Geschichtete Auswahl, Klumpenauswahl, mehrstufige Auswahl
- Achtung: Willkürliche Auswahl d.h. keine Kontrolle durch Auswahlplan (Beispiel: Interviews in der Fußgängerzone)
- Stichprobe des Umfangs n aus der Grundgesamtheit G:
  n-elementige Teilmenge der Grundgesamtheit

$$\{e_1, e_2, \ldots, e_n\} \subseteq G.$$

Einfache Stichprobe

- Sei X eine Zufallsvariable, die für jedes Element einer Stichprobe des Umfangs n erhoben wird, dann heißt

  $$\mathbf{X} = (X_1, \ldots, X_n)$$

  n-dimensionale Stichprobenvariable.

- Einfache Zufallsstichprobe: Alle X_i $(i = 1, 2, \ldots, n)$ sind stochastisch unabhängig und identisch verteilt.
  - Unabhängige Zufallsstichprobe: Alle X_i $(i = 1, 2, \ldots, n)$ sind stoch. unabhängig, d. h. insbesondere, dass das Ergebnis des i–ten Zuges unabhängig von den Ergebnissen anderer Züge ist.
  - Identisch verteilte Zufallsstichprobe: Alle X_i $(i = 1, 2, \ldots, n)$ besitzen dieselbe Verteilung

- Beispiel: Es werden aus dem Einwohnermelderegister der Großstadt XY zufällig $n = 40$ Personen zwischen 16-60 Jahren ausgewählt, die verpflichtet werden an einem IQ-Test teilzunehmen.

Parametrische Statistik

- Verteilung in der Grundgesamtheit folge einem parametrischen Verteilungsgesetz

$$F_X(x) = F_X(x;\theta), \qquad \theta \in \Theta,$$

- X: "IQ eines Einwohners zwischen 16-60 Jahren" folgt einer Normalverteilung mit unbekanntem Mittelwert μ und Varianz σ^2.
- Parametrische Statistik: Rückschlüsse (Inferenz) auf einen unbekannten Parameter θ der GG anhand der beobachteten Stichprobenwerte.
- Mithilfe der Stichprobe von 40 Personen sollen Rückschlüsse auf Mittelwert und Varianz in der Grundgesamtheit gezogen werden.
- Gesucht: Schätzwert $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$
- Gesucht: Schätzwerte $\widehat{\mu}$ und $\widehat{\sigma}^2$ für unbekannte Parameter μ und σ^2 der GG als Funktion der Stichprobenwerte $x_1, x_2, \ldots, x_{40}$.
- Beachte: Stets wird per Annahme vorausgesetzt, dass die betrachtete Stichprobe einfach ist (identisch unabhängig verteilt)

Stichprobenmomente

- WICHTIG: Jede Zufallsstichprobe ist zufällig, d.h. die Stichprobenvariablen sind zufällig.
- Stichprobenmomente d.h. Funktionen der Stichprobenvariablen, werden zur Schätzung der unbekannten Parameter in der Grundgesamtheit herangezogen.
- Stichprobenmittel

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

- Stichprobenvarianz

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

- WICHTIG: Wenn die Stichprobenvariablen zufällig sind, dann sind auch Stichprobenmittel und -varianz Zufallsvariablen.

Beispiel für Stichprobenmomente

- Zufallsvariable X: IQ eines Einwohners der Stadt XY
- Zufallsstichprobe 1 mit Umfang $n = 40$
  $\Rightarrow$ 76.27, 96.33, 85.80, 124.93 ... 99.08, 111.32, 83.04 mit Stichprobenmittelwert 94.61 und Stichprobenstandardabw. 14.46
- Zufallsstichprobe 2 mit Umfang $n = 40$
  $\Rightarrow$ 91.61, 109.91, 100.35, 109.02 ... 110.69, 90.65, 85.57 mit Stichprobenmittelwert 96.34 und Stichprobenstandardabw. 16.46
- Zufallsstichprobe 3 mit Umfang $n = 40$
  $\Rightarrow$ 99.08, 118.81, 114.10, 95.74 ... 108.68, 100.46, 83.55 mit Stichprobenmittelwert 100.13 und Stichprobenstandardabw. 13.76
- Zufallsstichprobe 4 mit Umfang $n = 40$
  $\Rightarrow$ 103.74, 99.53, 133.07, 87.02 ... 99.12, 107.85, 88.72 mit Stichprobenmittelwert 103.57 und Stichprobenstandardabw. 16.15

Stichprobenmittelwert

- Für jede Zufallsstichprobe aus einer Grundgesamtheit kann der Stichprobenmittelwert berechnet werden. [Jede Grundgesamtheit besitzt einen Erwartungswert.]
- Stichprobenmittelwert ist Zufallsvariable. [Erwartungswert der Grundgesamtheit ist KEINE Zufallsvariable.]
- Stichprobenmittelwert folgt einer bestimmten Verteilung (wie jede andere Zufallsvariable auch).
- Im Folgenden interessiert: Erwartungswert des Stichprobenmittelwerts, Varianz des Stichprobenmittelwerts und Verteilung des Stichprobenmittelwerts

Erwartungswert des Stichprobenmittels und in der Grundgesamtheit

Seien $X_1, \ldots, X_n$ stochastisch unabhängig, identisch verteilt mit Mittelwert μ.

$$E[\overline{X}_n] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu.$$

Interpretation: Wenn man sehr viele einfache Stichproben des Umfangs n zieht, für jede Stichprobe das Stichprobenmittel berechnet, dann aus diesen Stichprobenmittelwerten wiederum den Mittelwert, dann entspricht dieser dem Mittelwert der Grundgesamtheit.

Zusammenhang zwischen Stichprobenmomenten und den Momenten der Grundgesamtheit

$$E[\overline{X}_n] = \mu_X \qquad VAR[\overline{X}_n] = \sigma_X^2/n$$

Zusätzlich gilt:

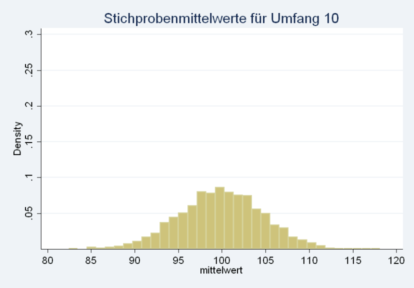
$$E[S_n^2] = \sigma_X^2$$

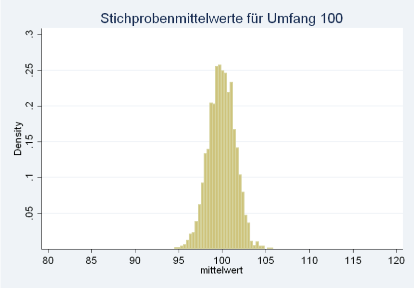
$\sqrt{n}$-Gesetz: Die Standardabweichung des Stichprobenmittels ist umso kleiner, je größer der Stichprobenumfang ist. Sie sinkt mit der Rate $\sqrt{n}$.

Oder: Das Stichprobenmittel liegt umso dichter beim Erwartungswert der Grundgesamtheit, je größer der Stichprobenumfang n ist, wobei der Abstand durch die mittlere quadratische Abweichung (d.h. die Varianz) gemessen wird.

Verteilung des Stichprobenmittels

Histogramm der Stichprobenmittelwerte bei 2000 Stichproben mit jeweils $n = 10$ bzw. $n = 100$ $(X \sim N(100, 15^2))$





Verteilung des Stichprobenmittels

- Wenn $(X_1, X_2, \ldots, X_n)$ eine einfache Stichprobe aus einer normalverteilten Grundgesamtheit ist, d. h. $X \sim N(\mu, \sigma^2)$, dann gilt

$$\overline{X}_n \sim N(\mu, \sigma^2/n) \quad \text{bzw.} \quad P\left(\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \leq x\right) = \Phi(x)$$

- Verteilung des Stichprobenmittels ermöglicht es, die Wahrscheinlichkeit zu berechnen, dass das Stichprobenmittel in ein bestimmtes Intervall fällt.
- Beispiel auf S. 69: Falls die IQ-Verteilung in der Stadt XY gleich der in Deutschland ist, wie wahrscheinlich ist es dann bei $n = 40$ ein Stichprobenmittel von kleiner als 95 bzw. 97 zu erhalten? Was folgern Sie aus Ihrem Ergebnis?

Verteilung des Stichprobenmittels bei unbekannter Varianz

- Ist allein der Mittelwert von X bekannt, nicht aber die Varianz σ^2, so kann verwendet werden, dass

$$\sqrt{n}\frac{\overline{X}_n - \mu}{S_n} \sim t(k = n - 1)$$

  t-verteilt ist mit $k = n - 1$ „Freiheitsgraden".

- t-Verteilung ist ähnlich der Standardnormalverteilung hat aber mehr Wahrscheinlichkeitsmasse (d.h. mehr Unsicherheit) in den Rändern.
- Verteilung des Stichprobenmittels in Abhängigkeit von der Stichprobenvarianz ermöglicht es, die Wahrscheinlichkeit zu berechnen, dass das Stichprobenmittel in ein bestimmtes Intervall fällt, selbst wenn die Varianz der Grundgesamtheit unbekannt ist.
- Beispiel auf S. 69: Nehmen Sie nun an, dass die Stadt XY den gleichen mittleren IQ besitzt wie Deutschland insgesamt, aber keine Informationen bezüglich der Standardabweichung der GG vorliegen. Wie wahrscheinlich ist es nun bei $n = 40$ ein Stichprobenmittel von kleiner als 95 bzw. 97 zu erhalten?

Asymptotische Eigenschaft des Stichprobenmittels

1. Chebychev-Gesetz der großen Zahlen
2. Zentraler Grenzwertsatz

Chebyshev-Gesetz der großen Zahlen

Seien $X_1, X_2, \ldots$ stochastisch unabhängig, identisch verteilt mit Mittelwert μ_X und Varianz σ_X^2.
Für alle (noch so winzig kleinen) $\varepsilon > 0$ gilt:

$$\lim_{n\to\infty} P(\mu_X - \epsilon < \overline{X}_n < \mu_X + \epsilon) = 1.$$

Interpretation: Wenn man nur den Stichprobenumfang genügend groß wählt, dann liegt das Stichprobenmittel jeder Stichprobe mit Sicherheit beliebig nahe beim Mittelwert der Grundgesamtheit.

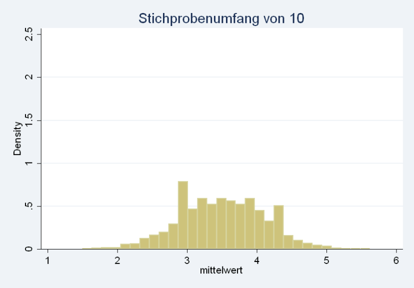
Noch ein Beispiel (Teil 1)

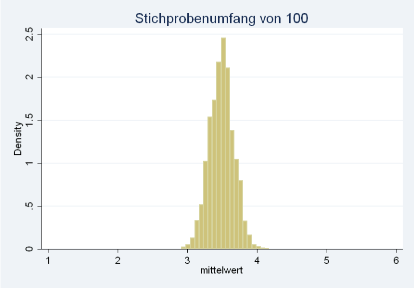
Die Zufallsvariable X: Augenzahl bei einfachen Würfelwurf ist gleichverteilt mit $N = 6$.
Der Erwartungswert beträgt 3.5.
Sie würfeln 5 Mal und erhalten folgendes Ergebnis: $\{$5,4,3,1,6$\}$.
Sie führen das Experiment noch 10 Mal durch:

Nr.	Zahl 1	Zahl 2	Zahl 3	Zahl 4	Zahl 5	MW
1	5	4	3	1	6	3.8
2	4	2	4	3	6	3.8
3	4	4	2	1	4	3
4	4	6	1	3	6	4
5	3	1	4	6	6	4
6	4	1	4	5	5	3.8
7	2	1	1	1	2	2
8	3	6	3	4	1	3.4
9	3	1	1	5	1	2.2
10	6	4	2	2	6	4

Noch ein Beispiel (Teil 2)

Histogramm der Stichprobenmittelwerte bei 3000 Stichproben mit jeweils $n = 5$ bzw. $n = 100$





Zentraler Grenzwertsatz

Sind $(X_1, X_2, \ldots, X_n)$ stochastisch unabhängige, in der Grundgesamtheit mit Mittelwert μ_X und Varianz σ_X^2 verteilte Zufallsvariable, dann ist das Stichprobenmittel für großen Stichprobenumfang n annähernd (approximativ) normalverteilt mit

$$E[\overline{X}_n] = \mu_X \qquad \text{und} \qquad VAR[\overline{X}_n] = \sigma_X^2/n.$$

d.h. es gilt

$$\overline{X}_n \overset{appr}{\sim} N(\mu_X, \sigma_X^2/n) \quad \Rightarrow \quad \frac{\overline{X}_n - \mu_X}{\sigma_X/\sqrt{n}} \overset{appr}{\sim} N(0,1)$$

Für Stichproben aus normalverteilten Grundgesamtheiten ist das Stichprobenmittel unabhängig vom Stichprobenumfang stets exakt normalverteilt (Reproduktivität der Normalverteilung)

Fazit

> Das schwache Gesetz der großen Zahlen besagt, dass das Stichprobenmittel gegen μ_X konvergiert, der Zentrale Grenzwertsatz besagt darüber hinaus, wie das Stichprobenmittel konvergiert (nämlich annähernd mit einer Normalverteilung).

Faustregel: Für $n > 30$ ist das arithmetische Mittel der Stichprobe in guter Näherung normalverteilt.

Stichprobenanteilswert

- Ausgangspunkt: Bernoulli-Experiment und Binomialverteilung
- Analog zum Stichprobenmittelwert und Stichprobenvarianz wird hier als Stichprobenfunktion der Stichprobenanteilswert A_n betrachtet.
- Beispiel:
  - Ein Versandhandelunternehmen möchte wissen, ob seine Kunden Interesse an einer persönlichen Stilberatung haben. Hierzu werden aus der Kundendatei zufällig $n = 100$ Personen telefonisch befragt.
  - X: "Anzahl der Personen in der Stichprobe, die Interesse an einer persönlichen Stilberatung haben" ist binomialverteilt mit $n = 100$ und $p =?$.
  - Es gilt: $E(X) = np = 100p$ und $Var(X) = 100p(1-p)$
  - In der Stichprobe haben 15 Personen Interesse an einer persönlichen Stilberatung, d.h. der Stichprobenanteilswert A_n beträgt $x/n = 15/100 = 0.15$

Stichprobenverteilung des Anteilswertes

- Der Stichprobenanteilswert $A_n = X/n$ ist eine Stichprobenfunktion und damit eine Zufallsvariable.
- Es gilt:

$$E(A) = E(X/n) = \frac{1}{n} \cdot E(X) = \frac{1}{n} np = p$$

$$Var(A) = Var(X/n) = \frac{1}{n^2} Var(X) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$$

- Auch hier gilt der zentrale Grenzwertsatz, d.h. A_n ist approximativ normalverteilt mit $\mu = p$ und $\sigma^2 = p(1-p)/n$.
- Beispiel: Wie groß ist die WS, dass der Stichprobenanteilswert bei einer Befragung von $n = 200$ Kunden kleiner als 0.12 ist, wenn tatsächlich 18% aller Kunden an einer persönlichen Stilberatung interessiert sind?

Parameterschätzung

- Zielsetzung der schließenden Statistik: Schätzen der unbekannten Parameter der GG und Testen von Hypothesen bezüglich der unbekannten Parameter
- Zwei Arten von Schätzung:
  1. Punktschätzung: Auf Basis der Stichprobe soll ein möglichst "guter" Schätzwert für unbekannten Parameter der GG berechnet werden (Schätzfunktion)
  2. Intervallschätzung: Auf Basis der Stichprobe soll ein bestimmter Bereich geschätzt werden, in dem der unbekannte Parameter der GG mit "bestimmter WS liegt".

Punktschätzung

- Schätzfunktion als Funktion der Stichprobenvariablen gibt an wie der unbekannte GG-Parameter geschätzt werden soll
- Symbol für Schätzfunktion: "Dach" z.B. $\widehat{\mu}$
- Beispiel: Punktschätzung für unbekannten Mittelwert der GG $\Rightarrow$ Es bieten sich an: Stichprobenmittelwert oder auch Stichprobenmedian
- Schätzwert ist der aus den Stichprobenrealisationen berechnete numerische Wert der Schätzfunktion.
- Beispiel: Es soll der durchschnittliche IQ der Studierenden in Erlangen geschätzt werden. Dazu werden zufällig 50 Studierende ausgewählt. Der Stichprobenmedian beträgt 109, der Stichprobenmittelwert 115.3
  - Schätzfunktion:

$$\widehat{\mu} = f(X_1, \ldots, X_{50}) = \overline{X}_{50} = \frac{1}{50} \sum_{i=1}^{50} X_i$$

  - Schätzwert:

$$\overline{x}_{50} = \frac{1}{50} \sum_{i=1}^{50} x_i = 115.3$$

Beurteilungskriterien für Punktschätzer

- Erwartungstreue
  - Im Mittel (über unendlich viele Stichproben) ist der Punktschätzer gleich dem unbekannten Parameter der Grundgesamtheit
  - Stichprobenmittel ist erwartungstreuer Schätzer für den GG-Mittelwert.
  - Stichprobenmedian ist nur bei symmetrischen Verteilungen ein erwartungstreuer Schätzer.
  - Stichprobenvarianz ist erwartungstreuer Schätzer für die GG-Varianz.
- Das Stichprobenmittel ist absolut effizient, d.h. es gibt keine andere erwartungstreue Schätzfunktion für den unbekannten Mittelwert der GG, die eine kleinere Varianz besitzt.

Schätzmethoden

- Allgemeine Prinzipien nach denen aus Stichprobenvariablen Schätzfunktionen gebildet werden.
- Standard-Methoden
  - Momentenmethode d.h. zur Schätzung der unbekannten Momente in der Grundgesamtheit werden die Momente der Stichprobe verwendet.
  - Methode der kleinsten Quadrate
  - Maximum-Likelihood-Methode

Maximum-Likelihood-Prinzip

Beim Maximum-Likelihood-Ansatz unterscheidet man grundsätzlich zwei Situationen:

A. Bevor die Stichprobe tatsächlich gezogen wird, ist nur die Verteilung der möglichen Stichprobenergebnisse in Abhängigkeit vom Parameterwert θ bekannt:

$$f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n;\theta)=\prod_{i=1}^{n} f_X(x_i;\theta)$$

D.h. die gemeinsame Verteilung hängt von $x_1,\ldots,x_n$ und dem unbekannten θ ab.

B. Nachdem die Stichprobe gezogen wurde, ist der Wert der Verteilung des konkreten Stichprobenergebnisses in Abhängigkeit vom Parameterwert θ bekannt:

$$L(\theta;x_1,x_2,\ldots,x_n)=\prod_{i=1}^{n} f_X(x_i;\theta)$$

D.h. die gemeinsame Verteilung hängt nur noch von dem unbekannten θ ab, da $x_1,\ldots,x_n$ bekannt sind.

Likelihoodfunktion

- Die Funktion

$$L(\theta) = L(\theta; x_1, \ldots, x_n)$$

heißt Likelihoodfunktion.

- Diskrete Zufallsvariable: Likelihhoodfunktion = Gemeinsame Wahrscheinlichkeitsfunktion.
- Stetige Zufallsvariable: Likelihoodfunktion = Gemeinsame Dichtefunktion (keine Wahrscheinlichkeit).

Likelihoodfunktion für eine Bernoulli-verteilte Zufallsvariable I

- Sei X die Zufallsvariable, die den Wert 1 annimmt, wenn beim Münzwurf "Kopf" oben liegt. Der unbekannte Parameter ist die Wahrscheinlichkeit p, dass "Kopf" oben liegt.
- Bevor eine Stichprobe gezogen wird, d.h. ein 10-maliger Münzwurf vorgenommen wird, kennt man in Abhängigkeit von dem unbekannten p die gemeinsame Verteilung der Stichprobenergebnisse:

$$P(X_1 = x_1, \ldots, X_n = x_n; p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}.$$

- Nach der Stichprobenziehung liegt z.B. ein Stichprobenbefund

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1	0	0	1	1	0	0	1	1	1

vor.

Likelihoodfunktion für eine Bernoulli-verteilte Zufallsvariable II

- Die Wahrscheinlichkeit dieses Stichprobenergebnisses in Abhängigkeit von p ist

$$L(p) = p^6(1-p)^4.$$

- Wahl von p: Wähle p so, dass die Wahrscheinlichkeit am größten ist, dass der konkrete Stichprobenbefund aus einer Grundgesamtheit mit diesem Parameterwert p stammt.
- D.h.: Maximiere die Likelihoodfunktion bez. p.
- Vereinfachung der Maximierung: Da nur die Stelle des Maximums und nicht der Wert der Zielfunktion an der Stelle des Maximums interessiert, kann auch die logarithmierte Likelihoodfunktion

$$\ln L(p) = \ln L(p; x_1, \ldots, x_n) = 6 \ln p + 4 \ln(1-p)$$

maximiert werden.

$L(p)$ und $\ln L(p)$ besitzen das Maximum an derselben Stelle p^*.

Likelihoodfunktion für eine Bernoulli-verteilte Zufallsvariable III

- Ergebnis der Maximierung:

$$\frac{\partial \ln L(p)}{\partial p} = \frac{6}{p} - \frac{4}{1-p} \stackrel{!}{=} 0$$

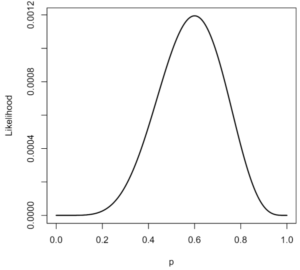
  liefert $\hat{p}_{ML} = 6/10 = 0.6$.

- D.h. am wahrscheinlichsten ist es, dass der Stichprobenbefund aus einer Grundgesamtheit mit $p = 0.6$ stammt, also mit 60%-iger Wahrscheinlichkeit "Kopf" oben liegt.
- Wie sähe das Ergebnis aus für eine 100malige Wiederholung des Münzwurf, wobei 55mal "Kopf" oben liegt?

Likelihoodfunktion in **R** I

- Plotten der Likelihoodfunktion für $n = 10$ und $\sum_{i=1}^{10} = 6$:

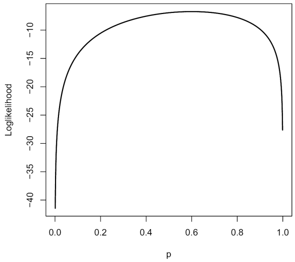
```
>p=seq(0,1,0.001)
>l=p^6*(1-p)^4
>plot(p,l,"l",lwd=2)
```



Likelihoodfunktion in **R** II

- Plotten der logarithmierten Likelihoodfunktion für $n = 10$ und $\sum_{i=1}^{10} = 6$:

```
>p=seq(0,1,0.001)
>l=p^6*(1-p)^4
>ll=log(l)
>plot(p,ll,"l",lwd=2)
```



Maximum-Likelihood-Schätzung für eine Bernoulli-verteilte Zufallsvariable I

- Es soll eine allgemeine Schätzformel angegeben werden, die nicht nur für den Stichprobenbefund $n = 10$ und $\sum_{i=1}^{10} x_i = 6$ einen Schätzwert liefert.
- Likelihoodfunktion für ein beliebiges Stichprobenergebnis $x_1, x_2, \ldots, x_n$:

$$L(p) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}$$

- Logarithmierte Likelihoodfunktion für ein beliebiges Stichprobenergebnis $x_1, x_2, \ldots, x_n$:

$$\ln L(p) = \sum_{i=1}^{n} (x_i \ln p + (1 - x_i) \ln(1-p)).$$

Maximum-Likelihood-Schätzung für eine Bernoulli-verteilte Zufallsvariable II

- Notwendige Maximumsbedingung:

$$\frac{\partial \ln L(p)}{\partial p} = \sum_{i=1}^{n} \left(\frac{x_i}{p} - \frac{1 - x_i}{1 - p} \right) \stackrel{!}{=} 0.$$

- Durchmultiplizieren mit $p(1-p) > 0$ und Auflösen nach p ergibt den ML-Schätzer:

$$\hat{p}_{ML} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- D.h. die unbekannte Wahrscheinlichkeit, mit der "Kopf" beim Münzwurf oben liegen bleibt, wird durch den Anteil der "Köpfe" in der Stichprobe geschätzt.
- Hinreichende Maximumsbedingung:

$$\frac{\partial^2 \ln L(p)}{\partial p^2} = \sum_{i=1}^{n} \left(-\frac{x_i}{p^2} - \frac{1 - x_i}{(1-p)^2} \right) < 0$$

für alle $p \in (0, 1)$. Somit besitzt die logarithmierte Likelihoodfunktion ein gleichmäßiges (konkaves) Krümmungsverhalten, so dass ein globales Maximum vorliegt.

Vom Punktschätzer zum Intervallschätzer

- Punktschätzer: Ausgehend von **einer** gezogenen Stichprobe wird ein Schätzwert (z.B. Stichprobenmittelwert oder Stichprobenanteilswert) als Punktschätzung (z.B. für unbekannten Mittelwert oder Anteilswert in der Grundgesamtheit) ermittelt.
- Erwartungstreuer Punktschätzer: Im Mittel über **unendlich viele** Stichproben wird der unbekannte Parameter exakt getroffen.
- Problem: Wie "vertrauenswürdig" ist der eine realisierte Schätzwert?
- Einschätzung mithilfe eines so genannten Konfidenzintervalls $\Rightarrow$ Intervallschätzer
- Als 95%-Konfidenzintervall wird dann der Bereich bezeichnet, der bei 95% aller möglichen Stichproben (der Größe n) den unbekannten (also gesuchten) Parameter der GG überdeckt.
- Ausgehend von **einer** gezogenen Stichprobe wird dann solch ein Konfidenzintervall um den Punktschätzer konstruiert.

Konfidenzintervall

- Beispiel: Die Zufallsvariable X: "IQ der Einwohner von XY" sei normalverteilt. Um den unbekannten Parameter μ_X der Grundgesamtheit zu schätzen, werden zufällig 150 Einwohner ausgewählt und deren IQ gemessen. Das Stichprobenmittel beträgt 97.2, die Stichprobenstandardabweichung 14.5. Wie muss ein Intervall gebildet werden, das den unbekannten Parameter der GG mit einer Sicherheit von 95% enthält?
- Das Konfidenzintervall wird in der Regel durch eine Unter- und Obergrenze gebildet. Gesucht ist also:

$$P(UG \leq \mu \leq OG) = 0.95 = 1 - \alpha$$

- α wird als Irrtumswahrscheinlichkeit bezeichnet, $1 - \alpha$ als Konfidenzintervall oder Sicherheit.
- Das Intervall liegt symmetrisch um das Stichprobenmittel.

Konfidenzintervall für unbekannten Mittelwert

- Ausgangspunkt: Verteilung des Stichprobenmittels für $n = 150$ und unbekannter Varianz der Grundgesamtheit.
- Es gilt:

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

- Es gilt weiterhin:

$$\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \sim t(n-1)$$

- Und schließlich gilt für $n \geq 150$ annähernd:

$$\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \sim N(0,1)$$

- D.h. es gilt approximativ:

$$P(-\lambda_{0.975} \leq \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \leq \lambda_{0.975}) = P(-1.96 \leq \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \leq 1.96) = 0.95$$

Interpretation des Konfidenzintervalls

- Für das (theoretische) 95%-Konfidenzintervall ergibt sich damit:

$$P(\overline{X}_n - 1.96 \cdot S_n/\sqrt{n} \leq \quad \mu_X \quad \leq \overline{X}_n + 1.96 \cdot S_n/\sqrt{n}) = 0.95$$

- WICHTIG: Die Intervallgrenzen sind Funktionen des Stichprobenmittelwerts und damit ebenfalls Zufallsvariablen
- Das realisierte Konfidenzintervall für die vorliegende Stichprobe lautet:

$$KI = [111.78; 114.82]$$

- Die Aussage, der unbekannte Parameter der GG μ_X fällt mit 95%er Wahrscheinlichkeit in das Konfidenzintervall, ist falsch.
- Korrekte Interpretation: Das aus der Stichprobe konstruierte KI überdeckt mit 95% WS den unbekannten Parameter der GG.
- Werden z.B. 1000 Stichproben gezogen und für jede Stichprobe das 95%-Konfidenzintervall berechnet, dann überdecken 95% also 950 der 1000 KI den unbekannten Parameter μ_X.

Noch mehr zum Konfidenzintervall

- Das Konfidenzintervall

$$P(\overline{X}_n - \lambda_{1-\alpha/2} \cdot S_n/\sqrt{n} \leq \mu_X \leq \overline{X}_n + \lambda_{1-\alpha/2} \cdot S_n/\sqrt{n}) = 1 - \alpha$$

ist umso enger, je

1. geringer die vorgegebene Sicherheit ist. Es gibt also einen Trade-off: Sicherheit gegen Intervallbreite.
2. geringer die Varianz in der Stichprobe ist
3. größer der Stichprobenumfang ist.

- Bei unbekannter Varianz der GG sind die möglichen Konfidenz- intervalle unterschiedlich lang. Sowohl die Lage $\bar{X}_n$ als auch die Länge S_n verändern sich von Stichprobe zu Stichprobe.

- KI für μ bei unbekannter Varianz der GG und $n < 150$.

$$P(\overline{X}_n - t_{n-1;1-\alpha/2} \cdot S_n/\sqrt{n} \leq \mu \leq \overline{X}_n + t_{n-1;1-\alpha/2} \cdot S_n/\sqrt{n}) = 1 - \alpha$$

Konfidenzintervall für den Anteilswert

- Für $nA_n(1 - A_n) > 9$ gilt:

$$A_n \sim N(p, p(1-p)/n)$$

- A_n ist ein erwartungstreuer Schätzer für p.
- Daraus folgt für das Konfidenzintervall:

$$P\left(A_n - \lambda_{1-\alpha/2}\frac{\sqrt{A_n(1-A_n)}}{\sqrt{n}} \leq p \leq A_n + \lambda_{1-\alpha/2}\frac{\sqrt{A_n(1-A_n)}}{\sqrt{n}}\right) = 1-\alpha$$

- Beispiel: Die Marketing-Abteilung eines Schokoladenherstellers möchte gern den Bekanntheitsgrad eines neuen Schokoriegels abschätzen. Dazu werden 70 Personen zufällig ausgewählt und telefonisch befragt. Der Schokoriegel war 50% der Befragten bekannt. Ermitteln und interpretieren Sie das 95%-Konfidenzintervall für den Bekanntheitsgrad p in der Grundgesamtheit.

Exaktes Konfidenzintervall für Anteilswerte in **R**

```
> binom.test(35,70)

        Exact binomial test

data:  35 and 70 number of successes = 35, number of trials = 70,
p-value = 1 alternative hypothesis: true probability of success is
not equal to 0.5 95 percent confidence interval:
 0.3780183 0.6219817
sample estimates: probability of success
                   0.5
```

Übergang vom Konfidenzintervall zum Hypothesentest

- Fortsetzung Beispiel IQ (S.96): Kann auf Basis des Stichprobenbefunds für $n = 150$ die Hypothese von $\mu_X = 100$ abgelehnt werden?
- Implizite Frage: Passt ein bestimmtes Stichprobenergebnis “gut” zu einer bestimmten Hypothese? $\Rightarrow$ Signifikanztest
- Logik des Signifikanztests am Beispiel: Wenn der IQ-Mittelwert μ_X in der Grundgesamtheit tatsächlich 100 beträgt, wie wahrscheinlich ist es dann in einer Stichprobe vom Umfang $n = 150$ ein Stichprobenmittel von 97.2 zu erhalten?
- Unsicherheit bleibt immer $\Rightarrow$ Stichprobe kann zufällig sehr extreme Realisationen aufweisen. $\Rightarrow$ WICHTIG: Unsicherheit wird quantifizierbar und damit “objektiv”.

Bestandteile eines Hypothesentests

1. Mögliche Hypothesen:
   - Nullhypothese H_0: $\mu_X = 100$
   - Alternativhypothese H_A: $\mu_X \neq 100$

2. Teststatistik z: Wenn H_0 stimmt ($\mu_X = 100$), wie wahrscheinlich ist es dann ein Stichprobenergebnis zu erhalten, das mindestens um 100-97.2=2.8 davon abweicht (egal ob nach oben oder unten)?

$$z = \frac{\overline{X}_n - \mu_0}{S_n/\sqrt{n}} = \frac{97.2 - 100}{14.5/\sqrt{150}} = -2.37$$

Bestandteile eines Hypothesentests

3. Irrtumswahrscheinlichkeit α: Welches Risiko bin ich bereit einzugehen, H_0 abzulehnen, obwohl diese korrekt ist (Fehler 1. Art).
   $\alpha = 0.01$ oder auch $\alpha = 0.05$

   Wichtig: Die Wahrscheinlichkeit β, H_0 nicht abzulehnen, obwohl H_0 falsch ist, kann nicht quantifiziert werden (Fehler 2. Art).
   Aber es gilt: Je kleiner α umso größer β.

4. Beibehaltungs- und Ablehnungsbereich (kritische Werte)
   - Beibehaltungsbereichs: Es wird ein Intervall gebildet, in dem sich der Wert der Teststatistik mit WS $1 - \alpha$ realisieren wird d.h. für $\alpha = 0.01$ das Intervall [-2.58;+2.58], für $\alpha = 0.05$ das Intervall [-1.96;+1.96]
   - Liegt z in diesem Intervall (d.h. passen Nullhypothese und Stichprobenergebnis zusammen), dann wird H_0 nicht abgelehnt.
   - Liegt z nicht im Intervall (d.h. passen Nullhypothese und Stichprobenergebnis nicht zusammen) wird H_0 abgelehnt.

Beispiel: Lebensdauer von Glühbirnen I

- Annahme bezüglich der Verteilung der Grundesamtheit: Lebensdauer in Tagen X von Glühbirnen sei normalverteilt mit unbekanntem Mittelwert μ und bekannter Varianz $\sigma^2 = 900$.

  Annahme bezüglich der Stichprobe: Aus der Produktion wird eine einfache Zufallsstichprobe des Umfangs $n = 16$ gezogen.
- Stichprobenergebnis: Mittlere Lebensdauer von $\overline{x}_{16} = 820$Tagen.
- Entscheidung zwischen zwei einfachen Hypothesen

$$H_0 : \ \mu = \mu_0 = 800 \quad \text{gegen} \quad H_1 : \ \mu = \mu_1 \ \text{ mit } \mu_1 > \mu_0.$$

- Verteilung der Prüfgröße unter H_0:

$$\overline{X}_{16} \sim \mathcal{N}(800, \frac{900}{16})$$

  bzw.

$$\frac{\sqrt{16}(\overline{X}_{16} - 800)}{\sqrt{900}} \sim \mathcal{N}(0,1).$$

- Vorgabe der Wahrscheinlichkeit für den Fehler erster Art: $\alpha = 0.05$.

Beispiel: Lebensdauer von Glühbirnen II

- Festlegung des kritischen Bereiches (der kritischen Schranke):

$$P\left(\frac{\sqrt{16}(\overline{X}_{16}-800)}{\sqrt{900}} > k|\mu = 800\right) = 1-\Phi(k) = \alpha = 0.05,$$

  so dass k das 95%-Quantil der Standardnormalverteilung ist, d.h. $k = 1.645$.

- Entscheidungsregel: Lehne die Nullhypothese mit 5%-Irtrtumswahrscheinlichkeit ab, wenn der Stichprobenbefund $(x_1, \ldots, x_{16})$ in den kritischen Bereich K fällt mit

$$K = \left\{(x_1, \ldots, x_n)|\frac{\sqrt{16}(\overline{X}_{16}-800)}{\sqrt{900}} > 1.645\right\}$$

- Testentscheidung: Setzt man den Stichprobenbefund von $\overline{x}_{16} = 820$ in die Prüfgröße ein, so ergibt sich als Stichprobenwert der Prüfgröße

$$\frac{\sqrt{16}(820-800)}{\sqrt{900}} = 2.667.$$

  Dieser Wert ist größer als die kritische Schranke $k = 1.645$, so dass die Stichprobe in den kritischen Bereich fällt und die Nullhypothese mit 5%-Irrtumswahrscheinlichkeit abzulehnen ist.

Beispiel: Lebensdauer von Glühbirnen III

- Die Testentscheidung hätte auch mit dem Stichprobenmittel direkt und nicht mit dem standardisierten Stichprobenmittel erfolgen können.

  Einfache Umrechnung ergibt

$$\frac{\sqrt{16}(\overline{X}_{16} - 800)}{\sqrt{900}} > k = 1.645 \quad \Longleftrightarrow \quad \overline{X}_{16} > k'$$

$$= 1.645 \frac{\sqrt{900}}{\sqrt{16}} + 800$$

$$= 812.3.$$

  Da das realisierte Stichprobenmittel $\overline{x}_{16} = 820$ größer ist als 812.3 ist die Nullhypothese mit mit 5%-Irrtumswahrscheinlichkeit abzulehnen.

Beispiel: Lebensdauer von Glühbirnen IV

- Testentscheidung mittels p-Wert
- Bislang wurde die Irrtumswahrscheinlichkeit α vorgegeben, die zugehörige kritische Schranke 1.645 errechnet und geschaut, ob der Stichprobenwert der Prüfgröße 2.667 diese Schranke "überschreitet".
- Man kann aber auch den Stichprobenwert der Prüfgröße zuerst berechnen und daraus Wahrscheinlichkeit ableiten, dass dieser Wert "überschritten" wird bei Gültigkeit der Nullhypothese:

$$P\left(\sqrt{n}\frac{\overline{X}_n - \mu_0}{\sigma} > 2.667\right) = 0.00383$$

- Ist dieser Wert kleiner als 0.05, so "überschreitet" der Stichprobenwert den kritischen Wert, und die Nullhypothese ist mit einer Irrtumswahrscheinlichkeit von 5% abzulehnen.

Spezielle Tests bei normalverteilten Grundgesamtheiten

- Mittelwerttest bei bekannten Varianzen
- Nittelwerttest bei unbekannten Varianzen
- Varianztest
- Mittelwertdifferenzentest für unverbundene Stichproben bei unbekannten und gleichen Varianzen
- Mittelwertdifferenzentest für verbundene Stichproben
- Varianzquotiententest für unverbundenen Stichproben

Mittelwerttest bei bekannter Varianz

- Prüfgröße:

$$T_n = \sqrt{n}\frac{\overline{X}_n - \mu_0}{\sigma} \overset{H_0}{\sim} \mathcal{N}(0,1)$$

- Kritische Bereiche für Hypothesenpaare:

Hypothesen	Kritischer Bereich
$H_0: \ \mu \geq \mu_0$ gegen $H_1: \ \mu < \mu_0$	$t_n < -\lambda_{1-\alpha}$
$H_0: \ \mu \leq \mu_0$ gegen $H_1: \ \mu > \mu_0$	$t_n > \lambda_{1-\alpha}$
$H_0: \ \mu = \mu_0$ gegen $H_1: \ \mu \neq \mu_0$	$\lvert t_n \rvert > \lambda_{1-\frac{\alpha}{2}}$

Mittelwerttest bei unbekannter Varianz

- Prüfgröße:

$$T_n = \sqrt{n}\frac{\overline{X}_n - \mu_0}{S_n} \overset{H_0}{\sim} t(n-1)$$

- Kritische Bereiche für Hypothesenpaare:

Hypothesen		Kritischer Bereich
$H_0: \ \mu \geq \mu_0$	gegen $H_1: \ \mu < \mu_0$	$t_n < -t_{1-\alpha;n-1}$
$H_0: \ \mu \leq \mu_0$	gegen $H_1: \ \mu > \mu_0$	$t_n > t_{1-\alpha;n-1}$
$H_0: \ \mu = \mu_0$	gegen $H_1: \ \mu \neq \mu_0$	$\lvert t_n \rvert > t_{1-\frac{\alpha}{2};n-1}$

Beispiel: Ist die Durchschnittsabiturnote signifikant schlechter als 2.3

```
> stat=read.table("c://stat.txt",header=TRUE)
      # Einlesen der Daten aus der Datei
> x1 =stat[1:946,2]
      # x1 enthaelt die Abinote
> t.test(x1,y=NULL,alternative="greater",mu=2.3,conf.level=0.95)}
      # Test, ob die Abinote
      # signifikant schlechter als 2.3 ist.

data:  stat[1:946, 2] t = 6.1329, df = 945, p-value = 6.335e-10
alternative hypothesis: true mean is greater than 2.3 95 percent
confidence interval:
 2.387459      Inf
sample estimates: mean of x
 2.419556
```

Ist die Durchschnittsabiturnote signifikant besser als 2.5?

```
> t.test(x1,y=NULL,alternative="less",mu=2.5,conf.level=0.95)
                                 # Test, ob die Abinote
                                 # signifikant besser als 2.5
                                 # ist.

        One Sample t-test

data:  stat[1:946, 2] t = -4.1265, df = 945, p-value = 2.004e-05
alternative hypothesis: true mean is less than 2.5 95 percent
confidence interval:
     -Inf 2.451653
sample estimates: mean of x
 2.419556
```

Varianztest bei unbekanntem Mittelwert

- Prüfgröße:

$$T_n = (n-1)\frac{S_n^2}{\sigma^2} \overset{H_0}{\sim} \chi^2(n-1)$$

- Kritische Bereiche für Hypothesenpaare:

Hypothesen	Kritischer Bereich
$H_0: \sigma^2 \geq \sigma_0^2$ gegen $H_1: \sigma^2 < \sigma_0^2$	$t_n < \chi^2_{\alpha;n-1}$
$H_0: \sigma^2 \leq \sigma_0^2$ gegen $H_1: \sigma^2 > \sigma_0^2$	$t_n > \chi^2_{1-\alpha;n-1}$
$H_0: \sigma^2 = \sigma_0^2$ gegen $H_1: \sigma^2 \neq \sigma_0^2$	$t_n < \chi^2_{\alpha/2;n-1}$; $t_n > \chi^2_{1-\alpha/2;n-1}$

Mittelwertdifferenzentest für unverbundene Stichproben bei unbekannten, aber identischen Varianzen

- Ausgangssituation: Seien $X_1, \ldots, X_n$ und $Y_1, \ldots, Y_m$ stochastisch unabhängige Zufallsstichproben aus zwei normalverteilten Grundgesamtheiten mit den Parametern μ_X, μ_Y, σ_X^2 und σ_Y^2, wobei die Mittelwerte und Varianzen unbekannt sein sollen.
- Prüfgröße:

$$T_n = \frac{\overline{X}_n - \overline{Y}_m - \delta_0}{\sqrt{\overline{S}^2(1/n + 1/m)}} \overset{H_0}{\sim} t(n+m-2)$$

- Kritische Bereiche für Hypothesenpaare:

Hypothesen $(\delta = \mu_X - \mu_Y)$		Kritischer Bereich
$H_0:\ \delta \geq \delta_0$	gegen $H_1:\ \delta < \delta_0$	$t_n < -t_{1-\alpha;n+m-2}$
$H_0:\ \delta \leq \delta_0$	gegen $H_1:\ \delta > \delta_0$	$t_n > t_{1-\alpha;n+m-2}$
$H_0:\ \delta = \delta_0$	gegen $H_1:\ \delta \neq \delta_0$	$\lvert t_n \rvert > t_{1-\frac{\alpha}{2};n+m-2}$

Beispiel zum Zweistichproben-t-Test I

- Stichprobe von $n = 16$ Glühbirnen mit $\overline{x}_{16} = 826$ als mittlere beobachtete Brenndauer X und $s^2_{X;16} = 870$.
- Zweite Stichprobe von $m = 20$ Glühbirnen, die mit einem anderen Produktionsverfahren hergestellt wurden, liefert $\overline{y}_{20} = 850$ als Mittelwert der Brenndauer Y und $s^2_{Y;20} = 890$.
- Es wird unterstellt, daß die Varianz der Brenndauer beider Glühbirnenarten gleich ist.
- Konstruieren Sie einen Test zum Niveau 10%, anhand dessen überprüft werden kann, ob die beiden Produktionsverfahren bezüglich der mittleren Brenndauer der Glühbirnen als ungleich eingestuft werden können.

Beispiel zum Zweistichproben-t-Test II

- Zu testendes Hypothesenpaar:

$$H_0: \ \delta = \mu_X - \mu_Y = \delta_0 = 0 \ \text{ gegen } \ H_1: \ \delta \neq \delta_0 = 0.$$

- Kritischer Bereich für 10%-Niveau:

$$\left| \frac{\overline{x}_{16} - \overline{y}_{20} - 0}{\sqrt{\overline{s}^2(1/16 + 1/20)}} \right| > t_{0.95;34} = 1.691.$$

- Testentscheidung: Da

$$\left| \frac{\overline{x}_{16} - \overline{y}_{20} - 0}{\sqrt{\overline{s}^2(1/16 + 1/20)}} \right| = \left| \frac{826 - 850}{\sqrt{881.177(1/16 + 1/20)}} \right| = |-2.410| > t_{0.95;34} = 1.691$$

ist, kann die Nullhypothese, dass beide Maschinen Glühbirnen mit identischer mittlerer Brennauer produzieren, mit einer Irrtumswahrscheinlichkeit von 10% abgelehnt werden.

Zwei-Stichproben-t-Test für unverbundene Stichproben in **R** mittels der Standardroutine **t.test** I

Unterscheiden sich die durchschnittlichen Abfüllmengen der beiden Zuckerabfüllmaschinen signifikant?

```
> x              # Vorhanden oder Einlesen
[1] 1005.04  992.97 1059.97  978.70 1023.75  997.75 1012.61 986.76
1014.55          # Stichprobe fuer aeltere Zuckerabfuellmaschine

> y              # Vorhanden oder Einlesen
 [1] 1014.96 1081.07 1049.75 1044.74 1058.31 1092.23 1092.57
                                                 1055.19 1075.13
[10] 1077.24 1088.48 1055.22 1054.65 1053.57 1026.25 1078.17
                 # Stichprobe fuer neue Zuckerabfuellmaschine

>t.test(x,y,alternative="two.sided",mu=0,paired=FALSE,
                                 var.equal=TRUE,conf.level=0.95)
                 # Zweiseitiger t-Test fuer unverbundene
                 # (paired=FALSE) Stichproben bei identischen
                 # (var.equal=TRUE) Varianzen
```

Zwei-Stichproben-t-Test für unverbundene Stichproben in R mittels der Standardroutine **t.test** II

Output:

```
        Two Sample t-test

data:  x and y t = -5.614, df = 23, p-value = 1.031e-05 alternative
hypothesis: true difference in means is not equal to 0 95 percent
confidence interval:
 -74.35593 -34.31310
sample estimates: mean of x mean of y
 1008.011  1062.346
```

Mittelwertdifferenzentest bei verbundenen Stichproben

- Ausgangspunkt: Seien $(X_1, Y_1), \ldots, (X_n, Y_n)$ stochastisch unabhängig, identisch bivariat normalverteilt mit Mittelwerte μ_X bzw. μ_Y und unbekannten Varianzen σ_X^2 bzw. σ_Y^2. Es werden die Differenzen $D_i = X_i = Y_i$, $i = 1, 2, \ldots, n$ betrachtet.
- Prüfgröße:

$$T_n = \sqrt{n}\frac{\overline{X}_n - \overline{Y}_n - \delta_0}{S_{D;n}} \overset{H_0}{\sim} t(n-1)$$

- Kritische Bereiche für alternative Hypothesenpaare:

Hypothesen $(\delta = \mu_X - \mu_Y)$	Kritischer Bereich
$H_0: \delta \geq \delta_0$ gegen $H_1: \delta < \delta_0$	$t_n < -t_{1-\alpha;n-1}$
$H_0: \delta \leq \delta_0$ gegen $H_1: \delta > \delta_0$	$t_n > t_{1-\alpha;n-1}$
$H_0: \delta = \delta_0$ gegen $H_1: \delta \neq \delta_0$	$\lvert t_n \rvert > t_{1-\alpha/2;n-1}$

Zwei-Stichproben-t-Test für verbundene Stichproben in **R** mittels der Standardroutine **t.test** I

Aufgabe: Ist die Mathenote signifikant schlechter als die Abinote?

```
> x=read.table("c://stat.txt",header=TRUE)
                                 # Einlesen der Daten aus der Datei
> x2 =x[1:946,4]                 # x1 enthaelt die Mathenote
> x1 =x[1:946,2]                 # x2 enthaelt die Abinote

> t.test(x1,x2,alternative="less",mu=0,paired=TRUE,
                                      var.equal=FALSE,conf.level=0.99)
                                   # Rechtsseitiger t-Test fuer
                                   # verbundene (paired) Stichproben
        Paired t-test

data:  x1 and x2 t = -23.7736, df = 945, p-value < 2.2e-16
alternative hypothesis: true difference in means is less than 0 99
percent confidence interval:
       -Inf -0.6391563
sample estimates: mean of the differences
             -0.7086152
```

Zwei-Stichproben-t-Test für verbundene Stichproben in **R** mittels der Standardroutine **t.test** II

Aufgabe: Unterscheiden sich Mathe- und Abinote signifikant ?

```
> t.test(x1,x2,alternative="two.sided",mu=0,paired=TRUE,
                              var.equal=FALSE,conf.level=0.99)
                                 # Zweiseitiger t-Test fuer
                                 # verbundene (paired) Stichproben
        Paired t-test

data:  x1 and x2 t = -23.7736, df = 945, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 -0.7855480 -0.6316824
sample estimates: mean of the differences
             -0.7086152
```

Varianzenquotiententest bei unverbundenen Stichproben

- Ausgangssituation: Seien $X_1, \ldots, X_n$ und $Y_1, \ldots, Y_m$ stochastisch unabhängige Zufallsstichproben aus zwei normalverteilten Grundgesamtheiten mit den Parametern μ_X, μ_Y, σ_X^2 und σ_Y^2, wobei die Mittelwerte und Varianzen unbekannt sein sollen.
- Prüfgröße:

$$\frac{S_n^2/\sigma_X^2}{S_m^2/\sigma_Y^2} \overset{H_0}{\sim} F(n-1, m-1).$$

- Kritische Bereiche für alternative Hypothesenpaare:

Hypothesen	Kritischer Bereich
$H_0: \sigma_X \geq \sigma_Y$ gegen $H_1: \sigma_X < \sigma_Y$	$S_n^2/S_m^2 < f_{\alpha;n-1,m-1}$
$H_0: \sigma_X \leq \sigma_Y$ gegen $H_1: \sigma_X > \sigma_Y$	$S_n^2/S_m^2 > f_{1-\alpha;n-1,m-1}$
$H_0: \sigma_X = \sigma_Y$ gegen $H_1: \sigma_X \neq \sigma_Y$	$S_n^2/S_m^2 < f_{\alpha;n-1,m-1}$
	oder
	$S_n^2/S_m^2 > f_{1-\alpha;n-1,m-1}$

Zwei-Stichproben-F-Test für unverbundene Stichproben in **R** mittels der Standardroutine **var.test**

Aufgabe: Produziert die neue Maschine signifikant genauer als die alte Maschine?

```
> x              # x bereits im Workspace; wenn nicht: Einlesen.
[1] 1005.04  992.97 1059.97  978.70 1023.75  997.75 1012.61 986.76
1014.55          # Stichprobe fuer aeltere Zuckerabfuellmaschine

> y              # y bereits im Workspace; wenn nicht: Einlesen.
 [1] 1014.96 1081.07 1049.75 1044.74 1058.31 1092.23 1092.57
                                                     1055.19 1075.13
[10] 1077.24 1088.48 1055.22 1054.65 1053.57 1026.25 1078.17
                 # Stichprobe fuer neue Zuckerabfuellmaschine

> var.test(x,y,ratio=1,alternative="greater",conf.level=0.95)
                 # Test der Hypothese, dass neuere Maschine
                 # praeziser
                 # produziert als die neuere Maschine

        F test to compare two variances

data:  x and y F = 1.1284, num df = 8, denom df = 15, p-value =
0.3992 alternative hypothesis: true ratio of variances is greater
than 1 95 percent confidence interval:
 0.4272979       Inf
sample estimates: ratio of variances
          1.128407
```

Asymptotische Tests

- Mittelwerttests:
  Wenn keine Normalverteilung unterstellt wird, Mittelwerte und Varianzen der Grundgesamtheit existieren, dann können aufgrund des zentralen Grenzwertsatzes für große Stichprobenumfänge die o.g. Mittelwerttests ebenfalls angewendet werden, wobei für $30 \leq n \leq 100$ die Quantile der t-Verteilung für die Bestimmung der kritischen Werte und für noch größere Stichprobenumfänge die Quantile der Standardnormalverteilung verwendet werden.
- Mittelwertdifferenzentest:
  Der vorgeschlagene Mittelwertdifferenzentest unverbundener Stichproben kann auch angewendet werden, wenn $n, m > 20$ sind, ohne dass Normalverteilung unterstellt werden muß.
- Varianztests:
  Varianzentests hängen äußerst sensibel von der Normalverteilungsannahme ab und sollten nur angewendet werden, wenn man sich dieser sicher ist.

Approximativer Anteilswerttest

- Prüfgröße:

$$T_n = \sqrt{n}\frac{\overline{X}_n - p_0}{\sqrt{p_0(1-p_0)}} \overset{appr}{\sim} N(0,1) \quad \text{(unter } H_0\text{)}.$$

- Faustregel für eine hinreichend gute Approximation:

$$np_0(1-p_0) > 9.$$

- "Approximative" kritische Bereiche für alternative Hypothesenpaare:

Hypothesen		Kritischer Bereich
$H_0 :\ p \geq p_0$	gegen $H_1 :\ p < p_0$	$t_n < -\lambda_{1-\alpha}$
$H_0 :\ p \leq p_0$	gegen $H_1 :\ p > p_0$	$t_n > \lambda_{1-\alpha}$
$H_0 :\ p = p_0$	gegen $H_1 :\ p \neq p_0$	$\|t_n\| > \lambda_{1-\alpha/2}$

Anteilswertetest in R I

Aufgabe: Es werden 2000 zufällig ausgewählte Personen gefragt, ob sie am nächsten Sonntag CDU/CSU wählen würden, wenn Wahl wäre. 840 Personen haben mit "Ja" geantwortet.
Zu testen ist die Hypothese, dass der CDU/CSU-Anteil in der Grundgesamtheit nicht 40% ist.
Exakter Binomialtest:

```
> binom.test(840,2000,alternative="two.sided",0.4)

        Exact binomial test

data:  840 and 2000 number of successes = 840, number of trials =
2000, p-value = 0.07137 alternative hypothesis: true probability of
success is not equal to 0.4 95 percent confidence interval:
 0.3982500 0.4419849
sample estimates: probability of success
                  0.42
```

Anteilswertetest in **R** II

Approximativer Anteilswerttest:

```
> x=c(rep(1,840),rep(0,1160))  # x ist Vektor mit 840 Einsen und
                               # 1160 Nullen

> t.test(x,y=NULL,mu=0.4,alternative="two.sided")

        One Sample t-test

data: x t = 1.8117, df = 1999, p-value = 0.07018 alternative
hypothesis: true mean is not equal to 0.4 95 percent confidence
interval:
 0.3983507 0.4416493
sample estimates: mean of x
     0.42
```

Anteilswertetest in **R** III

Zu testen ist die Hypothese, dass der CDU/CSU-Anteil in der Grundgesamtheit größer als 40% ist.
Exakter Binomialtest:

```
> binom.test(840,2000,alternative="greater",0.4)

        Exact binomial test

data:  840 and 2000 number of successes = 840, number of trials =
2000, p-value = 0.03595 alternative hypothesis: true probability of
success is greater than 0.4 95 percent confidence interval:
 0.4016924 1.0000000
sample estimates: probability of success
                  0.42
```

Anteilswertetest in **R** IV

Approximativer Anteilswerttest:

```
> t.test(x,y=NULL,mu=0.4,alternative="greater")

        One Sample t-test

data:  x t = 1.8117, df = 1999, p-value = 0.03509 alternative
hypothesis: true mean is greater than 0.4 95 percent confidence
interval:
 0.4018339       Inf
sample estimates: mean of x
     0.42
```