Mitschrieb zur Vorlesung

Stochastik für Informatiker und Bioinformatiker

Markus List

Sommersemester 2007

Inhaltsverzeichnis

Da dieser Aufschrieb während der Vorlesung entsteht, schleichen sich leicht Fehler ein. Wenn ihr welche findet dürft ihr mich gerne darauf hinweisen: Einfach eine kurze E-Mail an: markus.list@student.uni-tuebingen.de

A. Wahrscheinlichkeitstheorie

A.1. Wahrscheinlichkeit

Naiver Zugang: Man unterscheidet bei den möglichen Ausgängen eines Experiments zwischen günstigen und ungünstigen Fällen und interpretiert die Zahl

$$W := \frac{\text{Anzahl günstiger Fälle}}{\text{Anzahl aller Fälle}} \tag{1}$$

als die Wahrscheinlichkeit (kurz W.keit oder W.) für das Ereignis, dass das Experiment günstig ausgeht. Dabei wird angenommen, dass alle möglichen Fälle „gleich wahrscheinlich" eintreten.

Bsp.: Idealer unverfälschter Würfel
Jede der Zahlen von 1 bis 6 wird gleich wahrscheinlich eintreten $\left(W = \frac{1}{6}\right)$
Nimmt man z.B. die geraden Zahlen als günstige Fälle, erhält man $\left(W = \frac{3}{6} = \frac{1}{2}\right)$

Realer Würfel:

1. Man glaubt der Würfel sei ideal
   Problem: Es stimmt vielleicht nicht (z.B. gezinkter Würfel)

2. Man würfelt z.B. 1000 mal und bestimmt die W.keiten via Formel (1)
   Problem: Das Resultat hängt vom konkreten Ausgang des Würfelns ab
   (z.B. tausend mal die 6)

A.2. Laplacesche Wahrscheinlichkeitsräume

Gegeben seien eine natürliche Zahl $n \in \mathbb{N}$ und ein Grundraum $\Omega = \{\omega_1, \ldots, \omega_n\}$ von Elementarereignissen ω_i.

Die Potenzmenge $\mathcal{P}(\Omega)$ von Ω, d.h. die Menge („Sammlung", „Familie") der Teilmengen von Ω, d.h. die Mengen $A \subset \Omega$ heißen Ereignisse in Ω.
Es gibt also 2^n Ereignisse in Ω.

Wir definieren nun eine Abb. $P : \mathcal{P}(\Omega) \to [0,1]$, („probability") wie folgt:

- Für Elementarereignisse $P(\{\omega_i\}) := \frac{1}{n}$
- Für Ereignisse $P(A) := \frac{\#\text{A}}{\text{n}} := \frac{\text{Anzahl Elementarereignisse in A}}{\text{n}}$

Interpretation: Man betrachtet die ω_i im Ereignis (in der Teilmenge) $A \subset \Omega$ als die günstigen Fälle.

Insbesondere gilt:

- $P(\emptyset) = 0$
- $P(\Omega) = 1$

Das Tupel (Ω, P) heißt Laplacescher Wahrscheinlichkeitsraum mit n Elementen.

Begriffe aus der Mengenlehre:

Vereinigung: $A \cup B := \{\omega \in \Omega : \omega \in A \text{ oder } \omega \in B\}$
„es tritt entweder A oder auch B ein"

$$\bigcup_{n=1}^{\infty} A_n := A_1 \cup A_2 \cup A_3 \cup \ldots$$

Schnitt: $A \cap B := \{\omega \in A \text{ und } \omega \in B\}$
„es tritt sowohl A wie auch B ein"

$$\bigcap_{n=1}^{\infty} A_n := A_1 \cap A_2 \cap A_3 \cap \ldots$$

Ähnlich: $k \in \mathbb{N}$

$$\bigcup_{n=1}^{k} A_n := A_1 \cup A_2 \cup \ldots \cup A_k$$

$$\bigcap_{n=1}^{k} A_n := A_1 \cap A_2 \cap \ldots \cap A_k$$

Weiter: Seien $A, B \subset \Omega$ zwei Ereignisse.

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

Insbesondere: $P(A \cap B) = 0$
$\Rightarrow P(A \cup B) = P(A) + P(B)$

Also auch: $A \cap B = \emptyset$, d.h. A und B disjunkt.
$\Rightarrow P(A \cup B) = P(A) + P(B)$

Grund: Venn-Diagramm $\frac{\#(A \cup B)}{n} = \frac{\#A}{n} + \frac{\#B}{n} - \frac{\#(A \cap B)}{n}$

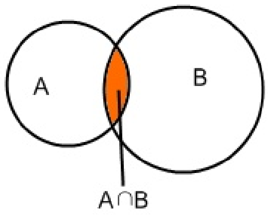


Abbildung 1: Venn-Diagramm

Bsp.:

- Münzwurf, $n = 2$, $\Omega = \{K, Z\}$ oder auch $\Omega = \{0, 1\}$
  Ereignisse in $\Omega = \{K, Z\} : \emptyset, \{K\}, \{Z\}, \{K, Z\}$

  W.keiten: $P(\emptyset) = 0, P(\{K, Z\}) = \frac{2}{2} = 1$
  $P(\{K\}) = P(\{Z\}) = \frac{1}{2}$
- Würfel, $n = 6$, $\Omega = \{1, 2, 3, 4, 5, 6\}$
  Anzahl der Ereignisse $= 2^6 = 64$
- Skat-Kartenspiel $n = 32$
- Roulette $n = 37$

Bsp.: Zweimaliges Würfeln

$\Omega = \{1, \ldots, 6\} \times \{1, \ldots, 6\}$ Kartesisches Produkt
$= \{1, \ldots, 6\}^2$
$= \{(1, 1), \ldots, (1, 6), (2, 1), \ldots, (2, 6), \ldots, (6, 1), \ldots, (6, 6)\}$

Anzahl Elementarereignisse $n = 36$

Sei A das Ereignis, dass eine 6 und eine weitere gerade Zahl gewürfelt werden.

$A := \{(6, 2), (6, 4), (6, 6), (2, 6), (4, 6)\}$

Dann gilt $P(A) = \frac{5}{36}$

Sei B das Ereignis, dass mind. eine 6 geworfen wird, also

$B := \{(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6), (1, 6), (2, 6), (3, 6), (4, 6), (5, 6)\}$

Dann gilt $P(B) = \frac{11}{36}$

Sei C das Ereignis, dass zwei gerade Zahlen geworfen werden, also

$C := \{(2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\}$

Dann gilt $P(B) = \frac{9}{36} = \frac{1}{4}$

Es gilt weiter: $A = B \cap C$

Wir erhalten also P („mind. eine 6 oder auch 2 gerade Zahlen“)

$$= P(B \cup C) = P(B) + P(C) - \underbrace{P(B \cap C)}_{P(A)} = \frac{11}{36} + \frac{9}{36} - \frac{5}{36} = \frac{5}{12}$$

A.3. Axiomatischer Zugang nach A.N.Kolmogoroff

Gegeben sei eine Menge $\Omega \neq \emptyset$.Man nennt Ω den Grundraum oder Ereignisraum. Die Elemente von Ω heißen Elementarereignisse. Gegeben sei weiter eine Familie $\mathcal{F} \subset \mathcal{P}(\Omega)$ von Teilmengen von Ω mit folgenden Eigenschaften:

1. Es gilt $\Omega \in \mathcal{F}$
2. Mit $A \in \mathcal{F}$ ist auch das Komplement
   $A^c = \{\omega \in \Omega : \omega \notin A\} \in \mathcal{F}$
3. gilt $A_1, A_2, \ldots \in \mathcal{F}$, so gelten auch
   $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{F}$ und $\bigcap\limits_{n=1}^{\infty} A_n \in \mathcal{F}$

Einfachstes Bsp.: $\mathcal{F} = \mathcal{P}(\Omega)$

Bemerkung: Sind 1., 2. und 3. erfüllt, so gilt auch folgende Aussage:

3.' gilt $A_1, \ldots, A_k \in \mathcal{F}$, so gelten auch
$\bigcup\limits_{n=1}^{k} A_n \in \mathcal{F}$ und $\bigcap\limits_{n=1}^{k} A_n \in \mathcal{F}$

Grund:

- Für $A_1, \ldots, A_k \in \mathcal{F}$ und $A_{k+1} := \Omega$, $A_{k+2} := \Omega, \ldots$ gelten auch $A_1, A_2, \ldots \overset{\text{nach 1}}{\in} \mathcal{F}$ und

$$\bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{k} A_n$$

Also gilt auch $\bigcap\limits_{n=1}^{k} A_n \in \mathcal{F}$ nach 3.

- Nach 1. und 2. gilt $\emptyset \in \mathcal{F}$, da $\Omega^c = \emptyset$
  Für $A_1, \ldots, A_k \in \mathcal{F}$ und $A_{k+1} := \emptyset, A_{k+2} := \emptyset, \ldots$ gelten auch $A_1, A_2, \ldots \in \mathcal{F}$ und

  $$\bigcup_{n=1}^{k} A_n = \bigcup_{n=1}^{\infty} A_n$$

  Also gilt auch $\bigcup\limits_{n=1}^{k} A_n \in \mathcal{F}$ nach 3.

Gegeben sei zusätzliche eine Abb. $P : \mathcal{F} \to [0,1]$ mit folgenden Eigenschaften:

i.) $P(\emptyset) = 0$ und $P(\Omega) = 1$

ii.) Seien $A_1, A_2, \ldots \in \mathcal{F}$ paarweise disjunkt, d.h. $A_i \cap A_j = \emptyset \; \forall \; i, j \in \mathbb{N}$ und $i \neq j$,

so gilt $P\left(\bigcup\limits_{n=1}^{\infty} A_n\right) = \sum\limits_{n=1}^{\infty} P(A_n) := P(A_1) + P(A_2) + \ldots$

Bemerkung: Sind i.) und ii.) erfüllt, so gilt auch folgende Aussage:

ii'.) Sind $A_1, \ldots, A_k \in \mathcal{F}$ paarweise disjunkt, d.h.

$A_i \cap A_j = \emptyset \; \forall \; i, j \in \{1, \ldots, k\}$ mit $i \neq j$, so gilt

$$P\left(\bigcup_{n=1}^{k} A_n\right) = \sum_{n=1}^{k} P(A_n) = P(A_1) + \ldots + P(A_k)$$

Grund: Sind $A_1, \ldots, A_k \in \mathcal{F}$ und $A_{k+1} := \emptyset, A_{k+2} := \emptyset, \ldots$
so gelten $A_1, A_2, \ldots \in \mathcal{F}$ paarweise disjunkt und $\bigcup\limits_{n=1}^{k} A_n = \bigcup\limits_{n=1}^{\infty} A_k$,
und somit

$$P\left(\bigcup_{n=1}^{k} A_n\right) = P\left(\bigcup_{n=1}^{\infty} A_n\right) \overset{ii}{=} \sum_{n=1}^{\infty} P(A_n) \overset{i}{=} \sum_{n=1}^{k} P(A_n)$$

Ein solches Tripel (3-Tupel) $(\Omega, \mathcal{F}, P)$ wird Wahrscheinlichkeitsraum genannt.

Für $A, B \in \mathcal{F}$ gilt z.B. Folgendes:

- $A \cup B = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B)$ paarweise disjunkt

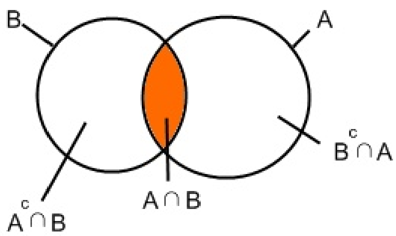


Abbildung 2: Venn-Diagramm

- $A = (A \cap B) \cup (A \cap B^c)$ disjunkt
- $B = (A \cap B) \cup (A^c \cap B)$ disjunkt

Nach 3' sind alle diese Mengen Ereignisse und nach ii.) gilt:

$$\begin{aligned} P(A \cup B) &= P(A \cap B) + P(A \cap B^c) + P(A^c \cap B) \\ &= P(A \cap B) + P(A \cap B^c) + P(A \cap B) + P(A^c \cap B) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

Abzählbar unendliche W.räume

Gegeben seien ein Grundraum $\Omega = \{\omega_1, \omega_2, \ldots\}$ oder $\Omega = \{\omega_0, \omega_1, \ldots\}$ und W.keiten $P(\{\omega_i\}) \in [0, 1]$ mit der Eigenschaft

$\sum\limits_{i=1}^{\infty} P(\{\omega_i\}) = 1$, bzw. $\sum\limits_{i=0}^{\infty} P(\{\omega_i\}) = 1$

Für $A \subset \Omega$ setzt man $P(A) := \sum\limits_{\omega \in A} P(\{\omega\})$

D.h. $P(A)$ ist die Summe der W.keiten der Elementarereignisse in A.

Dann ist $(\Omega, \mathcal{P}(\Omega), P)$ ein W.raum.

Wir nennen (Ω, P) einen abzählbar unendlichen W.raum

Endliche W.räume

Gegeben seien ein Grundraum $\Omega = \{\omega_1, \dots \omega_n\}$ oder $\Omega = \{\omega_0, \dots, \omega_n\}$ und W.keiten $P(\{\omega_i\})$ mit $\sum_{i=1}^{n} P(\{\omega_i\}) = 1$, bzw $\sum_{i=0}^{n} P(\{\omega_i\}) = 1$

Für $A \subset \Omega$ setzt man $P(A) = \sum_{\omega \in A} P(\{\omega\})$

Dann ist $(\Omega, \mathcal{P}(\Omega), P)$ ein W.raum.

Wir nennen (Ω, P) einen endlichen W.raum

Bsp.: Ein Laplacescher W.raum mit n Elementen ist ein endlicher W.raum mit $P(\{\omega_i\}) = \frac{1}{n}$ für jedes Elementarereignis ω_i.

A.4. Bedingte W.keiten

Wir betrachten einen Laplaceschen W.raum (Ω, P) mit n Elementen und $A, B \subset \Omega$

$P(A) = \frac{\#A}{n}$ beschreibt die W.keit dafür, dass ein beliebiges Elementarereignis von Ω aus A ist.

Wie groß ist die W.keit dafür, dass ein solches Elementarereignis (wenn es bereits aus A ist), dann auch aus B ist?

Naiver Ansatz:
Mögliche Fälle A
Günstige Fälle $A \cap B$

Also: $W = \frac{\#(A\cap B)}{\#A} = \frac{\frac{\#(A\cap B)}{n}}{\frac{\#A}{n}} = \frac{P(A\cap B)}{P(A)}$

Dies lässt sich auf allg. W.räume erweitern:

Seien $(\Omega, \mathcal{F}, P)$ ein W.raum, $A, B \in \mathcal{F}$ und $P(A) > 0$
Dann definiert man die bedingte W.keit von B unter A als

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

Für $P(A) = 0$ setzt man $P(B|A) := 0$

Analog: $P(A|B) := \frac{P(A \cap B)}{P(B)} \quad , P(B) > 0$
$P(A|B) := 0 \quad , P(B) = 0$

Multiplikationssatz:

$$\begin{aligned} P(A \cap B) &= P(B) \cdot P(A|B) \\ &= P(A) \cdot P(B|A) \end{aligned}$$

Bayessche Regel: Es sei $P(B) > 0$

Dann gilt: $P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)}$

Bsp.: Würfeln $A :=$ „Augenzahl ≥ 3“, $B :=$ „Augenzahl ungerade“

$P(A) = \frac{4}{6} = \frac{2}{3}, P(B) = \frac{3}{6} = \frac{1}{2} \Rightarrow P(B|A) = \frac{2}{4} = \frac{1}{2}$

$P(A|B) = \frac{\frac{4}{6} \cdot \frac{2}{4}}{\frac{3}{6}} = \frac{2}{3}$

Satz von der totalen W.keit

$P(B) = P(A) \cdot P(B|A) + P(A^c) \cdot P(B|A^c)$

Allgemeiner: Es seien A_n endlich oder abzählbar viele paarweise disjunkte Ereignisse mit $\bigcup\limits_n A_n = \Omega$.
Dann gilt: $P(B) = \underbrace{\sum_n P(A_n)}_{\text{endliche oder unendliche Summe}} \cdot P(B|A_n)$

A.5. Unabhängigkeit

Gegeben W.raum $(\Omega, \mathcal{F}, P)$

Definition:
Zwei Ereignisse $A, B \in \mathcal{F}$ heißen <u>unabhängig</u>, wenn $P(A \cap B) = P(A) \cdot P(B)$ gilt.

Interpretation:
Für $P(A) > 0$ und A, B unabh. gilt $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(B) \cdot P(A)}{P(A)} = P(B)$

d.h. die W.keit von B hängt nicht davon ab, ob A eingetroffen ist.

Analog: Für $P(B) > 0$, A, B unabhängig gilt $P(A|B) = P(A)$

Bsp.: Zweimaliges Ziehen einer Skatkarte

Seien A = „As beim 1. Zug“
B = „As beim 2. Zug“

- mit zurücklegen:

  $P(A) = P(B) = \frac{4}{32}, P(A \cap B) = \frac{4 \cdot 4}{32^2}$

  Also: A, B sind unabhängig

- ohne zurücklegen:

  $P(A) = \frac{4}{32} = P(B), P(A \cap B) = \frac{4 \cdot 3}{32 \cdot 31}$

  Also: A, B sind nicht unabhängig.

A.6. Zufallsvariablen

Gegeben sei ein endlicher oder abzählbar unendlicher W.raum (Ω, P). Wir nennen eine Abbildung $X : \Omega \to \mathbb{R}$ eine Zufallsvariable (kurz ZV) auf (Ω, P).

Notation: Für $a \in \mathbb{R}$ schreiben wir $\{X = a\} = \{\omega \in \Omega : X(\omega) = a\}$.

Es bezeichnet also

$P(X = a) := P(\{X = a\})$ die W.keit, dass X der Wert $a \in \mathbb{R}$ zugewiesen wird.

Bsp.: Würfeln $\Omega = \{1, \dots, 6\}$

$P(A) = \frac{\#A}{n}$ für $A \subset \Omega$

$X(\omega) := \omega$ für $\omega \in \{1, \dots, 6\}$ die „Augenzahl“

$P(X = \omega) = \frac{1}{6}$ für jedes $\omega \in \{1, \dots, 6\}$

Bsp.: Zweimaliges Würfeln

$$\begin{aligned}\Omega &= \{1, \dots, 6\} \times \{1, \dots, 6\} \\ &= \{(1,1), \dots, (1,6) \dots (6,1), \dots, (6,6)\}\end{aligned}$$

$P(A) = \frac{\#A}{36}, A \subset \Omega$

$X((\omega_1, \omega_2)) := \omega_1$ Augenzahl des 1.Wurfes
$Y((\omega_1, \omega_2)) := \omega_2$ Augenzahl des 2.Wurfes

Wir setzen $Z := X + Y$, also $Z((\omega_1, \omega_2)) := \omega_1 + \omega_2$ „die Summe der Augenzahlen“
Es gilt z.B.

$$P(Z = 2) = \frac{1}{36}$$
$$P(Z = 3) = \frac{2}{36}$$
$$\dots$$
$$P(Z = 7) = \frac{6}{36}$$

Gegeben sei ein beliebiger W.raum $(\Omega, \mathcal{F}, P)$
Eine Abb. $X : \Omega \to \mathbb{R}$ ist eine Zufallsvariable,
wenn sie folgende Eigenschaft erfüllt:

Für jede Wahl $a \leq b$ gilt $\{a \leq X \leq b\} := \{\omega \in \Omega : a \leq X(\omega) \text{ und } X(\omega) \leq b\} \in \mathcal{F}$

Man sagt, X ist messbar bezgl. $\mathcal{F}$ (vgl. „Masstheorie“ in der Analysis)

Die Gesamtheit der W.keiten $P(a \leq X \leq b)$ nennt man die Verteilung von X unter P.
Die Verteilung von X unter P ist bereits festgelegt durch die Funktion

$$F(t) := P(X \leq t) := P(\{X \leq t\}), t \in \mathbb{R}$$

wobei $\{X \leq t\} = \{\omega \in \Omega : X(\omega) \leq t\}$, die sogenannte Verteilungsfunktion von X unter P ist. F ist eine monoton wachsende, rechtsseitig stetige Funktion mit Werten im Intervall $[0, 1]$.

Bsp.: Würfeln

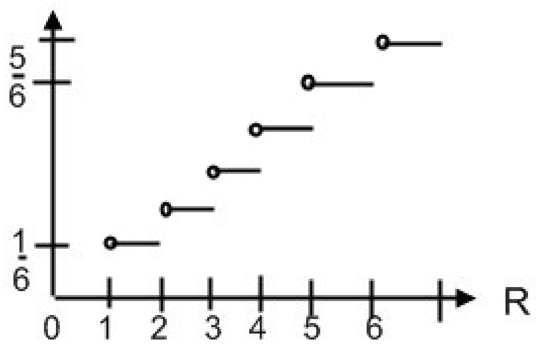


Abbildung 3: Verteilungsfunktion beim Würfeln

Definition: Zwei ZVen X, Y auf einem W.raum $(\Omega, \mathcal{F}, P)$ heißen unabhängig, wenn

$P(X \leq a, Y \leq b) := P(\{X \leq a\} \cap \{Y \leq b\})$

$= P(X \leq a) \cdot P(Y \leq b)$ gilt $\forall\ a, b \in \mathbb{R}$.

Bsp.: zweimaliges Würfeln

$\Omega = \{1, \ldots, 6\} \times \{1, \ldots, 6\}$

$P(A) = \frac{\#A}{36}$

X: Augenzahl 1. Wurf
Y: Augenzahl 2. Wurf

Betrachten $a, b \in \{1, \ldots, 6\}$ (die anderen Fälle folgen einfach)

$\{X \leq a\} \cap \{Y \leq b\} = \{(1,1), \ldots, (1,b), \ldots, (a,1), \ldots, (a,b)\}$

Also: $P(X \leq a, Y \leq b) = \frac{a \cdot b}{36} = \frac{a}{6} \cdot \frac{b}{6}$

$= P(X \leq a) \cdot P(y \leq b)$

Somit: X, Y sind unabhängig.

A.7. Diskrete Zufallsvariablen

Definition: Eine ZV heißt diskret (verteilt), wenn sie endlich viele Werte $x_1, \ldots, x_n \in \mathbb{R}$ oder abzählbar unendlich viele Werte $x_1, x_2, \ldots \in \mathbb{R}$ annimmt.

Die Gesamtheit der Zahlenpaare (x_i, p_i), wobei $p_i := P(X = x_i)$ nennen wir die Verteilung von X unter P (die Verteilung im Sinne von Kap. 6 erhält man durch Aufsummieren der entsprechenden p_i).

Erwartungswert: Falls die (ev. unendliche) Summe

$E(x) := \sum_i x_i \cdot p_i$

existiert, nennen wir $E(x)$ den Erwartungswert.

Falls die Summe nicht existiert oder gleich unendlich ist, sagt man $E(x)$ ex. nicht (vgl. Übung).

Falls $E(x)$ existiert, nennen wir

$Var(x) := \sum_i (x_i - E(x))^2 \cdot p_i$ die Varianz von X unter P, und
$\sigma(X) := \sqrt{Var(x)}$ die Standardabweichung oder Streuung von X.

Bemerkungen:

- $E(X)$ beschreibt in einem gewissen Sinne die „Mitte“ der Verteilung von X.
- Für $Y := (X - E(X))^2$ gilt $Var(X) = E(Y)$

  $Var(X)$ beschreibt die „mittlere“ quadratische Abweichung von X zum Erwartungswert $E(X)$.
- $Var(X) = \sum_i (x_i^2 - 2x_i \cdot E(X) + E(X)^2) \cdot p_i$
  $= \underbrace{\sum_i x_i^2 \cdot p_i}_{E(X^2)} - 2E(X) \underbrace{\sum_i x_i \cdot p_i}_{E(X)} + E(X) \cdot \underbrace{\sum_i p_i}_{=1}$
  $= E(X^2) - E(X)^2$

Um die Verteilung von X mit Verteilungen anderer Zufallsvariablen zu vergleichen ist es sinnvoll die Standardisierte $X^* := \frac{X - E(X)}{\sigma(X)}$, falls $\sigma(X) > 0$, von X zu betrachten. Es gilt $E(X^*) = 0, Var(X^*) = 1$

A.8. Binomialverteilung

Wir betrachten das n-malige Werfen einer Münze (ev. gezinkt).

Grundraum$\{0,1\}^n = \{\underbrace{(\omega_1, \ldots, \omega_n)}_{\text{„}n\text{-Tupel“}} : \omega_i \in \{0,1\} \text{ für } i = 1, \ldots, n\}$

1: „Erfolg“, 0: „Misserfolg“

Beim einmaligen Würfeln sei die W.keit für Erfolg gleich p,
für Misserfolg also $q := 1 - p \in [0,1]$

Sei nun $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$ und $k := \omega_1 + \ldots + \omega_n$ die Gesamtzahl der Erfolge.

Dann setzen wir $P(\omega) := P(\{\omega\}) = p^k \cdot q^{n-k}$

und $P(A) = \sum_{\omega \in A} p(\omega)$

Weiter sei $X :=$ „Gesamtzahl der Erfolge".

Dann gilt $p_k = P(X = k) = \underbrace{\binom{n}{k}}_{*} \cdot p^k \cdot q^{n-k}$

* Anzahl der $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$ mit $\omega_1 + \ldots + \omega_n = k$

Definition: Eine solche ZV X heißt binomialverteilt mit Parametern $n \in \mathbb{N}$ und $p \in [0,1]$, kurz $B(n,p)$

A.9. Erzeugende Funktionen

Gegeben sei ZV X mit Werten in $\mathbb{N}_0$
und Verteilung $p_k := P(X = k)$ für $k \in \mathbb{N}_0$. Für $s \in [0,1]$ definieren wir
$f_X(s) := \sum\limits_{k=0}^{\infty} p_k \cdot s^k$

Es gilt $f_X(s) \leq \sum\limits_{k=0}^{\infty} p_k = 1$, also insbesondere $f_X(s) < \infty$

Die Abb. $f : [0,1] \to \mathbb{R}$ heißt erzeugende Funktion von X.
Erwartungswert und Varianz von X lassen sich mit Hilfe von f_X berechnen:

Für $s \in [0,1)$ gilt:

- $f'_X = \frac{d}{ds}\left(\sum\limits_{k=0}^{\infty} p_k s^k\right) \underbrace{=}_{\text{Analysis}} \sum\limits_{k=0}^{\infty} \frac{d}{ds}(p_k \cdot s^k)$
  $= \sum\limits_{k=1}^{\infty} p_k \cdot k \cdot s^{k-1}$
- $f''_X = \frac{d}{ds}\left(\sum\limits_{k=1}^{\infty} p_k \cdot k \cdot s^{k-1}\right) \overset{\text{Analysis}}{=} \sum\limits_{k=1}^{\infty} \frac{d}{ds}(p_k \cdot k \cdot s^{k-1}) = \sum\limits_{k=2}^{\infty} p_k \cdot k(k-1) \cdot s^{k-2}$

Setzt man $s = 1$ ein, erhält man:

- $f'_X(1) = \sum\limits_{k=1}^{\infty} p_k \cdot k = E(X)$
- $f''_X(1) = \sum\limits_{k=2}^{\infty} p_k \cdot k(k-1) = \sum\limits_{k=1}^{\infty} p_k \cdot k^2 - \sum\limits_{k=1}^{\infty} p_k \cdot k = E(X^2) - E(X)$

Also:

$$\begin{aligned} Var(X) &= E(X^2) - E(X)^2 \\ &= f''_X(1) + f'_X(1) - (f'_X(1))^2 \end{aligned}$$

Bsp.: Binomialverteilung

X := Anzahl der Erfolge beim n-maligen Werfen einer ev. unfairen Münze mit Erfolgsw.keit p (Bernoulli-Experiment mit Erfolgsw.keit p)

Es gilt also $p_k = P(X = k) = \binom{n}{k} p^k \cdot q^{n-k}$ mit $q = 1 - p$ und $k \in \{0, \dots, n\}$

$$\begin{aligned} f_X(s) &= \sum_{k=0}^{n} \binom{n}{k} p^k \dot{q}^{n-k} \cdot s^k \\ &= \sum_{k=0}^{n} \binom{n}{k} (p \cdot s)^k \cdot q^{n-k} \\ &= (p \cdot s + q)^n \text{ \underline{Binomische Formel}} \end{aligned}$$

$f'_X(s) = n(p \cdot s + q)^{n-1} \cdot p$

$f''_X(s) = n(n-1)(p \cdot s + q)^{n-2} p^2$

Also: $E(X) = n \cdot p$

$$\begin{aligned} Var(X) &= n(n-1) \cdot p^2 + n \cdot p - n^2 \cdot p^2 \\ &= n \cdot p(-p + 1) \\ &= n \cdot p \cdot q \end{aligned}$$

Alternative: X binomialverteilt mit Parametern n, p.

$\Omega := \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \{0, 1\}$ für $i = 1, \dots, n\}$

$P(\omega) = p^k \cdot q^{n-k}$, wobei $k = \omega_1 + \ldots + \omega_n$

Sei $X_i(\omega) := \omega_i$ „Ausgang des i-ten Wurfes“

Dann gilt $X = X_1 + \ldots + X_n$

Definition: Die ZVen $X_1, \dots, X_n$ heißen unabhängig, wenn

$P(X_1 \leq a_1, \dots, X_n \leq a_n) = P(X_1 \leq a_1) \cdot \ldots \cdot (P(X_n \leq a_n)\ \forall a_1, \dots, a_n \in \mathbb{R}$

Bemerkung: Nehmen $X_1, \dots, X_n$ nur Werte in $\mathbb{N}_0$ an, so sind sie genau dann unabhängig, wenn

$P(X_1 = a_1, \ldots, X_n = a_n) = P(X_1 = a_1) \cdot \ldots \cdot P(X_n = a_n)\ \forall a_1, \ldots, a_n \in \mathbb{N}_0$
(für $n = 2$ vergleiche Übung)

Beim n-maligen Werfen gilt:

$$P(X_i = a_i) = \begin{cases} p & \text{für } a_i = 1 \\ q & \text{für } a_i = 0 \\ 0 & \text{sonst} \end{cases}$$
berechenbar mit der Formel $P(A) = \sum\limits_{\omega \in A} p(\omega)$ mit $A = \{X_i = a_i\}$
Es gilt also für $\omega = (a_1, \ldots, a_n), k = a_1 + \ldots + a_n$

$$\begin{aligned} P(X_1 = a_1, \ldots, X_n = a_n) &= P(\omega) \\ &= P(X_1 = a_1) \cdot \ldots \cdot P(X_n = a_n) \end{aligned}$$

(mit $a_i \in \{0, 1\}$, alle anderen W.keiten sind Null)

Also sind $X_1, \ldots, X_n$ unabhängig.

Satz: Seien $X_1, \ldots, X_n$ ZVen mit existierenden Erwartungswerten und Varianzen.

i.) $E(x_1 + \ldots + x_n) = E(x_1) + \ldots + E(x_n)$

ii.) Sind $X_1, \ldots, X_n$ unabhängig, so gilt $Var(X_n + \ldots + X_n) = Var(X_1) + \ldots + Var(X_n)$ (vgl. Übungen für $n = 2$ und Werte in $\mathbb{N}_0$).

Für die Binomialverteilung gilt also:

$E(X) = E(X_1) + \ldots + E(X_n) = n \cdot p,$, da $E(X_i) = p$

$Var(X) = Var(X_1) + \ldots + Var(X_n) = n \cdot p \cdot q$, da $Var(X_i) = p \cdot q$

A.10. Poisson-Verteilung

Ein bestimmtes „Ereignis“ treffe im Laufe der Zeit beliebig oft und zu zufälligen Zeitpunkten ein. Mit X_I beschreiben wir wie oft ein solches Ereignis in einem Intervall I eintritt.
Der Erwartungswert $E(X_I) =: \mu > 0$ sei vorgegeben (z.B. durch Messungen) Was ist die Verteilung $p_k = P(X_I = k)$ von X? Wie soll sie gewählt werden?

Dazu teilen wir I auf in kleinere Intervalle $I_1, \ldots, I_n$ von identischer Länge.
Mit X_{I_i}, für $i = 1, \ldots, n$, bezeichnen wir die Anzahl der „Ereignisse“ in I_i

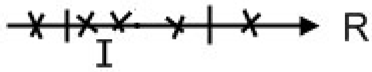


Abbildung 4: $X_I = 3$

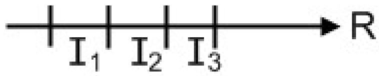


Abbildung 5: Einteilung von I in Intervalle

Wir nehmen an:

i.) Die ZVen $X_{I_1}, \ldots, X_{I_n}$ seien unabhängig und identisch verteilt („independent, identically distributed", kurz: iid)

ii.) Die Intervalle $I_1, \ldots, I_n$ seien derart klein (bzw. n derart groß), dass höchstens ein „Ereignis" pro Intervall I_i statt findet.

$X_I = X_{I_1} + \ldots + X_{I_n}$ ist dann binomialverteilt und es gilt: $\mu = E(X_I) = n \cdot E(X_i)$

Also gilt: $p := P_n(X_{I_i} = 1) = \frac{\mu}{n}$ mit n Anzahl der Invervalle I_i

Was passiert nun für $n \to \infty$?

$$\begin{aligned}
P_n(X_I = k) &= \binom{n}{k} \cdot p^k \cdot q^{n-k} \\
&= \frac{n(n-1) \cdot (n-k+1)}{k!} \cdot p^k \cdot \frac{(1-p)^n}{(1-p)^k} \\
&= \frac{1}{k!} \cdot \frac{n}{n} \cdot \frac{(n-1)}{n} \cdot \ldots \cdot \frac{(n-k+1)}{n} \cdot \mu^k \cdot \frac{(1-\frac{\mu}{n})^n}{(1-\frac{\mu}{n})^k}
\end{aligned}$$

Also: $\lim\limits_{n\to\infty} P_n(X_I = k) \to \frac{\mu^k}{k!} \cdot e^{-\mu}$, da $\lim\limits_{n\to\infty} (1 - \frac{x}{n})^n = e^{-x}$ ist für $x \in \mathbb{R}$

Definition: Eine ZV X mit Verteilung $P(X = k) = \frac{\mu^k}{k!} \cdot e^{-\mu}$ für $k \in \mathbb{N}_0$ heißt Poisson-verteilt mit Parameter $\mu > 0$. Die Verteilung von X wird kurz mit $P_n(\mu)$ bezeichnet.

Beachte: $\sum\limits_{k=0}^{\infty} \frac{\mu^k}{k!} = e^\mu$ für $\mu \in \mathbb{R}$ „Reihendarstellung der e-Funktion".
Übung: Bestimmen Sie $f_X, E(X)$ und $Var(X)$.

A.11. Geometrische Verteilung

Wir betrachten das Bernoulli-Experiment mit Erfolgsw.keit $p > 0$ und Misserfolgsw.keit $q = 1 - p$. Die ZV X zähle die Misserfolge bis zum erstmaligen Erfolg. Die Verteilung von X ist also

$P(X = k) = q^k \cdot p$ für $k \in \mathbb{N}_0$

Kann das Experiment unendlich lange dauern?

$$\begin{aligned} P(X = \infty) &= 1 - P(X < \infty) \\ &= 1 - \sum_{k=0}^{\infty} P(X = k) \\ &= 1 - p \underbrace{\sum_{k=0}^{\infty} q^k}_{*} \\ &= 0 \end{aligned}$$

$(*) = \frac{1}{1-q} = \frac{1}{p}$ für $q \in [0, 1)$ („geometrische Reihe“)

Definition: Eine ZV X mit Verteilung $P(X = k) = p \cdot q^k$ für $p \in [0, 1], q = 1-p, k \in \mathbb{N}_0$ heißt <u>geometrisch verteilt</u> mit Parameter p. Die zugehörige Verteilung wird kurz mit $G(p)$ bezeichnet

A.12. Zufallsvariablen mit Dichten

Definition: Eine Funktion $f : \mathbb{R} \to \mathbb{R}$ heißt <u>Dichte</u> (W.keitsdichte, Verteilungsdichte), falls

- $f(x) \geq 0$ für jedes $x \in \mathbb{R}$
- f ist integrierbar
- $\int\limits_{-\infty}^{\infty} f(x)dx = 1$

Ist nun X eine ZV, deren Verteilungsfunktion $F(t) = P(X \leq t)$ für $t \in \mathbb{R}$ gegeben ist durch
$F(t) = \int\limits_{-\infty}^{t} f(x)dx$

dann sagt man f sei eine Dichte von X oder X habe die Dichte f. Ist f stetig in $t \in \mathbb{R}$, so gilt $F'(t) = f(t)$ (Hauptsatz der Differential-/Integralrechnung).

Für $a < b$ gilt $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a) = \int\limits_a^b f(x)dx$

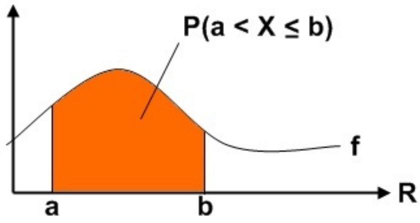


Abbildung 6: Verteilungsfunkton $F(t)$

Bemerkung zur Existenz: Gegeben sei eine bel. Dichte f.

Es existiert eine eindeutige kleinste Familie $\mathcal{B} \subset \mathcal{P}(\mathbb{R})$, die sogenannte Borel-$\sigma$-Algebra (unabhängig von f), und eine Abbildung $P : \mathcal{B} \to [0,1]$, so dass

- $\mathcal{B}$ alle Intervalle enthält.
- $(\mathbb{R}, \mathcal{B}, P)$ ein W.raum ist.
- $P((-\infty, t]) = \int\limits_{-\infty}^{t} f(x)dx$ gilt $\forall\ t \in \mathbb{R}$.

Die Abbildung $X := id_{\mathbb{R}}$ ist dann eine ZV mit Dichte f.

Es gilt $\mathcal{B} \neq \mathcal{P}(\Omega)$. Würde man statt dessen

$\mathcal{E} := \mathcal{P}(\mathbb{R})$ wählen, so existierte im Allgemeinen keine W.keit P, so dass obige Bedingungen erfüllt sind.

Definition: Sei X eine ZV mit Dichte f. Dann heißen

$E(X) := \int\limits_{-\infty}^{\infty} x \cdot f(x)dx$ der Erwartungswert von X (falls das Integral existiert) und

$Var(X) := \int\limits_{-\infty}^{\infty} (x - E(X))^2 f(x)dx$ die Varianz von X (falls das Integral existiert).

Interpretation:

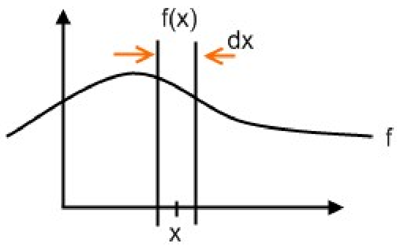


Abbildung 7: $f(x)dx \approx \int\limits_{x-\frac{dx}{2}}^{x+\frac{dx}{2}} f(x)dx = P(|X-x| \leq \frac{dx}{2}) \approx P(X \approx x)$

A.13. Gleichverteilung

Definition: Eine ZV X heißt gleichverteilt auf dem Intervall $[a, b]$ für $a \leq b$, wenn

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

eine Dichte von X ist. Die Gleichverteilung auf $[a, b]$ wird kurz mit $U(a, b)$ bezeichnet. Es gelten

$$\begin{aligned} E(X) &= \int\limits_{-\infty}^{\infty} x \cdot f(x)\ dx \\ &= \int\limits_{a}^{b} \frac{x}{b-a}\ dx \\ &= \frac{1}{2 \cdot (b-a)} x^2 \Big|_a^b \\ &= \frac{b^2 - a^2}{2 \cdot (b-a)} \\ &= \frac{b+a}{2} \end{aligned}$$

$$\begin{aligned} Var(X) &= \frac{1}{b-a}\int_a^b \left(x-\frac{b+a}{2}\right)^2 \, dx \\ &= \frac{1}{b-a}\int_a^b x^2 \, dx - \underbrace{2\cdot\frac{b+a}{2}\int_a^b x \, dx}_{=\frac{(b+a)^2}{2}} + \frac{(b+a)^2}{4} \\ &= \frac{1}{b-a}\left(\frac{b^3-a^3}{3}\right) - \frac{(a+b)^2}{4} \\ &= \frac{(b-a)^2}{12} \end{aligned}$$

A.14. Exponentialverteilung

Definition: Eine ZV X heißt exponentialverteilt mit Parameter $\alpha > 0$, wenn

$$f(x) = \begin{cases} 0 & \text{für } x < 0 \\ \alpha \cdot e^{-\alpha x} & \text{für } x \geq 0 \end{cases}$$

eine Dichte von X ist. Kurz: $Exp(\alpha)$.

Es gilt $F(t) = P(X \leq t) = \begin{cases} 0, & t < 0 \\ 1 - e^{-\alpha t}, & t \geq 0 \end{cases}$

Interpretation: Für alle $x, t \geq 0$ gilt

$$P(X > t + x | X > t) \overset{(*)}{=} P(X > x)$$

D.h. X beschreibt die Zeitdauer, bis zum Eintreten eines „Ereignisses“, wobei die Verteilung dieser Zeitdauer nicht davon abhängt, wann man zu messen beginnt.

Bemerkung: Umgekehrt kann gezeigt werden, dass jede ZV (mit stetiger Verteilungsfunktion), welche (*) erfüllt, exponentiell verteilt ist mit Parameter $\alpha := -\log P(X > 1)$;

A.15. Normalverteilung

Definition: Eine ZV X heißt standardnormalverteilt, wenn

$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $x \in \mathbb{R}$ eine Dichte von X ist.

Bemerkung: Es kann gezeigt werden, dass

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1$$

Für $\mu \in \mathbb{R}$ und $\sigma > 0$ betrachten wir die ZV $Y := \mu + \sigma X$, wobei X standardnormalverteilt sei.

$$\begin{aligned} P(Y \leq t) &= P(X \leq \frac{t-\mu}{\sigma}) \\ &= \int_{-\infty}^{\frac{t-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\ dx && \Big| \text{ Substitution mit } x = \frac{y-\mu}{\sigma} \\ &= \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \end{aligned}$$

$\Rightarrow Y$ hat eine Dichte $\varphi_{\mu,\sigma^2}(y) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}}$

Definition: Eine ZV Y mit Dichte φ_{μ,σ^2} heißt <u>normalverteilt</u> mit Parameter μ und σ^2. Kurzschreibweise $\mathcal{N}(\mu, \sigma^2)$.

Bemerkungen:

- Die Standardnormalverteilung ist $\mathcal{N}(0, 1)$
- Ist X normalverteilt mit Parameter μ und σ^2, dann ist die Standardisierte $X^* := \frac{X-\mu}{\sqrt{\sigma^2}}$ von X standardnormalverteilt. Es gilt also

  $P(X \leq t) = \Phi\left(\frac{t-\mu}{\sqrt{\sigma^2}}\right)$ mit

  $\Phi(t) := \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$
- Ist X normalverteilt mit Parametern μ und σ^2, so gilt $E(X) = \mu, Var(X) = \sigma^2$ (Übung)

A.16. Tschebyscheffsche Ungleichung

Sei Y eine ZV mit Dichte f, und $\epsilon > 0$. Es gilt

$$\begin{aligned}\epsilon^2 P(|Y| \geq \epsilon) &= \epsilon^2(P(Y < -\epsilon) + P(Y > \epsilon)) \\ &= \int_{-\infty}^{-\epsilon} \epsilon^2 f(x)\ dx + \int_{\epsilon}^{\infty} \epsilon^2 f(x)\ dx \\ &\leq \int_{-\infty}^{-\epsilon} x^2 f(x)\ dx + \int_{\epsilon}^{\infty} x^2 f(x)\ dx \\ &= \int_{-\infty}^{\infty} x^2 f(x)\ dx\end{aligned}$$

Also: $P(|Y| \geq \epsilon) \leq \frac{\int_{-\infty}^{\infty} y^2 f(y)\ dy}{\epsilon^2} \overset{*}{=} \frac{E(Y^2)}{\epsilon^2}$

(* vgl. Übung 22)

Tschebyscheffsche Ungleichung:
Sei X eine bel. ZV mit existierendem Erwartungswert und existierender Varianz. Dann gilt

$P(|X - E(X)| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2}\ \forall\ \epsilon > 0$

A.17. Gesetz der großen Zahlen

Gegeben seien iid ZV $X_1, X_2, ...$ mit ex. Erwartungswerten $\mu = E(X_i)$ und Varianzen $\sigma^2 = Var(X_i)$.
Wir setzen $S_n := X_n + ... + X_n$ und

$\overline{X}_n := \frac{S_n}{n}$ arithmetisches Mittel

Es gelten:

$E(\overline{X}_n) = \frac{n \cdot E(X_i)}{n} = \mu$

$Var(\overline{X}_n) = \frac{n \cdot Var(X_i)}{n^2} = \frac{\sigma^2}{n}$

Das (schwache) Gesetz der großen Zahlen:

$\lim\limits_{n\to\infty} P(|\overline{X}-\mu| \geq \epsilon) \overset{*}{\leq} \lim\limits_{n\to\infty} \frac{\sigma^2}{n\cdot\epsilon^2} = 0$

(*) Tschebyscheff

Bsp.: Sei $Y_1, Y_2, ...$ die endlichwertigen Ausgänge eines wiederholten Zufallsexperimentes und

$X_i := \begin{cases} 1 & \text{für } Y_i = y \\ 0 & \text{sonst} \end{cases}$

Dann ist $\overline{X}_n$ die gemittelte Anzahl der Ausgänge mit y.

$E(\overline{X}_n) = E(X_i) = P(Y_i = y)$

Somit $P(|\overline{X}_n - P(Y_i = y)| \geq \epsilon) \to 0$

Das heißt, dass der axiomatische Zugang von Kolmogoroff den „empirischen Zugang beinhaltet“.

A.18. Zentraler Grenzwertsatz

Seien $X_1, X_2, \dots$ unabhängige, identisch verteilte ZVen mit Erwartungswert

$\mu := E(X_i) \in \mathbb{R}$ und Varianz $\sigma^2 = Var(X_i) \in (0, \infty)$.

Das Gesetz der großen Zahlen besagt dann, dass für jedes $\epsilon > 0$ gilt:

$\lim\limits_{n\to\infty} P(|\overline{X}_n - \mu| > \epsilon) = 0$, wobei $\overline{X} := \frac{X_1+...+X_n}{n}$ für $n \in \mathbb{N}$.

D.h. $\overline{X}_n$ weicht nur „wenig“ ab vom Erwartungswert μ, wenn n „groß“ ist.
Wie ist $\overline{X}_n$ um μ herum verteilt?

Genauer: wie verhält sich $P(\frac{a\cdot\sigma}{\sqrt{n}} \leq \overline{X}_n - \mu \leq \frac{b\cdot\sigma}{\sqrt{n}}) = P\Big(a \leq \frac{X_1+...+X_n - n\ \mu}{\sqrt{n}\ \sigma} \leq b\Big)$, $a \leq b$ für große n?

Satz(Zentraler Grenzwertsatz)

Unter obigen Bedingungen an $X_1, X_2, \dots$ und mit

$S_n := X_1 + \ldots + X_n$ und $S_n^* := \frac{S_n - n\ \mu}{\sqrt{n\ \sigma^2}}$
(S_n^* ist die Standardisierte von S_n), gilt

$\lim\limits_{n\to\infty} P\Big(a \leq S_n^* \leq b\Big) = \Phi(b) - \Phi(a)\ \forall\ a \leq b\ (a \in \mathbb{R} \cup \{-\infty\}, b \in \mathbb{R} \cup \{\infty\})$,

wobei $\Phi(t) := \int\limits_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}\ dx$ für $t \in \mathbb{R}$

die Verteilungsfunktion der Standardnormalverteilung ist.

Es kann gezeigt werden, dass die Konvergenz sogar gleichmäßig ist, d.h. es gilt:

$\lim\limits_{n\to\infty} \sup\limits_{a\leq b} \Big| P\Big(a \leq S*_n \leq b\Big) - (\Phi(b) - \Phi(a))\Big| = 0$

Bsp.: Binomialverteilung

Betrachten das Bernoulli-Experiment:

$S_n := X_1 + \ldots + X_n$, wobei $X_1, \ldots, X_n$ identisch verteilt, unabhängig und mit

$P(X_i = 1) = p$ und $P(X_i = 0) = q = (1 - p)$.

Dann ist S_n binomialverteilt mit Parametern n und p. Es gilt:

$E(X_i) = p$ und $Var(X_i) = p \cdot q$

Für $a, b \in \mathbb{N}$ gilt dann:

$P(a \leq S_n \leq b) = \sum\limits_{k=a}^{b} P(S_n = k) = \sum\limits_{k=a}^{b} \binom{n}{k} p^k \cdot q^{n-k}$

$P\Big(\frac{a-n\ p}{\sqrt{n\ p\ q}} \leq S_n^* \leq \frac{b-n\ p}{\sqrt{n\ p\ q}}\Big) \underset{\approx}{\text{ZGS}} \Phi\Big(\frac{b-n\ p}{\sqrt{n\ p\ q}}\Big) - \Phi\Big(\frac{a-n\ p}{\sqrt{n\ p\ q}}\Big)$

für große n (Faustregel $n\ p\ q \geq 9$).

Diese Approximation lässt sich verbessern, indem man eine Diskretheitskorrektur einführt:

$P(a \leq S_n \leq b) \approx \Phi\Big(\frac{b+\frac{1}{2}-n\ p}{\sqrt{n\ p\ q}}\Big) - \Phi\Big(\frac{a-\frac{1}{2}-n\ p}{\sqrt{n\ p\ q}}\Big)$

Es gilt z.B. $\underbrace{P(S_{2n} = 2n)}_{=p^{2n}} = P(2n \leq S_n \leq 2n) > 0$

Bemerkung: Für $t < 0$ gilt $\Phi(t) = 1 - \Phi(-t)$ (vgl. Aufgabe 17).

Es reicht also, wenn $\Phi(t)$ für $t > 0$ tabelliert ist.

A.19. Mehr zur Unabhängigkeit

Erinnerung: Zwei ZV X,Y heißen unabhängig, wenn

$$P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y) \; \forall \; x, y \in \mathbb{R}$$

Diskrete ZVen:
Wir nehmen an X nehme höchstens abzählbar viele (endlich viele oder abzählbar unendlich viele) Werte x_i mit $i \in I$ ($I = \{1, \ldots, n\}$ oder $I = \mathbb{N}$) an und Y nehme höchstens abzählbar viele Werte y_j mit $j \in J$ an.

Satz: X, Y sind genau dann unabhängig, wenn

$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j) \; \forall \; i \in I, j \in J$.
(vgl. Übung 17 für $x_i, y_j \in \mathbb{N}$).

Betrachten die Matrix $P = (p_{ij})_{i \in I \, , \, j \in J}$ mit $p_{ij} := P(X = x_i, Y = y_j)$,
sowie die Zeilenvektoren

$p := (p_i)_{i \in I}$ mit $p_i := P(X = x_i)$ und

$q := (q_j)_{j \in J}$ mit $q_j := P(Y = y_j)$

Dann gilt X, Y sind genau dann unabhängig, wenn

$P = p^t \cdot q = (p_i \cdot q_j)$, $i \in I$, $j \in J$ (Matrixmultiplikation)

$$\Bigg(\qquad\quad\Bigg) = \Bigg(\quad\Bigg) \cdot \big(\qquad\quad\big)$$

Definition: Für zwei beliebige ZVen X, Y heißt

$F : \mathbb{R}^2 \to [0, 1]$, definiert durch

$F(x, y) := P(X \leq x, Y \leq y)$ für $x, y \in \mathbb{R}$,

die <u>gemeinsame Verteilungsfunktion</u> von (X, Y).

Kennt man die gemeinsame Verteilungsfunktion von (X, Y), so erhält man die Verteilungsfunktion F_1 von X und die Verteilungsfunktion F_2 von Y durch

$$\begin{aligned} F_1(x) := P(X \leq x) &= P(X \leq x, Y \leq \infty) \\ &= F(x, \infty) := \lim_{y \to \infty} F(x, y) \end{aligned}$$

und analog $F_2(y) = \lim_{x\to\infty} F(x,y)$

F_1 und F_2 sind die Randverteilungen von (X,Y).

Es gilt: X,Y sind genau dann unabhängig, wenn $F(x,y) = F_1(x)\cdot F_2(y)$ gilt $\forall\ x,y\in\mathbb{R}$

ZVen mit Dichten:
Man sagt, dass zwei ZVen X,Y eine gemeinsame Dichte $f:\mathbb{R}^2\to[0,\infty)$ besitzen, wenn

$F(x,y) = P(X\le x, Y\le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)\ dt\ ds$ gilt für $x,y\in\mathbb{R}$

Es gilt:

$$\begin{aligned}P(X\le x) = P(X\le x, Y\le\infty) &= \int_{-\infty}^{x}\int_{-\infty}^{\infty} f(s,t)\ dt\ ds\\ &= \int_{-\infty}^{x} f_1(s)\ ds\end{aligned}$$

mit $f_1(s) := \int_{-\infty}^{\infty} f(s,t)\ dt$

Also ist f_1 eine Dichte von X und analog ist $f_2(t) := \int_{-\infty}^{\infty} f(s,t)\ ds$ eine Dichte von Y.
f_1, f_2 sind die Randdichten von (X,Y), bzw. von f.

Wir nehmen nun an, es gelte $f(s,t) = f_1(s)\cdot f_2(t)$ für alle $s,t\in\mathbb{R}$.

Dann gilt:

$$\begin{aligned}F(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f_1(s)f_2(t)\ dt\ ds\\ &= \int_{-\infty}^{x} f_1(s)\underbrace{\Big(\int_{-\infty}^{y} f_2(t)\ dt\Big)}_{=F_2(y)} ds\\ &= F_2(y)\cdot\underbrace{\int_{-\infty}^{x} f_1(s)\ ds}_{=F_1(x)}\end{aligned}$$

Man kann zeigen, dass auch die Umkehrung gilt:

Satz: Sind X und Y ZVen mit Dichten f_1 und f_2, so sind X, Y genau dann unabhängig, wenn $f(x,y) := f_1(x) \cdot f_2(y)$ eine gemeinsame Dichte von (X,Y) ist.

A.20. Faltung

Es seien X, Y zwei unabhängige ZVen. Q_1 sei die Verteilung von X, Q_2 sie die Verteilung von Y.

Mit der Faltung $Q := Q_1 * Q_2$ von X und Y bezeichnet man die Verteilung der ZV $X+Y$.

Betrachte zwei Spezialfälle:

1. Fall: X, Y nehmen nur Werte in $\mathbb{N}_0$ an.
Betrachten $Q_1(n) := P(X = n), Q_2(n) := P(Y = n)$

Dann gilt:

$$\begin{aligned} Q(n) := P(X+Y=n) &= \sum_{k=0}^{n} P(X=k, Y=n-k) \\ &= \sum_{k=0}^{n} P(X=k) \cdot P(Y=n-k) \\ &= \sum_{k=0}^{n} Q_1(k) \cdot Q_2(n-k) \end{aligned}$$

2. Fall: X besitzt eine Dichte f_1 und Y besitzt eine Dichte f_2.

Betrachten $Q_1(x) := P(X \leq x) = \int\limits_{-\infty}^{x} f_1(t)\ dt$

$Q_2(y) := P(Y \leq y) = \int\limits_{-\infty}^{x} f_2(t)\ dt$

$Q(z) := P(X+Y \leq z)$

Heuristisch:

$P(X \approx x, Y \leq z-x) = P(X \approx x) \cdot P(Y = z-x) \approx f_1(x)\ dx \cdot \int\limits_{-\infty}^{z-x} f_2(s)\ ds$

Durch „Aufsummieren“, bzw. „Aufintegrieren“ erhält man

$$P(X+Y=Z) = \int_{-\infty}^{\infty} f_1(x) \cdot \underbrace{\int_{-\infty}^{z-x} f_2(s)\ ds}_{=\int_{-\infty}^{z} f_2(y-x)\ dy\ (*)} \quad dx$$

$$= \int_{-\infty}^{z} \int_{-\infty}^{\infty} f_1(x) \cdot f_2(y-x)\ dx\ dy$$

(*) nach Substitution $s = y - x$

Also: $g(y) = \int_{-\infty}^{\infty} f_1(x) f_2(y-x)\ dx$ ist eine Dichte von $X+Y$.

A.21. Poisson-Punktprozess

Der Poisson-Punktprozess beschreibt ein in der Zeit wiederkehrendes Ereignis (z.B. Telefonanrufe, Teilchenemission beim radioaktiven Zerfall, Meteoriteneinschläge,...)

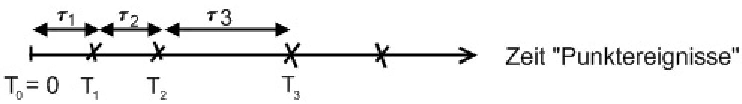


Abbildung 8: Punktereignisse

Ansatz: Die Zeitspannen $\tau_1, \tau_2, ...$ zwischen den Punktereignissen seien unabhängig und identisch exponential verteilt zum Parameter $\alpha > 0$ (Begründung: siehe Kap. 14).

Setzen nun $T_0 := 0$ und $T_n := \tau_1 + ... + \tau_n$ für $n \in \mathbb{N}$
„Zeitpunkt des n-ten Punktereignisses".

Wie ist τ_n verteilt?

Aufgabe 31 a): $f_n(x) := \begin{cases} \frac{\alpha^n \cdot x^{n-1} \cdot e^{-\alpha \cdot x}}{(n-1)!} & , x \geq 0 \\ 0 & , x < 0 \end{cases}$ ist eine Dichte von T_n.

Die Verteilung von T_n heißt Gamma-Verteilung, benannt nach der Gamma-Funkion $/Gamma(n) := \int_0^{\infty} x^n \cdot e^{-x}\ dx = (n-1)!$

Beweisschema:

Vollständige Induktion nach $n \in \mathbb{N}$

Ind.verankerung $n = 1 : T_1 = \tau_1$ ist exp. verteilt mit Parameter α nach Voraussetzung.

	Also ist f_1 eine Dichte von T_1
Ind.schritt	$n \to n+1$
Ind.annahme:	f_n ist eine Dichte von T_n.
	Zeigen Sie, dass dann auch f_{n+1} eine Dichte von T_{n+1} ist.
Hinweis:	Aus der Unabhängigkeit der $\tau_1, ..., \tau_n + 1$ folgt auch die Unabhängigkeit von $T_n = \tau_1 + ... + \tau_n$ und τ_{n+1}

Aufgabe 31 b): Für $t > 0$ gilt: $P(T_n > t) = \sum\limits_{k=0}^{n-1} \frac{(\alpha t)^k}{k!} \cdot e^{-\alpha t}$

Beweisschema:
Vollständige Induktion nach $n \in \mathbb{N}$

Ind.verankerung: $n = 1$ Zeigen Sie $P(T_1 > t) = e^{-\alpha t}$

Hinweis:	$P(T_1 > t) = 1 - P(T_1 \leq t) = 1 - \int\limits_{-\infty}^{t} f_1(x)\ dx = \int\limits_{t}^{\infty} f_1(x)\ dx$
Ind.Schritt:	$n \to n+1$
Ind.Annahme:	$P(T_n > t) = \sum\limits_{k=0}^{n-1} \frac{\alpha t)^k}{k!} \cdot e^{-\alpha t}$
	Zeigen Sie, dass dann $P(T_{n+1} > t) = \sum\limits_{\alpha=0}^{n} \frac{(\alpha t)^k}{k!} \cdot e^{-\alpha t}$
Hinweis:	$P(T_{n+1} > t) = \int\limits_{t}^{\infty} f_{n+1}(x)\ dx$, partielle Integration.

Es sei nun $X_t :=$ „Anzahl der Punktereignisse im Intervall $I := [0, t]$“

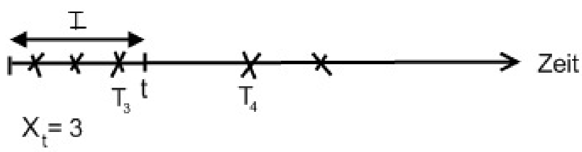


Abbildung 9: Punktereignisse im Intervall

In Kapitel 10 wurde angenommen, dass eine solche ZV Poisson-verteilt ist. Wie sieht es hier aus?

Offensichtlich gilt: $X_t = n \Leftrightarrow T_n \leq t$ und $T_{n+1} > t$
Somit gilt $P(X_t = n) = P(T_n \leq t < T_{n+1})$

außerdem $P(t < T_{n+1}) = P(T_n \leq t < T_{n+1}) + \underbrace{P(t < T_n, t < T_{n+1})}_{=P(t<T_n)\ ,\ \text{da}\ T_n<T_{n+1}}$

Also:

$$\begin{aligned}
P(X_t = n) &= P(T_{n+1} > t) - P(T_n > t) \\
&= \sum_{k=0}^{n} \frac{(\alpha t)^k}{k!} \cdot e^{-\alpha t} - \sum_{k=0}^{n-1} \frac{(\alpha t)^k}{k!} \cdot e^{-\alpha t} \\
&= \frac{(\alpha t)^n}{n!} \cdot e^{-\alpha t}
\end{aligned}$$

D.h. X_t ist Poisson-verteilt mit Parameter $\alpha \cdot t$ und es gilt:

$$E(X_t) = Var(X_t) = \alpha \cdot t$$

Insbesondere ist $\alpha = E(X_1)$ die erwartete Anzahl Punktereignisse pro Zeiteinheit.

B. Statistik

Die Statistik ist ein Konzept, bei dem man von einer sehr großen Grundgesamtheit ausgeht, die nicht komplett zugänglich ist, und deren Struktur man durch Daten (Stichprobe) aufklären will. Dabei geht man davon aus, dass die Grundgesamtheit groß genug ist, so dass sie von der Stichprobe nicht beeinflusst wird.

Die Statistik baut auf der W.theorie auf. Während die W.keitstheorie ohne Daten auskommt, sind diese gerade ein wesentliches Merkmal der Statistik.

B.1. Statistisches Modell und Stichproben

Die Grundgesamtheit wird in Form einer ZV beschrieben. In der parametrischen Statistik wird angenommen, dass die Verteilung P_ϑ dieser ZV durch einen Parameter ϑ einer Parametermenge Θ charaktierisiert ist. Ein statistisches Modell ist also von der Form $\{P_\vartheta : \vartheta \in \Theta\}$.

Nun werden Daten $\underline{x} := (x_1, \ldots, x_n)$ erhoben (Stichprobe vom Umfang $n \in \mathbb{N}$). Man nimmt an, dass $x_1, \ldots, x_n$ Realisierungen von (meistens) ZVen $X_1, \ldots, X_n$ sind, wobei diese auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P_\vartheta)$ für ein geeignetes $\vartheta \in \Theta$ definiert sein soll.

Aus der Stichprobe $\underline{x}$ berechnet man sogenannte Kennzahlen, die charakteristisch sein sollen für die Grundgesamtheit.

Sichprobenmittelwert: $\overline{x} := \frac{1}{n} \sum_{i=1}^{n} x_i$

(unverzerrte) Stichprobenvarianz: $s^2 := \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \overline{x})^2$

Diese Kennzahlen sind Realisierungen (d.h. Auswertungen in einem $\omega \in \Omega$) der ZVen
$\overline{X} := \frac{1}{n} \sum\limits_{i=1}^{n} X_i$, $S^2 := \frac{1}{(n-1)} \sum\limits_{i=1}^{n} (X_i - \overline{X})^2$

Beachte: $\overline{x}_i s^2$ sind auf konkreten Daten beruhende Zahlen. $\overline{X}, S$ sind auf dem statistischen Modell beruhende ZVen.

Offenbar gilt: $E(\overline{X}) \quad = E(x_i)$
$Var(\overline{X}) = \frac{1}{n} Var(x_i)$„Mitteln verringert die Varianz“

Wir fassen $X_1, \ldots, X_n$ zu einem Zufallsvektor $\underline{X} := (X_1, \ldots, X_n)$ zusammen, welcher auch Stichprobe vom Umfang $n \in \mathbb{N}$ genannt wird.

B.2. Schätzen von Parametern

Gegeben sei ein statistisches Modell $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$, welches die Grundgesamtheit (approximativ) beschreibt. Mit Hilfe von Daten $x_1, \ldots, x_n$ soll der Parameter ϑ oder irgend eine andere Kennzahl $q = q(\vartheta)$ geschätzt werden.

Definition: Gegeben sei eine Abbildung $T : \mathbb{R}^n \to \mathbb{R}$.
Dann heißt die ZV $U_n := T(\underline{X}) := T \circ \underline{X}$, $\Omega \to \mathbb{R}$ ein Schätzer für q.

Die Zahl T($\underline{x}$) heißt dann Schätzwert für q.

Was sind „gute“ Schätzer?

Definition:

i.) Schätzer U_n heißt erwartungstreu für q, wenn $E(U_n) = q$ gilt

ii.) Eine Folge $(U_n)_{n\in\mathbb{N}}$ von Schätzern heißt konsistent,
wenn $\lim\limits_{n\to\infty} P(|U_n - q| > \epsilon) = 0$ für jedes $\epsilon > 0$

Bsp.: Münzwurf

$p := P[\text{Kopf}] \in [0, 1]$ unbekannt

Wie schätzt man p?

ZV X mit $X = \begin{cases} 1 & \text{falls Kopf} \\ 0 & \text{sonst} \end{cases}$
X binomialverteilt mit $(1, p)$

Statistisches Modell:

$\mathcal{P} = \{Bin(1,p) : p \in [0,1]\}$

Allgemeiner:

$X_1, \ldots, X_n$ = Ergebnisse von n unabhängigen Münzwürfen. (Zufallsvariablen)

Kanonischer Schätzer für p:

$T(x_1, \ldots, x_n) = \frac{\#\text{ Einsen}}{\#\text{ Würfe}} = \frac{x_1+\ldots+x_n}{n}$

$T(0,1,1,1,0,1,0) = \frac{4}{7}$

$T : \mathbb{R}^n \to \mathbb{R}$ mit $\mathbb{R}^n$ Menge der möglichen Daten

$U_n := T(X_1, \ldots, X_n) = \frac{X_1+\ldots+X_n}{n} = \bar{X}$

$U_n : \Omega \to \mathbb{R}$ der Schätzer als Zufallsvariable

U_n ist erwartungstreuer Schätzer von p:

$$E[U_n] = E\left[\frac{X_1+\ldots+X_n]}{n}\right] = \frac{1}{n}(\underbrace{E[X_1]}_{0\cdot P[X_1=0]+1\cdot P[X_1=1]} + \ldots + E[X_n]) = \frac{n\cdot p}{n} = p$$

$(U_n)_n$ ist konsistenter Schätzer von p:

$0,1,1,1,0,1,0,1,1,\ldots \Big| (X_n)_n$

$0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{3}{5}, \frac{4}{6}, \ldots \Big| (U_n)_n$ sollte gegen p konvergieren im Sinne der Konvergenz des schwachen Gesetzes der großen Zahlen.

Schwaches Gesetz der großen Zahlen: $\forall\ \epsilon > 0\ P[\ |\ U_n - p| \geq \epsilon\ ] \to 0$ für $n \to \infty$.

Konvergenzgeschwindigkeit ist approximativ berechenbar:

$$P\left[\ \left|\frac{X_1+\ldots+X_n}{n} - p\right| \geq \epsilon\right] = P[|(X_1 + \ldots + X_n) - \underbrace{n\ p}_{E[X_1+\ldots+X_n]}\ | \geq n\ \epsilon]$$

$$= P\left[\ |\underbrace{\frac{(X_1 + \ldots + X_n) - n\ p}{\sqrt{Var(X_1 + \ldots + X_n)}}}_{\approx \text{ standardnormalverteilt}}| \geq \frac{n\ \epsilon}{\sqrt{Var(X_1+\ldots+X_n)}}\right] = P\left[\ \left|\frac{X_1+\ldots+X_n - n\ p}{\sqrt{n\ p(1-p)}}\right| \geq \frac{n\ \epsilon}{\sqrt{n\ p(1-p)}}\right]$$

$$\overset{\text{ZGWS}}{\approx} \Phi(-z) + 1 - \Phi(z) = 2(1-\Phi(z)) = 2(1-\Phi\Big(\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\Big)$$

B.3. Erwartungstreue Schätzer von Erwartungswert und Varianz

Im letzten Kapitel haben wir gesehen:

Falls $X_1, \ldots, X_n \sim Bin(1,p)$ unabhängig, so ist $\bar{X}$ ein erwartungstreuer Schätzer für $p = E[X_1]$

Nach der gleichen Rechnung ergibt sich:

$\bar{X}$ schätzt $E[X_1]$ erwartungstreu, egal welche Verteilung $X_1, \ldots, X_n$ wir haben (unter der Annahme, dass alle X_i die gleiche Verteilung haben).

Behauptung:

$S^2 := \frac{1}{n-1} \sum\limits_{i=1}^{n} (X_i - \bar{X})^2$ schätzt $\sigma^2 := Var(X_1)$ erwartungstreu.

Beweis:

$Y_i = X_i - E[X_i]$

$E[\sum\limits_{i=1}^{n}(X_i - \bar{X})^2] = E[\sum\limits_{i=1}^{n}(Y_i - \bar{Y})^2]$

$$\begin{aligned} E[\sum_{i=1}^{n} Y_i^2 - 2\bar{Y}Y_i + \bar{Y}^2] &= \sum_{i=1}^{n}(E[Y_i^2] - 2E[Y_i \frac{1}{n}\sum_{j=1}^{n} Y_j] + E[Y^2]) \\ &= \sum_{i=1}^{n} \sigma^2 - \frac{2}{n}(E[Y_i^2] + \sum_{j=1, i \neq j}^{n} E[Y_i Y_j] + \frac{\sigma^2}{n}) \\ &= \sum_{i=1}^{n}(\sigma^2 - \frac{2}{n}\sigma^2 + \frac{\sigma^2}{n}) \\ &= \sum_{i=1}^{n}(1 - \frac{1}{n} \\ &= \sigma^2 n(\frac{n-1}{n} = (n-1)\sigma^2 \end{aligned}$$

Also $E[S^2] = \sigma^2$

Wie kann man Schätzer konstruieren?

B.4. Maximum-Likelihood-Methode und Momentmethode

Maximum-Likelihood-Methode

Idee: Wähle ein $\vartheta \in \Theta$, für das die Wahrscheinlichkeit, genau das zu beobachten , was tatsächlich beobachtet wurde, maximal ist.

1. Fall: Alle $P_\vartheta, \vartheta \in \Theta$ diskret. Wähle ein ϑ, mit $L(\vartheta) := P_\vartheta(x_1) \cdot P_\vartheta(x_2) \ldots P_\vartheta(x_n)$ maximal. L heißt Likelihood-Funktion.

Bsp.: Münzwurf
$x_1, \ldots, x_n \in \{0,1\}$ beobachtet, (z.B. $0,1,1,1,0,1,0,0$)

Parameter $\underbrace{p}_{=\vartheta} \in \underbrace{[0,1]}_{=\Theta}$ unbekannt.

$L(p) = p^x(1-p)^{n-x}$ mit $x = \#\ 1$ in $(x_1, \ldots, x_n) = x_1 + \ldots + x_n$

(Hier $L(\vartheta) = p^4(1-p)^5$). Maximiere L.

Trick: Oft einfacher statt L, $\log L$ zu maximieren. (beachte log streng monoton) $\mathcal{L}$ heißt Log-Likelihood-Funktion.

$\mathcal{L}(p) = \log L(p) = x\ \log\ p + (n-x)\log(1-p)$

$\mathcal{L}'(p) = \frac{x}{p} - \frac{n-x}{1-p} \overset{!}{=} 0$ „soll Null sein"

$\frac{x}{p} \overset{!}{=} \frac{n-x}{1-p},\ (1-p)x \overset{!}{=} (n-x)p),\ p = \frac{x}{n} = \overline{x}$

($\overline{x} = \frac{4}{9} =$ ML-Schätzer für p).

Der ML-Schätzer $\hat{p}$ für p ist $\overline{X}$

2. Fall Alle P_ϑ haben eine Dichte $f_\vartheta : \mathbb{R} \to [0, \infty)$

Bsp.: $x_1 = 1.3,\ x_2 = 2.4,\ x_3 = 0.9$ mit $X_1, X_2, X_3 \sim Exp(\alpha)$ Lebensdauer

In diesem Fall maximiere $L(\vartheta) := f_\vartheta(x_1) \cdot \ldots \cdot f_\vartheta(x_n)$

Statistisches Modell: $\mathcal{P} = \{Expo(\alpha) : a > 0\}$

$$\begin{aligned} L(\alpha) &= \alpha \cdot e^{-\alpha\ x_1} \cdot \ldots \cdot \alpha e^{-\alpha\ x_n} \\ &= \alpha^n \cdot e^{-\alpha\ x} \text{ mit } x = x_1 + \ldots + x_n \end{aligned}$$

$\mathcal{L}(\alpha) = \log L(\alpha) = n \ \log \ \alpha - \alpha \ x$

$\mathcal{L}'(\alpha) = \frac{n}{\alpha} - x \stackrel{!}{=} 0$

$\alpha \stackrel{!}{=} \frac{n}{x} = \frac{1}{\overline{x}}, \hat{\alpha} = \frac{1}{\overline{x}}$ ist ML-Schätzer für α.

$\hat{\alpha} = \frac{3}{4{,}6}$

Momentmethode
Wähle ein $\vartheta \in \Theta$ aus, für das der „empirische Mittelwert $\overline{X}$" gleich dem theoretischen Mittelwert, d.h. dem Erwartungswert $E[X_1]$ ist.

Bsp.: $X_1, \ldots, X_n \sim Exp(\alpha)$ unabh. wie oben, so soll $\overline{X} \stackrel{!}{=} E[X_1] = \frac{1}{\alpha}$
Also $\alpha = \frac{1}{\overline{X}}$

$\hat{\alpha} = \frac{1}{\overline{x}}$ im Beispiel $\hat{\alpha} = \frac{3}{4{,}6}$, d.h. ML-Methode und Momentmethode liefer hier den selben Schätzer.

Rückblick:

Gegeben seien iid ZVen $X_1, \ldots, X_n \ \Omega \to \mathbb{R}$ mit Verteilung $P_\vartheta, \vartheta \in \Theta$.

Betrachten Stichprobe $(x_1, \ldots, x_n) = (X_1(\omega), \ldots, X_n(\omega)), \omega \in \Omega$, vom Umfang n.

Likelihood-Funktion $L : \Theta \to [0, \infty)$

i.) $L(\vartheta) = P_\vartheta(x_1) \cdot \ldots \cdot P_\vartheta(x_n)$ im diskreten Fall

ii.) $L(\vartheta) = f_\vartheta(x_1) \cdot \ldots \cdot f_\vartheta(x_n)$ falls f_ϑ eine Dichte der X_i ist.

Dann ist $\hat{\vartheta}$ so gewählt, dass

$L(\hat{\vartheta}) = \max\limits_{\vartheta \in \Theta} L(\vartheta)$ gilt, der Maximum-Likelihood-Schätzer für ϑ.

Beachte: $\hat{\vartheta}$ hängt von der Stichprobe $(x_1, \ldots, x_n)$ respektive von $\omega \in \Omega$ ab, ist also eine ZV.

Schätzer der Momentenmethode:

$\hat{\vartheta}$ sei so gewählt, dass $\overline{x} = \overline{x}(\omega) == E_\vartheta(X_1)$ gilt.

Bei der Exponentialverteilung (vgl. oben) und der geometrischen Verteilung (vgl. Übung) stimmen die beiden Schätzer überein. Dies gilt im Allgemeinen jedoch nicht.

Bsp.: Betrachten ZV X_1 mit Dichte $f_\theta(x) = \begin{cases} 0 & , x < 0 \\ \vartheta\ x\ e^{-\frac{\vartheta}{2}\ x^2} & , x \geq 0 \end{cases}$ für $\vartheta > 0$.

- f_ϑ ist eine Dichte:

  $$\int\limits_0^\infty \vartheta x e^{-\frac{\vartheta}{2}x^2}\ dx = \Big[- e^{-\frac{\vartheta}{2}x^2} \Big]_0^\infty = -0 + 1 = 1$$

- Erwartungswert:

  $$E_\vartheta[X_1] = \int_0^\infty \vartheta x^2 e^{-\frac{\vartheta}{2}x^2}\ dx = \tfrac{1}{2} \int\limits_{-\infty}^{\infty} \vartheta\ x^2\ e^{-\frac{\vartheta}{2}\ x^2}\ dx$$

  mit $\sigma := \frac{1}{\sqrt{\vartheta}}$

  $$E_\vartheta[X_1] = \tfrac{\sqrt{2\pi\sigma^2}}{\sigma^2 \cdot 2} \underbrace{\int\limits_{-\infty}^{\infty} x^2 \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\ dx}_{=Var(Y)=\sigma^2 \text{ für } Y \overset{d}{=} \mathcal{N}(0,\sigma^2)}$$

- Momenten-Methode: $\overline{X} \overset{!}{=} E_{\hat{\vartheta}}[X_1] = \sqrt{\frac{\pi}{2\hat{\vartheta}}}$

  Also: $\hat{\vartheta} = \frac{\pi}{2\overline{X}^2} \overset{n=1}{=} \frac{\pi}{2X_1^2}$

  ML-Schätzer für $n = 1$:

  Log-Likelihood-Funktion $\log L(\vartheta) = \log f_\vartheta$

  $\log L(\hat{\vartheta})' = (\log \hat{\vartheta} + \log X_1 - \frac{\hat{\vartheta}}{2} X_1^2)' = \frac{1}{\vartheta} - \frac{X_1^2}{2} \overset{!}{=} 0$

  Also: $\hat{\vartheta} = \frac{2}{X_1^2} \neq \frac{n}{2x_1^2}$

B.5. Konfidenzintervalle normalverteilter ZVen

Seien $\underline{X} = (X_1, \ldots, X_n)$ eine Stichprobe zum statistischen Modell $\mathcal{P} = \{P_\vartheta, \vartheta \in \Theta\}$
Oft will man ϑ gar nicht konkret schätzen, sonder ein Intervall angeben, in welchem ϑ mit hoher W.keit liegt.

Definition: Gegeben sei eine Konfidenzzahl (Vertrauenszahl) $\gamma \in (0,1)$.

Ein Intervall $[U_n, O_n]$ heißt γ-Konfidenzintervall für ϑ, wenn U_n, O_n zwei Schätzer sind, so dass $P_\vartheta(U_n \leq \vartheta \leq O_n) \geq \gamma$ gilt $\forall\ \vartheta \in \Theta$.

(Wenn die Verteilung P_ϑ die „richtige“ ist, dann liegt ϑ mit „hoher“ W.keit im Intervall $[U_n, O_n]$).

Für den Rest des Kapitels sollen die X_i normalverteilt sein.

Wir beträchten zunächst den Fall, in dem die Varianz $\sigma^2 > 0$ vorgegeben ist. (Bsp.weise bei Produktionsprozessen mit variabler Einstellung und bekannter und konstanter Toleranz).

Das statistische Modell ist also $\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}\}$

Zu gegebenem γ bestimmt man das $\frac{1+\gamma}{2}$-Quantil, d.h. die Zahl z für die gilt:

$\Phi(z) = \frac{1+\gamma}{2}$, wobei $\Phi(z) := \int\limits_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\ dx$

Man setzt $U_n := \overline{X} - z\frac{\sigma}{\sqrt{n}}, O_n := \overline{X} + z\frac{\sigma}{\sqrt{n}}$.

Dann gilt: $P_\mu(U_n \leq \mu \leq O_n) = P_\mu(|\overline{X} - \mu| \leq z\frac{\sigma}{\sqrt{n}} = P_\mu(\underbrace{\sqrt{n}\frac{|\overline{X} - \mu|}{\sigma}}_{\text{Standardisierte von } \overline{X}} \leq z) =$

$\Phi(z) - \Phi(-z) = 2\Phi(z) - 1 = \gamma$

Also: $[U_n, O_n]$ ist ein γ-Konfidenz-intervall für den Erwartungswert μ.

Betrachten nun den Fall, bei dem σ^2 nicht bekannt ist.

Definition: Seien $Y_1, \ldots, Y_n$ unabhängige, $\mathcal{N}(0,1)$-verteilte ZVen. Dann heißt die Verteilung von $Y := Y_1^2 + \ldots + Y_n^2$ die Chi-Quadrat-Verteilung mit n Freiheitsgraden. (Notation: χ_n^2)

Durch vollständige Induktion nach n kann gezeigt werden, dass Y eine Dichte g_n besitzt, gegeben durch

$$g_n(x) := \begin{cases} \frac{x^{\frac{n}{2}} e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} & , x \geq 0 \\ 0 & , x < 0 \end{cases},$$

wobei $\Gamma(\frac{n}{2}) = \int\limits_0^\infty x^{\frac{n}{2}} e^{-x}\ dx$ die Gamma-Funktion ist.

Für $m := \frac{n}{2}, \alpha := \frac{1}{2}$ gilt:

$g_n(x) = \frac{\alpha^m x^{m-1} e^{-\alpha x}}{\Gamma(m)}, x \geq 0$

g_n ist also die Dichte einer „erweiterten“ Gamma-Verteilung (vgl. Kap. 21).

Definition: Seien $Y \stackrel{d}{=} \chi_n^2$ und $z \stackrel{d}{=} \mathcal{N}(0,1)$ unabhängig. Dann heißt die Verteilung von

$T := \sqrt{n}\frac{z}{\sqrt{Y}}$ die <u>Students-t-Verteilung</u> mit n Freiheitsgraden (Notation: t_n).

Man kann zeigen, dass $f_n(x) := \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\frac{x^2}{n})^{-\frac{n+1}{2}}, x \in \mathbb{R}$, eine Dichte von T ist.

Beachte: Da $f_n(-x) = f_n(x)$ ist für alle $x \in \mathbb{R}$ gilt auch

$P(|T| \leq t) \leq F_n(t) - F_n(-t) = 2F_n(t) - 1, t \geq 0,$

wobei $F_n(t) = \int\limits_{-\infty}^{t} f_n(x)\ dx$

Wir erzeugen die Varianz σ^2 nun durch ihren erwartungstreuen Schätzer

$S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$

und setzen nun $U_n := \overline{X} - t\frac{S}{\sqrt{n}}, O_n := \overline{X} + t\frac{S}{\sqrt{n}},$

wobei t das $\frac{1+\gamma}{2}$-Quantil mit $n-1$ Freiheitsgraden der t-Verteilung sein soll. Es muss also gelten $F_n(t) = \frac{1+\gamma}{2}$ wobei F_{n-1} die Verteilungsfunktion der t-Verteilung mit $n-1$ Freiheitsgraden ist.

Es gilt Folgendes:

Die Standardisierte $Z := \frac{\overline{X}-\mu}{\sigma}\sqrt{n}$ von $\overline{X}$ ist $\mathcal{N}(0,1)$.

Man kann zeigen, dass

$Y := \frac{n-1}{\sigma^2}s^2$

χ^2_{n-1}-verteilt und unabhängig ist von Z.

Also: $T := \sqrt{n-1}\frac{z}{\sqrt{y}} = \frac{\sqrt{n}}{s}(\overline{X} - \mu)$ ist t-verteilt mit $n-1$ Freiheitsgraden (nach Definition). Es gilt dann:

$P(U_n \leq \mu \leq O_n) = P(|\overline{X} - \mu| \leq t\frac{S}{\sqrt{n}}) = P(|T| \leq t) = F_{n-1}(t) - F_{n-1}(-t) = \gamma$

Also: $[U_n, O_n]$ ist ein γ-Konfidenzintervall für den Mittelwert μ.

B.6. Testen von Hypothesen, Binomialtest

Durch n-maliges Werfen einer Münze soll getestet werden, ob die Münze fair ist (d.h. ob $P(\text{Kopf}) = P(\text{Zahl})$).

Nullhypothese	$H_0 : P \stackrel{d}{=} B(1, \frac{1}{2})$
Alternativhypothese	$H_A : P \in \{B(1,p) : p \neq \frac{1}{2}\}$

Allgemeiner Ansatz:
Gegeben sei ein statistisches Modell $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$

Weiter sei Θ aufgeteilt in $\Theta_{H_0} \subset \Theta$ und $\Theta_{H_A} := \Theta$
$\Theta_{H_0} = \Theta_{H_0}^c \subset \Theta$

Nullhypothese	$H_0 : P \in \{P_\vartheta : \vartheta \in \Theta_{H_0}\}$
Alternativhypothese	$H_A : P \in \{P_\vartheta : \vartheta \in \Theta_{H_A}\}$

Wie soll entschieden werden, ob die Nullhypothese verworfen werden kann/darf?

Es sei $X_i = 1$ bei „Kopf" und $X_i = 0$ bei „Zahl" des i-ten Wurfes für $i = 1, \dots, n$.

Weiter sei $T(\underline{x}) := \sum\limits_{i=1}^{n} x_i$ für $\underline{x} = (x_1, \dots, x_n) \in \{0,1\}^n$

Dann ist $T(\underline{X}) = T_0(X_1, \dots, X_n)$ die Anzahl der „Kopf-Würfe".

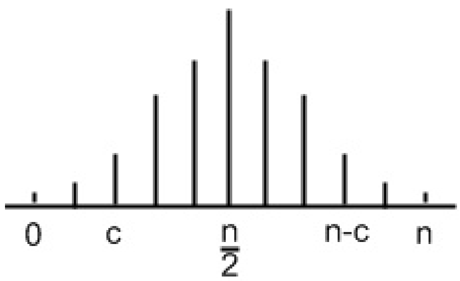


Abbildung 10: Es gilt $T(\underline{X}) \stackrel{d}{=} B(n,p)$ mit $p = P(x_i = 1)$

$E(T(\underline{X})) = \frac{n}{2}$ für $p = \frac{1}{2}$

Wählen $c \leq \frac{n}{2}$ und sagen die Münze sei gezinkt, wenn $T(\underline{x}) \leq c$ oder $T(\underline{x}) \geq n - c$

Definition:

Ein Test der Nullhypothese H_0 mit Verwerfungsbereich $V \subset \mathbb{R}$ ist eine Abbildung $T : \mathbb{R}^n \to \mathbb{R}$ mit folgender Interpretation:

- H_0 wird verworfen für $T(\underline{x}) \in V$
- H_0 wird nicht verworfen für $T(\underline{x}) \in V$

Beim Münzwurf ist $V = [0, c] \cup [n-c, n]$, wobei $c > 0$ noch genauer bestimmt werden soll.

Fehler 1. Art:
H_0 liegt vor, wird aber verworfen
(Die Münze wird als gezinkt betrachtet, obwohl sie fair ist).

Fehler 2. Art:
H_0 liegt nicht vor, wird aber auch nicht verworfen
(Die Münze ist gezinkt, man kann es aber nicht „beweisen").

Das Ziel ist meistens einen Test zu finden, für den die W.keit für den Fehler 1. Art ein vorgegebenes Signifikanzniveau α nicht überschreitet, die W.keit für einen Fehler 2. Art aber möglichst gering hält. Dabei wird $\alpha > 0$ klein gewählt („Im Zweifelsfalle für den Angeklagten").

Definition:
Die Abbildung $\Pi_T : \Theta \to [0, 1]$, definiert durch $\Pi_T(\vartheta) := P_\vartheta(T(\underline{X}) \in V)$, heißt Gütefunktion des Test T mit Verwerfungsbereich V.
Die Zahl $\sup\limits_{\vartheta \in \Theta_{H_0}} \Pi_T(\vartheta)$ heißt Niveau des Tests.

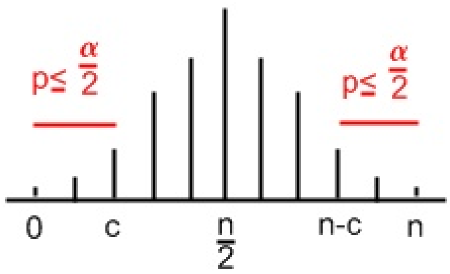
Das Niveau eines Tests soll also ein vorgegebenes Signifikanzniveau α nicht überschreiten.

Beim Münzwurf: Wie ist c zu wählen?

$$\begin{aligned}
\Pi_T(p) &= P_p(\sum_{i=0}^{n} x_i \leq c \text{ oder } \sum_{i=1}^{n} x_i \geq n - c) \\
&= \sum_{i=0}^{c} \binom{n}{i} p^i (1-p)^{n-1} + \sum_{i=n-c}^{n} \binom{n}{i} p^i (1-p)^{n-i}
\end{aligned}$$

Niveau $\Pi_T(\frac{1}{2}) = \frac{1}{2^{n-1}} \sum\limits_{i=0}^{c} \binom{n}{i}$

Es ist also (durch ausprobieren) c möglichst groß zu wählen, so dass aber noch gilt $\Pi(\frac{1}{2}) \leq \alpha$.



Umgekehrt existiert für eine Stichprobe $\underline{x}$ ein kleinstes Signifikanzniveau $\hat{\alpha}(\underline{x})$ für welches die Nullhypothese gerade noch abgelehnt werden darf.

Beim Münzwurfbsp.: $\hat{\alpha}(\underline{x}) = \frac{1}{2^{k-1}} \sum\limits_{i=0}^{\min(k,n-k)} \binom{n}{i}$

$k := x_1 + \ldots + x_n = T(\underline{x})$

H_0 wird genau dann verworfen, wenn $\hat{\alpha}(x) \leq \alpha$ gilt.

$\hat{\alpha}(\underline{x})$ heißt $\underline{p\text{-Wert}}$ von $\underline{x}$, und gibt an, wie groß die W.keit eines Fehlers 1. Art ist, wenn H_0 auf Grund von $\underline{x}$ verworfen wird.

Zweiseitiger Binomialtest

Statistisches Modell $\mathcal{P} = \{B(1,p) : p \in [0,1]\}$

$H_0 : P \stackrel{d}{=} B(1,p_0), p_0 \in [0,1]$ fest
$H_A : P \in \{B(1,p) : p \neq p_0\}$

Verwerfungsbereich $V := [0, c_1] \cup [n - c_2, n]$

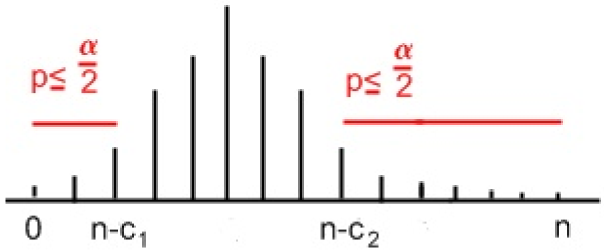
Also, verwerfe H_0, wenn $T(\underline{x}) \leq c_1$ oder $T(\underline{x}) \geq n - c_2$

$$\begin{aligned}\Pi_T(p) &= P_p(\sum_{i=0}^{n} x_i \leq c_1 \text{ oder } \sum_{i=1}^{n} x_i \geq n - c_2) \\ &= \sum_{i=0^c_1} \binom{n}{i} p^i (1-p)^{n-i} + \sum_{i=n-c_2}^{n} \binom{n}{i} p^i (1-p)^{n-i}\end{aligned}$$

Setze c_1, c_2 möglichst groß an, so dass aber noch gilt

$\sum\limits_{i=0}^{c_1} \binom{n}{i} p_0^i (1-p_0)^{n-1} \leq \frac{\alpha}{2}$

$\sum\limits_{i=n-c_2}^{n} \binom{n}{i} p_0^i (1-p_0)^{n-i} \leq \frac{\alpha}{2}$



Einseitiger Binomialtest:

Statistisches Modell $\mathcal{P} = \{B(1,p) : p \geq p_0\}, p_0$ fest

$H_0 : P \stackrel{d}{=} B(1,p_0)$
$H_A : P \in \{B(1,p) : p > p_0\}$

Verwerfungsbereich $V = [n-c, n]$, wobei c möglichst groß zu wählen ist, aber so dass noch gilt $P_{p_0}(T(\underline{X}) \geq n-c) \leq \alpha$

Gütefunktion $\Pi_T(p) = \sum\limits_{i=n-c}^{n} \binom{n}{i} p^i (1-p)^{n-i}$
Niveau des Test $\Pi_T(p_0)$

Näherung für Signifikanzniveau $\alpha = 5\%$ und große n (Faustregel $n\ p_0\ q_0 > 9,\ q_0 := 1-p_0)$:

Dann ist der Verwerfungsbereich V ungefähr gegeben durch

$V^c = [n\ p_0 - 2\sqrt{n\ p_0\ q_0}, n\ p_0 + \sqrt{n\ p_0\ q_0}]$

beim zweiseitigen Test mit $H_0 : p = p_0$.

Warum?

Setzen $Z := \frac{T - n\ p_0}{\sqrt{n\ p_0\ q_0}} \approx \mathcal{N}(0,1)$ nach zentralem Grenzwertsatz.

Also gilt:

$P_{p_0}(n\ p_0 - 1,96\sqrt{n\ p_0\ q_0} \leq T(\underline{X}) \leq n\ p_0 + 1,96\sqrt{n\ p_0\ q_0})$

$= P_{p_0}(-1,96 \leq Z \leq 1,96) \approx 0,95$ (mit Hilfe der Tabelle)

Außerdem: $1,96 \approx 2$

Analog für den einseitigen Binomialtest:

$H_0 : p = p_0$
$H_A : p > p_0$

$V \approx [n\ p_0 + 1,64\sqrt{n\ p_0\ q_0}, n]$

B.7. Weitere Tests

B.7.1. z-Test:

Statistisches Modell $\mathcal{P} := \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$

zweiseitig: $H_0 : \mu = \mu_0$
$H_A : \mu \neq \mu_0$

Test $T(\underline{x}) := \overline{x}$ arithmetisches Mittel

Verwerfungsbereich $V := (-\infty, \mu_0 - c] \cup [\mu_0 + c, \infty)$

Dabei ist $c > 0$ möglichst klein zu wählen, aber so dass noch gilt:

$P_{\mu_0}(\overline{X} \in V) \leq \alpha$

Also:

$P_{\mu_0}(\overline{X} \in V)$
$= P_{\mu_0}(\overline{X} \leq \mu_0 - c) + P_{\mu_0}(\overline{X} \geq \mu_0 + c)$

$= P_{\mu_0}\left(\frac{\overline{X}-\mu_0}{\sigma}\sqrt{n} \leq \frac{-c}{\sigma}\sqrt{n}\right) + P_{\mu_0}\left(\frac{\overline{X}-\mu_0}{\sigma}\sqrt{n} \geq \frac{c\sqrt{n}}{\sigma}\right)$

$= \Phi(\frac{-c}{\sigma}\sqrt{n}) + 1 - \Phi(\frac{n}{\sigma}\sqrt{n})$

$= 2 - 2\Phi(\frac{c}{\sigma}\sqrt{n}) \stackrel{!}{=} \alpha$

Somit: $c = z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

wobei $z_{1-\frac{\alpha}{2}}$ das $(1 - \frac{\alpha}{2})$-quantil von $\mathcal{N}(0,1)$ sein soll (also $\Phi(z_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$)

Vergleich Konfidenzintervalle

Wir haben gesehen

$P_{\mu_0}[\overline{X} - z\frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \overline{X} + z\frac{\sigma}{\sqrt{n}}] = \gamma$

für $z = z_{\frac{1+\gamma}{2}}$, also $\Phi(z) = \frac{1+\gamma}{2}$

Dies stimmt für $\gamma := 1 - \alpha$ und $c := z\frac{\sigma}{\sqrt{n}}$ mit obigem Resultat überein (nachprüfen).

einseitig: $H_0 : \mu = \mu_0$
$H_A : \mu > \mu_0$

Verwerfungsbereich für T ist $V = [\mu_0 + c, \infty)$ mit $c = z_{1-\alpha}\frac{\sigma}{\sqrt{n}}$,
$z_{1-\alpha}(1-\alpha)$-quantil von $\mathcal{N}(0,1)$

Nachtrag:

$X_1, \ldots, X_n \stackrel{d}{=} \mathcal{N}(\mu, \sigma^2)$ unabhängig, $\mu = E(X)$

$\overline{X} = \frac{X_1 + .. + X_n}{n}$ arithmetisches Mittel

Dann gilt: Standardisierte $\frac{\overline{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \stackrel{d}{=} \mathcal{N}(0,1)$

Satz: Seien $X_1 \stackrel{d}{=} \mathcal{N}(\mu_1, \sigma_1^2), \ldots, X_n = \mathcal{N}(\mu_n, \sigma_n^2)$ unabhängig. Dann gilt

$X_1 + \ldots + X_n \stackrel{d}{=} \mathcal{N}(\mu_1 + \ldots + \mu_n, \sigma_1^2 + \ldots + \sigma_n^2)$

Beweis: Vollständige Induktion nach $n \in \mathbb{N}$

Ind. Verankerung $n = 1 :$ $\surd$
Ind. schritt $n \to n+1 :$
Ind.annahme: Es gelte $X_1 + \ldots + X_n \stackrel{d}{=} \mathcal{N}(\mu, \sigma^2)$, wobei
$\mu := \mu_1 + \ldots + \mu_n, \sigma^2 := \sigma_1^2 + \ldots + \sigma_n^2$

$X_1 + \ldots + X_n$ hat also eine Dichte φ_{μ,σ^2}, und x_{n+1} hat eine Dichte $\varphi_{\mu_{n+1},\sigma_{n+1}^2}$, wobei

$\varphi_{EW,Var}(x) = \frac{1}{\sqrt{2\pi Var}} e^{-\frac{(x-EW)^2}{2\ Var}}$

Außerdem sind $X_1 + \ldots + X_n$ und X_{n+1} unabhängig (anschaulich klar).

Faltungsformel für unabhängige ZVen mit Dichte: $(X_1 + \ldots + X_n) + X_{n+1}$ hat eine Dichte

$g(y) = \int\limits_{-\infty}^{\infty} \varphi_{\mu,\sigma^2}(x)\varphi_{\mu_{n+1},\sigma_{n+1}^2}(y-x)dx, y \in \mathbb{R}$

Es gilt $\varphi_{\mu,\sigma^2}(x)\varphi_{\mu_{n+1},\sigma_{n+1}}(y-x) = \underbrace{\varphi_{\mu+\mu_{n+1},\sigma^2+\sigma^2_{n+1}}(y)}_{\text{gesuchter Faktor}} \cdot \underbrace{\varphi_{\tilde{\mu},\tilde{\sigma}^2}(x)}_{\text{ergibt sich}}$,

wobei $\tilde{\mu} = \frac{\mu\sigma^2_{n+1}+(\mu_{n+1}-y)\sigma^2}{\sigma^2+\sigma^2_{n+1}}$

$\tilde{\sigma} = \frac{\sigma^2\cdot\sigma^2_{n+1}}{\sigma^2+\sigma^2_{n+1}}$

Somit $g(y) = \varphi_{\mu+\mu_{n+1},\sigma^2+\sigma^2_{n+1}}(y) \cdot \underbrace{\int\limits_{-\infty}^{\infty} \varphi_{\tilde{m}u,\tilde{\sigma}^2}(x)\ dx}_{=1}$

Also: $X_1 + \ldots + X_{n+1}$ ist $\mathcal{N}(\mu + \mu_{n+1}, \sigma^2 + \sigma^2_{n+1})$
$= \mu_1 + \ldots + \mu_{n+1}, \sigma_1^2 + \ldots + \sigma^2_{n+1}\square$

B.7.2. t-Test

Beim t-Test wird im Gegensatz zum z-Test nicht angenommen, dass die Varianz σ^2 bekannt ist.

zweiseitig: statistisches Modell $P = \{\mathcal{N}(\mu,\sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$

$H_0 : P \in \{\mathcal{N}(\mu_0,\sigma^2) : \sigma^2 > 0\}, \mu_0 \in \mathbb{R}$ fest
$H_A : P \in \{\mathcal{N}(\mu,\sigma^2) : \mu \neq \mu_0, \sigma^2 > 0\}$

Verwerfungsbereich für $\overline{X} : V := (-\infty, \mu_0 - c] \cup [\mu_0 + c, \infty)$

Wie ist c zu wählen für ein vorgegebenes Signifikanzniveau α?

Test $T(\underline{X}) := \sqrt{n}\frac{\overline{X}-\mu_0}{\sqrt{S^2}}$

$S^2 = \frac{1}{n-1}((X_1 - \overline{X})^2 + \ldots + (X_n - \overline{X})^2)$

Betrachten $T(\underline{X}) = \frac{(\overline{X}-\mu_0)}{\sqrt{\frac{\sigma^2}{n}}}\sqrt{n-1}\frac{1}{\sqrt{Y}}$ mit

$Y = \frac{n-1}{\sigma^2}S^2, \sigma^2$ die „echte“Varianz

Satz: Unter H_0 gilt: $Z \stackrel{d}{=} \mathcal{N}(0,1), Y \stackrel{d}{=} X^2_{n-1}$, Z und Y sind unabhängig, d.h. T ist t-verteilt mit $n-1$ Freiheitsgraden.

$$\begin{aligned}\text{Durch} P_{\mu_0}(\overline{X} \in V) &= P_{\mu_0}(|T(\underline{X})| \leq c\frac{n}{S}) \\ &= 2 - 2P_{\mu_0}(T(\underline{X}) \geq c\frac{\sqrt{n}}{S}) \\ &\overset{!}{=} \alpha\end{aligned}$$

erhalten wir $c := \frac{S}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}}$

$t_{n-1,1-\frac{\alpha}{2}}$ ist das $(1-\frac{\alpha}{2})$-quantil der t-Verteilung mit $n-1$ Freiheitsgraden.

Aufgabe: Überprüfen Sie, dass dies äquivalent ist zu den Resultaten beim Konfidenzintervall.

B.7.3. Chi-Quadrat-Anpassungstest

Der Ausgang eines Zufallsexperiments sei aufgeteilt in k Klassen.

H_0: Die k Klassen treffen mit W.keiten $p_1, \ldots, p_k$ ein (wobei $p_1 + \ldots + p_k = 1$)
H_A: H_0 stimmt nicht

Seien n_i : Anzahl „Treffer" i-te Klasse
$n := n_1 + \ldots + n_k$ Umfang der Stichprobe

Test $T := \sum\limits_{i=1}^{k} \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$

Dabei ist $n \cdot p_i$ die erwartete Anzahl „Treffer" der i-ten Klasse.

Verwerfen H_0 wenn T „zu groß" ist.

Satz: T konvergiert für $n \to \infty$ gegen χ^2_{k-1}-Verteilung.

Eine genauere Analyse ergibt folgende Faustregel für $\alpha = 5\%$:

Gilt $n \cdot p_i \geq 1$ für alle $i \in \{1, \ldots, k\}$ und $n \cdot p_i \geq 4$ für mind. 80% der $i \in \{1, \ldots, k\}$,so ist H_0 zu verwerfen, falls $T \geq \nu + 2\sqrt{2\nu}$ ist, wobei $\nu := k - 1$.

Falls die Bedingungen nicht erfült sind, fasst man mehrere Klassen zu einer zusammen (wodurch k kleiner wird).

B.7.4. Chi-Quadrat-Unabhängigkeitstest:

Der Ausgang eines Zufallsexperiments wird nach klassifizierten Merkmalen untersucht. (Bsp.: Haarfarbe und Augenfarbe eines zufällig gewählten Menschen)

H_0: Die beiden Merkmale sind unabhängig voneinander
H_A: H_0 stimmt nicht

Seien r: Anzahl Klassen 1. Merkmal
s: Anzahl Klassen 2. Merkmal
$n_{i,j}$:Anzahl „Treffer“ i-te Klasse 1. Merkmal und j-te Klassen 2. Merkmal

Gegeben seien $p_i^1 := P$(i-te Klasse 1.Merkmal)
$p_j^2 := P$(j-te Klasse 2.Merkmal)

z.B. ML-Schätzer:
n_i^1: Anzahl Treffer i-te Klasse 1.Merkmal
n_j^2: Anzahl Treffer j-te Klasse 2.Merkmal

$n := n_1^1 + \ldots + n_r^1 = n_1^2 + \ldots + n_s^2$

$\hat{p}_i^1 = \frac{n_i^1}{n}, \hat{p}_i^2 = \frac{n_j^2}{n}$

Unter H_0 gilt $p_{ij} := P(1.\text{Merkmal i-te Klasse}, 2.\text{Merkmal j-te Klasse})$
$= p_i^1 \cdot p_j^2$

χ^2-Anpassungstest: $T = \sum\limits_{i=1}^{r} \sum\limits_{j=1}^{s} \frac{(n_{ij} - \overbrace{np_i^1 \cdot p_j^2}^{=np_{ij}})^2}{np_i^1 \cdot np_j^2}$

z.B. ML-Schätzer $T = n \cdot \sum\limits_{i=1}^{r} \sum\limits_{j=1}^{s} \frac{(n_{ij} - \frac{n_i^1 n_j^2}{n})^2}{n_i^1 \ldots n_j^2} = n \sum\limits_{i=1}^{r} \sum\limits_{j=1}^{s} \frac{n_{ij}^2}{n_i^1 n_j^2} - n$

Faustregel: Verwerfe H_0 falls $T \geq \nu + 2\sqrt{2\nu}$, wobei
$\nu = (r-1)(s-1) = r \cdot s + 1 - (r-1) - (s-1)$

B.7.5. Vorzeichentest:

Sei X eine ZV mit Dichte f und Verteilungsfunktion F.

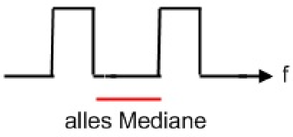
Ein Median von X, bzw. von F, ist eine Zahl $\mu \in \mathbb{R}$ mit $F(\mu) = P(X \leq \mu) = \frac{1}{2}$

- Ist μ ein Median von X, so gilt

  $P(X \leq \mu) = P(X \geq \mu) = \frac{1}{2}$, da $P(X = \mu) = 0$

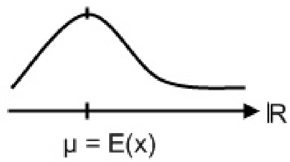
- Ein Median muss nicht eindeutig sein.

  Bsp.:



- Existiert $E(X)$ und ist f symm. bzgl. $E(X)$, dann ist $E(X)$ ein Median.

  Bsp.: $\mathcal{N}(\mu, \sigma^2)$



H_0 : μ_0 ist ein Median
H_A : nicht H_0

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

Test $T(\underline{X}) := \#\{i \in \{1, \ldots, n\}; X_i > \mu_0\}$

Unter H_0 gilt $T(\underline{X}) \stackrel{d}{=} B(n, \frac{1}{2})$

Weiter mit Binomialtest

B.7.6. Wilcoxon-Test:

Voraussetzung: f ist symm. bzgl. eines Parameters μ (μ ist dann automatisch ein Median und $\mu = E(X)$ falls existent)

$H_0 : \mu = \mu_0$
$H_A : \mu \neq \mu_0$

Stichprobe $\underline{x} = (x_1, \ldots, x_n)$

Ordnen $|\mu_0 - x_i|$ der Größe nach (aufsteigend) und weisen i den entsprechenden Rang zu

$T(\underline{x})$ = Summe der Ränge derjenigen $i \in \{1, \ldots, n\}$ für die gilt $x_i - \mu_0 > 0$

Bsp.: $n = 5, \mu_0 = 4$

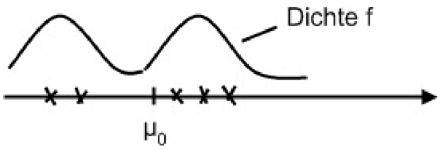
x_i	3,2	4,8	5,6	4,4	6,2
$\|x_i - \mu_0\|$	1,8	0,2	0,6	0,6	1,2
Vorzeichen	-	-	+	-	+
Rang	5	1	2,5	2,5	4

$T(\underline{x}) = 6,5$

T nimmt Werte in $\{0, \ldots, \frac{n(n+1)}{2}\}$

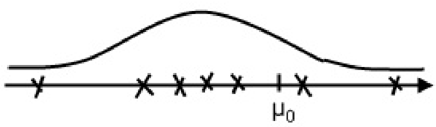
Man verwirft H_0 falls T zu „klein“ oder zu „groß“ ist.(siehe Tabelle)

Vergleiche Vorzeichentest:



Bei einer solchen Dichte lässt sich H_0 mit dem Wilcoxon-Test eher verwerfen (vgl. Übung 46)

Vergleiche mit t-Test:



S muss klein sein beim t-Test um H_0 zu verwerfen. Bei Ausreißern lässt sich H_0 mit dem Wilcoxon-Test eher verwerfen (vgl. Übung 45).

Klausurhinweise

Klausur:	Datum:	18.07.
	Zeit:	13.15 Uhr (pünktlich da sein)
	Dauer:	90 Minuten
	Raum:	N6 Hörsaalzentrum
	Hilfsmittel:	zwei A4-Blätter (handbeschrieben, beidseitig), **kein** Taschenrechner

Klausurteilnahme (oder offizielle Entschuldigung) ist Voraussetzung für die Zulassung zur Nachklausur.

Sprechstunde: Di, 14-17 Uhr, C-Bau A17