



Inferenzstatistik

BSc Psychologie SoSe 2021

Prof. Dr. Dirk Ostwald

(0) Mathematische Grundlagen

Als bekannt vorausgesetzte Symbole

$A := B$	A ist definiert als B	$A = B$	A ist gleich B
$A \Leftrightarrow B$	A ist äquivalent mit B	$A \Rightarrow B$	Aus A folgt B
$a < b$	a ist kleiner als b	$a \leq b$	a ist kleiner oder gleich b
$a > b$	a ist größer als b	$a \geq b$	a ist größer oder gleich b
$n!$	Fakultät von n	$ a $	Betrag von a

Definition. Eine Definition (oder auch ein *Axiom*) ist ein Grundsatz eines logischen Systems, der innerhalb dieses Systems weder begründet noch deduktiv abgeleitet wird. Definitionen können nur nach ihrer Nützlichkeit innerhalb eines logischen Systems bewertet werden. Eine Definition lernt man am besten erst einmal auswendig und hinterfragt sie erst dann, wenn man ihren Nutzen in der Anwendung verstanden hat oder von diesem nicht überzeugt ist. Etwas Entspannung beim Umgang mit Definitionen ist generell hilfreich.

Theorem. Ein Theorem (oder auch *Satz*) ist eine Aussage, die mittels eines Beweises als richtig erkannt, das heißt, aus Definitionen und/oder bereits bekannten Sätzen hergeleitet werden kann. Theoreme sind die “empirischen Ergebnisse” der Mathematik. Theoreme sind in der angewandten Mathematik oft für Berechnungen hilfreich, es lohnt sich also, sie auswendig zu lernen.

Beweis. Ein Beweis ist eine logische Argumentationskette, die auf bekannte Definitionen und Theoreme zurückgreift, um die Richtigkeit eines Theorems zu belegen. Kurze Beweise tragen oft zum Verständnis eines Theorems bei, lange Beweise eher nicht. **Beweise sind die Antwort auf die Frage warum eine mathematische Aussage gilt (“Warum ist das so?”).** Beweise lernt man nicht auswendig. Wenn sie lang sind, ist es sinnvoller sie beim ersten Lesen zu übergehen, um sich nicht in Details zu verlieren.

Mathematische Grundlagen

- Mengen
- Summen, Produkte, Potenzen
- Funktionen
- Differentialrechnung
- Analytische Optimierung
- Integralrechnung
- Selbstkontrollfragen

Literaturhinweise

Die Lektüre von Bärwolff (2017) wird empfohlen. Speziell werden zum Thema Mengenlehre die Abschnitte 1.2 und 1.5, zum Thema Funktionen die Abschnitte 1.3, 2.1 und 2.3, zum Thema Differentialrechnung Abschnitt 2.6, zum Thema Analytische Optimierung Abschnitt 2.10, und zum Thema Integralrechnung Abschnitt 2.13 (ohne 2.13.3) empfohlen.

Mathematische Grundlagen

- **Mengen**
- Summen, Produkte, Potenzen
- Funktionen
- Differentialrechnung
- Analytische Optimierung
- Integralrechnung
- Selbstkontrollfragen

Definition (Mengen und Mengendefinition)

Nach Cantor (1895) ist eine *Menge* definiert als “eine Zusammenfassung M von bestimmten wohlunterschiedenen Objekten m unsere Anschauung oder unseres Denken (welche die Elemente der Menge genannt werden) zu einem Ganzen”. Wir schreiben

$$m \in M \text{ bzw. } m \notin M \quad (1)$$

um auszudrücken, dass m ein Element bzw. kein Element von M ist. Zur Definition von Mengen gibt es mindestens folgende Möglichkeiten:

- (1) Auflisten der Elemente in geschweiften Klammern, z.B. $M := \{1, 2, 3\}$.
- (2) Angabe der Eigenschaften der Elemente, z.B. $M := \{x \in \mathbb{N} | x < 4\}$.
- (3) Gleichsetzen mit einer anderen eindeutig definieren Menge, z.B. $M := \mathbb{N}_3$.

Bemerkungen

- (2) wird als “ $x \in \mathbb{N}$, für die gilt, dass $x < 4$ ist” gelesen.
- Die Bedeutung von $\mathbb{N}$ und $\mathbb{N}_3$ wird im Folgenden erläutert.
- Mengen sind *ungeordnet*, d.h. $\{1, 2, 3\} = \{1, 3, 2\} = \{2, 3, 1\}$ etc.

Definition (Teilmengen und Mengengleichheit)

- Eine Menge A heißt *Teilmenge* einer Menge B , wenn für jedes Element $a \in A$ gilt, dass auch $a \in B$. Ist A eine Teilmenge von B , so schreibt man

$$A \subseteq B \quad (2)$$

und nennt A *Untermenge* von B und B *Obermenge* von A .

- Eine Menge A heißt *echte Teilmenge* einer Menge B , wenn für jedes Element $a \in A$ gilt, dass auch $a \in B$, es aber zumindest ein Element $b \in B$ gibt, für das gilt $b \notin A$. Ist A eine echte Teilmenge von B , so schreibt man

$$A \subset B. \quad (3)$$

- Zwei Mengen A und B heißen *gleich*, wenn für jedes Element $a \in A$ gilt, dass auch $a \in B$, und wenn für jedes Element $b \in B$ gilt, dass auch $b \in A$. Sind die Mengen A und B gleich, so schreibt man

$$A = B. \quad (4)$$

Bemerkung

- Seien $A := \{1\}$, $B := \{1, 2\}$, $C := \{1, 2\}$. Dann gilt $A \subset B$, $B \subseteq C$, $C \subseteq B$ und $B = C$.

Definition (Kardinalität, leere Menge)

Die Anzahl der Elemente einer Menge M heißt *Kardinalität* und wird mit $|M|$ bezeichnet. Eine Menge mit Kardinalität 0 heißt *leere Menge* und wird mit $\emptyset$ bezeichnet.

Bemerkung

- Für $M := \{1, 2, 3\}$ gilt $|M| = 3$.

Definition (Potenzmenge)

Die Menge aller Teilmengen einer Menge M heißt *Potenzmenge von M* und wird mit $\mathcal{P}(M)$ bezeichnet. Die leere Untermenge von M und M selbst sind immer Elemente von $\mathcal{P}(M)$.

Bemerkungen

- Für $M := \{1, 2, 3\}$ gilt $\mathcal{P}(M) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$.
- Allgemein gilt $|M| = n \Rightarrow |\mathcal{P}(M)| = 2^n$.

Definition (Mengenoperationen)

M und N seien zwei Mengen.

- Die *Vereinigung von M und N* ist definiert als die Menge

$$M \cup N := \{x | x \in M \text{ oder } x \in N\}, \quad (5)$$

wobei *oder* im inklusiven Sinne als *und/oder* zu verstehen ist.

- Der *Durchschnitt von M und N* ist definiert als die Menge

$$M \cap N := \{x | x \in M \text{ und } x \in N\}. \quad (6)$$

- Die *Differenz von M und N* ist definiert als die Menge

$$M \setminus N := \{x | x \in M \text{ und } x \notin N\}. \quad (7)$$

- Die *symmetrische Differenz von M und N* ist definiert als die Menge

$$M \Delta N := \{x | x \in M \text{ oder } x \in N, \text{ aber } x \notin M \cap N\}, \quad (8)$$

wobei *oder* hier also im exklusiven Sinne zu verstehen ist.

Bemerkung

- Für $M := \{1, 2, 3\}$ und $N := \{2, 3, 4, 5\}$ gilt

$$M \cup N = \{1, 2, 3, 4, 5\}, M \cap N = \{2, 3\}, M \setminus N = \{1\}, N \setminus M = \{4, 5\}, M \Delta N = \{1, 4, 5\}.$$

Definition (Zahlenmengen)

Es bezeichnen

- $\mathbb{N} := \{1, 2, 3, \dots\}$ die *natürlichen Zahlen*,
- $\mathbb{N}_n := \{1, 2, 3, \dots, n\}$ die *natürlichen Zahlen der Ordnung n* ,
- $\mathbb{N}^0 := \mathbb{N} \cup \{0\}$ die *natürlichen Zahlen* und Null,
- $\mathbb{Z} := \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$ die *ganzen Zahlen*,
- $\mathbb{Q} := \{\frac{p}{q} | p, q \in \mathbb{Z}, q \neq 0\}$ die *rationalen Zahlen*,
- $\mathbb{R}$ die *reellen Zahlen*, und
- $\mathbb{C} := \{a + ib | a, b \in \mathbb{R}, i := \sqrt{-1}\}$ die *komplexen Zahlen*.

Bemerkungen

- $\mathbb{R}$ umfasst die rationalen Zahlen und die irrationalen Zahlen $\mathbb{R} \setminus \mathbb{Q}$ wie z.B. e , π und $\sqrt{2}$.
- Es gilt $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$.

Definition (Intervalle)

Zusammenhängende Teilmengen der reellen Zahlen heißen *Intervalle*. Für $a, b \in \mathbb{R}$ unterscheidet man

- das *abgeschlossene Intervall*

$$[a, b] := \{x \in \mathbb{R} | a \leq x \leq b\}, \quad (9)$$

- das *offene Intervall*

$$]a, b[:= \{x \in \mathbb{R} | a < x < b\}, \quad (10)$$

- die *halboffenen Intervalle*

$$]a, b] := \{x \in \mathbb{R} | a < x \leq b\} \text{ und } [a, b[:= \{x \in \mathbb{R} | a \leq x < b\}. \quad (11)$$

Bemerkung

- Positiv Unendlich (∞) und negativ Unendlich ($-\infty$) sind keine Element von $\mathbb{R}$.
- Es gilt also immer $] - \infty, b]$ oder $] - \infty, b[$ bzw. $]a, \infty[$ oder $[a, \infty[$, sowie $\mathbb{R} =] - \infty, \infty[$.

Definition (Kartesische Produkte)

M und N seien zwei Mengen. Dann ist das *Kartesische Produkt der Mengen M und N* die Menge aller geordneten Tupel (m, n) mit $m \in M$ und $n \in N$, formal

$$M \times N := \{(m, n) | m \in M, n \in N\}. \quad (12)$$

Das Kartesische Produkt einer Menge M mit sich selbst wird bezeichnet mit

$$M^2 := M \times M. \quad (13)$$

Seien weiterhin $M_1, M_2, \dots, M_n$ Mengen. Dann ist das *Kartesische Produkt der Mengen $M_1, \dots, M_n$* die Menge aller geordneten n -Tupel $(m_1, \dots, m_n)$ mit $m_i \in M_i$ für $i = 1, \dots, n$, formal

$$\prod_{i=1}^n M_i := M_1 \times \dots \times M_n := \{(m_1, \dots, m_n) | m_i \in M_i \text{ für } i = 1, \dots, n\}. \quad (14)$$

Das n -fache Kartesische Produkt einer Menge M mit sich selbst wird bezeichnet mit

$$M^n := \prod_{i=1}^n M := \{(m_1, \dots, m_n) | m_i \in M\}. \quad (15)$$

Definition (Die Menge $\mathbb{R}^n$)

Das n -fache Kartesische Produkt der reellen Zahlen mit sich selbst wird bezeichnet mit

$$\mathbb{R}^n := \prod_{i=1}^n \mathbb{R} := \{x := (x_1, \dots, x_n) | x_i \in \mathbb{R}\} \quad (16)$$

und " $\mathbb{R}$ hoch n " gesprochen. Wir schreiben die Elemente von $\mathbb{R}^n$ typischerweise als Spalten

$$x := \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad (17)$$

und nennen sie *n-dimensionale Vektoren*. Zu Abgrenzung nennen wir die Elemente von $\mathbb{R}^1 = \mathbb{R}$ auch *Skalare*.

Bemerkung

- Ein Beispiel für $x \in \mathbb{R}^4$ ist $x = \begin{pmatrix} 0.16 \\ 1.76 \\ 0.23 \\ 7.10 \end{pmatrix}$.

Mathematische Grundlagen

- Mengen
- **Summen, Produkte, Potenzen**
- Funktionen
- Differentialrechnung
- Analytische Optimierung
- Integralrechnung
- Selbstkontrollfragen

Definition (Summenzeichen)

Summen werden oft mithilfe des *Summenzeichens*

$$\sum_{i=1}^n x_i = x_1 + x_2 + \cdots + x_n \quad (18)$$

dargestellt. Dabei stehen

- Σ für das griechische Sigma, mnemonisch für Summe,
- das Subskript $i = 1$ für den Laufindex der Summanden und den Startindex,
- das Superskript n für den Endindex,
- $x_1, x_2, \dots, x_n$ für die Summanden.

Bemerkung

- Nur mit Subskript und Superskripten macht das Summenzeichen Sinn.
- Die Bezeichnung des Laufindex ist irrelevant, es gilt $\sum_{i=1}^n x_i = \sum_{j=1}^n x_j$.
- In seltenen Fällen wird der Laufindex auch als Element einer Indexmenge angegeben. Ist z.B. die Indexmenge $I := \{1, 5, 7\}$ definiert, so ist $\sum_{i \in I} x_i := x_1 + x_5 + x_7$.

Theorem (Rechenregeln für Summen)

Summen gleicher Summanden

$$\sum_{i=1}^n x = nx \quad (19)$$

Assoziativität bei Summen gleicher Länge

$$\sum_{i=1}^n x_i + \sum_{i=1}^n y_i = \sum_{i=1}^n (x_i + y_i) \quad (20)$$

Distributivität bei Multiplikation mit einer Konstante

$$\sum_{i=1}^n ax_i = a \sum_{i=1}^n x_i \quad (21)$$

Aufspalten von Summen mit $1 < m < n$

$$\sum_{i=1}^n x_i = \sum_{i=1}^m x_i + \sum_{i=m+1}^n x_i \quad (22)$$

Umindizierung

$$\sum_{i=0}^n x_i = \sum_{j=m}^{n+m} x_{j-m} \quad (23)$$

Definition (Produktzeichen)

Produkte werden oft mithilfe des *Produktzeichens*

$$\prod_{i=1}^n x_i = x_1 \cdot x_2 \cdot \dots \cdot x_n \quad (24)$$

dargestellt. Dabei stehen

- $\prod$ für das griechische Π , mnemonisch für *Produkt*,
- das Subskript $i = 1$ für den Laufindex der Produkterme und den Startindex,
- das Superskript n für den Endindex,
- $x_1, x_2, \dots, x_n$ für die Produkterme.

Bemerkungen

- Nur mit Subskript und Superskripten macht das Produktzeichen Sinn.
- Die Bezeichnung des Laufindex ist irrelevant, es gilt $\prod_{i=1}^n x_i = \prod_{j=1}^n x_j$.
- In seltenen Fällen wird der Laufindex auch als Element einer Indexmenge angegeben.
Ist z.B. die Indexmenge $J := \mathbb{N}_2^0$ definiert, so ist $\sum_{j \in J} x_j := x_0 + x_1 + x_2$.

Definition (Potenz)

Für $a \in \mathbb{R}$ und $n \in \mathbb{N}^0$ ist die n -te *Potenz von a* definiert durch

$$a^0 := 1 \text{ und } a^{n+1} := a^n \cdot a. \quad (25)$$

Weiterhin ist für $a \in \mathbb{R} \setminus 0$ und $n \in \mathbb{N}^0$ die *negative n -te Potenz von a* definiert durch

$$a^{-n} := (a^n)^{-1} := \frac{1}{a^n}. \quad (26)$$

a wird dabei *Basis* und n wird *Exponent* genannt.

Theorem (Rechenregel für Potenzen)

Für $a, b \in \mathbb{R}$ und $n, m \in \mathbb{Z}$ mit $a \neq 0$ bei negativen Exponenten gelten

$$a^n a^m = a^{n+m} \quad (27)$$

$$(a^n)^m = a^{nm} \quad (28)$$

$$(ab)^n = a^n b^n \quad (29)$$

Definition (n -te Wurzel)

Für $a \in \mathbb{R}$ und $n \in \mathbb{N}$ ist die n -te Wurzel von a definiert als die Zahl r , so dass

$$r^n = a. \quad (30)$$

Theorem (Potenzschreibweise der n -ten Wurzel)

Es sei $a \in \mathbb{R}$, $n \in \mathbb{N}$, und r die n -te Wurzel von a . Dann gilt

$$r = a^{\frac{1}{n}} \quad (31)$$

Beweis

Es gilt

$$\left(a^{\frac{1}{n}}\right)^n = a^{\frac{1}{n}} \cdot a^{\frac{1}{n}} \cdot \dots \cdot a^{\frac{1}{n}} = a^{\sum_{i=1}^n \frac{1}{n}} = a^1 = a. \quad (32)$$

Also gilt mit der Definition der n -ten Wurzel, dass $r = a^{\frac{1}{n}}$.

□

Bemerkung

- Das Rechnen mit Quadratwurzeln wird durch $\sqrt{x} = x^{\frac{1}{2}}$ sehr erleichtert.

Mathematische Grundlagen

- Mengen
- Summen, Produkte, Potenzen
- **Funktionen**
- Differentialrechnung
- Analytische Optimierung
- Integralrechnung
- Selbstkontrollfragen

Definition (Funktion)

Eine *Funktion* (oder *Abbildung*) f ist eine Zuordnungsvorschrift, die jedem Element einer Menge D genau ein Element einer Zielmenge Z zuordnet. D wird dabei *Definitionsmenge* von f und Z wird *Zielmenge* von f genannt. Wir schreiben

$$f : D \rightarrow Z, x \mapsto f(x), \quad (33)$$

wobei $f : D \rightarrow Z$ gelesen wird als “die Funktion f bildet alle Elemente der Menge D eindeutig auf Elemente in Z ab” und $x \mapsto f(x)$ gelesen wird als “ x , welches ein Element von D ist, wird durch die Funktion f auf $f(x)$ abgebildet, wobei $f(x)$ ein Element von Z ist”. Der Pfeil $\rightarrow$ steht für die Abbildung zwischen den Mengen D und Z , der Pfeil $\mapsto$ steht für die Abbildung zwischen einem Element von D und einem Element von Z .

Bemerkungen

- Es ist zentral, zwischen der *Funktion* f als Zuordnungsvorschrift und einem *Wert* der Funktion $f(x)$ als Element von Z zu unterscheiden.
- x ist der Funktionsinput (auch *Argument* der Funktion genannt), $f(x)$ der Funktionsoutput.
- Üblicherweise folgt $f(x)$ die Definition der *funktionalen Form* von f , z.B.

$$f : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, x \mapsto f(x) := x^2. \quad (34)$$

Definition (Bildmenge, Urbildmenge)

Es sei $f : D \rightarrow Z, x \mapsto f(x)$ eine Funktion und es seien $D' \subseteq D$ und $Z' \subseteq Z$.

- Die Menge

$$f(D') := \{z \in Z \mid \text{Es gibt ein } x \in D' \text{ mit } z = f(x)\} \quad (35)$$

heißt die *Bildmenge von D'* und $f(D) \subseteq Z$ heißt der *Wertebereich* von f .

- Die Menge

$$f^{-1}(Z') := \{x \in D \mid f(x) \in Z'\} \quad (36)$$

heißt die *Urbildmenge von Z'* . $x \in D$ mit $z = f(x) \in Z$ heißt auch *Urbild von z* .

Bemerkung

- Wertebereich $f(D)$ und Zielmenge Z sind nicht notwendigerweise identisch.

Definition (Injektivität, Surjektivität, Bijektivität)

Es sei $f : D \rightarrow Z, x \mapsto f(x)$ eine Funktion.

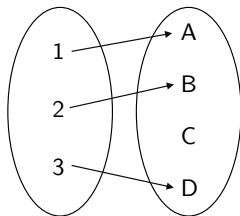
- Die Funktion f heißt *injektiv*, wenn es zu jedem Bild $z \in f(D)$ genau ein Urbild $x \in D$ gibt. Äquivalent gilt, dass f injektiv ist, wenn aus $x_1, x_2 \in D$ mit $x_1 \neq x_2$ folgt, dass $f(x_1) \neq f(x_2)$ ist.
- Die Funktion f heißt *surjektiv*, wenn $f(D) = Z$ gilt, wenn also jedes Element der Zielmenge Z ein Urbild in der Definitionsmenge D hat.
- Die Funktion f heißt *bijektiv*, wenn f injektiv und surjektiv ist.

Bemerkungen

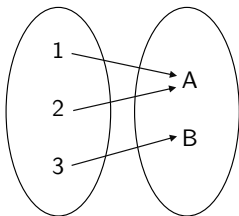
- $f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := x^2$ ist nicht injektiv, weil z.B. für $x_1 = 2 \neq -2 = x_2$ gilt, dass $f(x_1) = 2^2 = 4 = (-2)^2 = f(x_2)$. Weiterhin ist f auch nicht surjektiv, weil z.B. $-1 \in \mathbb{R}$ kein Urbild unter f hat.
- $f : [0, \infty[\rightarrow [0, \infty[, x \mapsto f(x) := x^2$ ist injektiv und surjektiv, also bijektiv.
- Bijektive Abbildungen heißen auch *eindeutige Funktionen* (engl. *one-to-one mappings*).

Beispiele

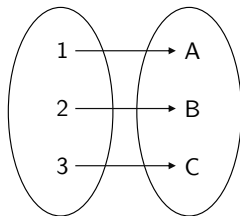
Nicht-surjektiv



Nicht-injektiv



Bijektiv



Definition (Verkettung von Funktionen)

Es seien $f : D \rightarrow R$ und $g : R \rightarrow S$ zwei Funktionen, wobei die Zielmenge von f mit der Definitionsmenge von g übereinstimmen. Dann ist durch

$$g \circ f : D \rightarrow S, x \mapsto (g \circ f)(x) := g(f(x)) \quad (37)$$

eine Funktion definiert, die die *Verkettung von f und g* genannt wird.

Bemerkungen

- $g \circ f$ bezeichnet die Funktion.
- $(g \circ f)(x)$ bezeichnet ein Element in S .
- Erst wird f auf x angewendet, dann wird g auf $f(x)$ angewendet.
- Für $f(x) := -x^2$ und $g(x) := \exp(x)$ ist $(g \circ f)(x) = \exp(-x^2)$.

Definition (Inverse Funktion)

Es sei $f : D \rightarrow R, x \mapsto f(x)$ eine bijektive Funktion. Dann heißt die Funktion f^{-1} mit

$$f^{-1} \circ f : D \rightarrow D, x \mapsto (f^{-1} \circ f)(x) := f^{-1}(f(x)) = x \quad (38)$$

inverse Funktion (oder Umkehrfunktion) von f .

Bemerkungen

- Weil f bijektiv ist, wird jedem $x \in D$ genau ein $z = f(x) \in Z$ zugeordnet.
- Jedem $z \in Z$ wird also auch genau ein $x \in D$ zugeordnet.
- Die inverse Funktion einer bijektiven Funktion ist also auch bijektiv.
- Die inverse Funktion von $f(x) := 2x =: y$ ist $f^{-1}(y) = y/2$.

Definition (Lineare Abbildung)

Eine Abbildung $f : D \rightarrow R, x \mapsto f(x)$ heißt *lineare Abbildung*, wenn für $x, y \in D$ und einen Skalar c gelten, dass

$$f(x + y) = f(x) + f(y) \text{ und } f(cx) = cf(x). \quad (39)$$

Eine Abbildung, für die obige Eigenschaften nicht gelten, heißt *nicht-lineare Abbildung*.

Bemerkungen

- Für $a \in \mathbb{R}$ ist die Abbildung $f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := ax$ linear, weil

$$f(x + y) = a(x + y) = ax + ay = f(x) + f(y) \text{ und } f(cx) = acx = cax = cf(x). \quad (40)$$

- Für $a, b \in \mathbb{R}$ dagegen ist die Abbildung $f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := ax + b$ nicht-linear, weil z.B. für $a := b := 1$ gilt, dass

$$f(x + y) = 1(x + y) + 1 = x + y + 1 \neq x + 1 + y + 1 = f(x) + f(y). \quad (41)$$

- Eine Abbildung der Form $f(x) := ax + b$ heißt *affin-lineare Abbildung*
- Abbildungen der Form $f(x) := ax + b$ werden auch als *lineare Funktionen* bezeichnet.

Definition (Funktionentypen)

In der statistischen Anwendung unterscheiden wir

- *univariate reellwertige Funktionen* der Form

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x), \quad (42)$$

- *multivariate reellwertige Funktionen* der Form

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto f(x) = f(x_1, \dots, x_n), \quad (43)$$

- *multivariate vektorwertige Funktionen* der Form

$$f : \mathbb{R}^n \rightarrow \mathbb{R}^m, x \mapsto f(x) = \begin{pmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_m(x_1, \dots, x_n) \end{pmatrix}. \quad (44)$$

Bemerkung

- In der Physik werden multivariate reellwertige Funktionen auch *Skalarfelder* genannt.
- In der Physik werden multivariate vektorwertige Funktionen auch *Vektorfelder* genannt.
- In manchen Anwendungen kommen auch *matrixvariante matrixwertige Funktionen* zum Tragen.

Elementare univariate reellwertige Funktion der Statistik sind

- die Polynomfunktionen,
- die Exponentialfunktion,
- der Logarithmusfunktion,
- die Gammafunktion.

Wir skizzieren diese Funktionen im Folgenden.

Definition (Polynomfunktionen)

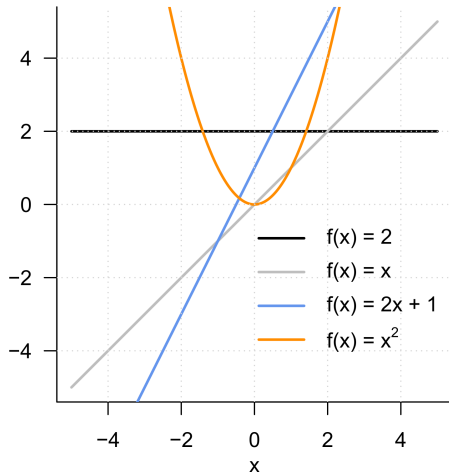
Eine Funktion der Form

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := \sum_{i=0}^k a_i x^i = a_0 + a_1 x^1 + a_2 x^2 + \cdots + a_k x^k \quad (45)$$

heißt *Polynomfunktion* k -ten Grades mit Koeffizienten $a_0, a_1, \dots, a_k \in \mathbb{R}$. Typische Polynomfunktionen sind in nachfolgender Tabelle aufgelistet

Name	Funktionale Form	Koeffizienten
Konstante Funktion	$f(x) = a$	$a_0 := a, a_i := 0, i > 0$
Identitätsfunktion	$f(x) = x$	$a_0 := 0, a_1 := 1, a_i := 0, i > 1$
Lineare Funktion	$f(x) = ax + b$	$a_0 := b, a_1 := a, a_i := 0, i > 1$
Quadratfunktion	$f(x) = x^2$	$a_0 := 0, a_1 := 0, a_2 := 1, a_i := 0, i > 2$

Graphen typischer Polynomfunktionen



Theorem (Exponentialfunktion und ihre Eigenschaften)

Die *Exponentialfunktion* ist definiert als

$$\exp : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \exp(x) := e^x := \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots \quad (46)$$

Die Exponentialfunktion hat folgende Eigenschaften:

Wertebereich	$x \in]-\infty, 0[\Rightarrow \exp(x) \in]0, 1[$ $x \in]0, \infty[\Rightarrow \exp(x) \in]1, \infty[$
Monotonie	$x < y \Rightarrow \exp(x) < \exp(y)$
Spezielle Werte	$\exp(0) = 1$ und $\exp(1) = e$
Exponentialeigenschaften	$\exp(x + y) = \exp(x) \exp(y)$ $\exp(x - y) = \frac{\exp(x)}{\exp(y)}$ $\exp(x) \exp(-x) = 1$

Bemerkungen

- Für Beweise der Eigenschaften wird auf die einschlägige Literatur verwiesen.
- Die Exponentialeigenschaften sind beim Rechnen mit der Normalverteilung zentral.
- $e \approx 2.71$ heißt *Eulersche Zahl*.

Theorem (Logarithmusfunktion und ihre Eigenschaften)

Die *Logarithmusfunktion* ist definiert als inverse Funktion der Exponentialfunktion,

$$\ln :]0, \infty[\rightarrow \mathbb{R}, x \mapsto \ln(x) \text{ mit } \ln(\exp(x)) = x \text{ für alle } x \in \mathbb{R}. \quad (47)$$

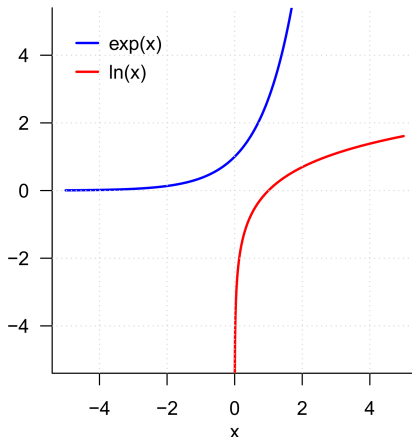
Die Logarithmusfunktion hat folgende Eigenschaften:

Wertebereich	$x \in]0, 1[\Rightarrow \ln(x) \in]-\infty, 0[$ $x \in]1, \infty[\Rightarrow \ln(x) \in]0, \infty[$
Monotonie	$x < y \Rightarrow \ln(x) < \ln(y)$
Spezielle Werte	$\ln(1) = 0$ und $\ln(e) = 1$.
Logarithmeneigenschaften	$\ln(xy) = \ln(x) + \ln(y)$ $\ln(x^c) = c \ln(x)$ $\ln\left(\frac{1}{x}\right) = -\ln(x)$

Bemerkungen

- Für Beweise der Eigenschaften wird auf die einschlägige Literatur verwiesen.
- Die Logarithmeneigenschaften sind beim Rechnen mit Log-Likelihood-Funktionen zentral.
- “Die Logarithmusfunktion wandelt Produkte in Summen und Potenzen in Produkte um.”

Graphen der Exponential- und Logarithmusfunktion



Theorem (Gammafunktion und ihre Eigenschaften)

Die *Gammafunktion* ist definiert durch

$$\Gamma : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \Gamma(x) := \int_0^{\infty} \xi^{x-1} \exp(-\xi) d\xi \quad (48)$$

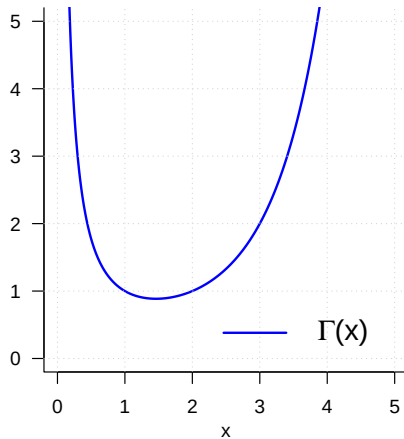
Die Gammafunktion hat folgende Eigenschaften:

Spezielle Werte	$\Gamma(1) = 1$ $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ $\Gamma(n) = (n-1)!$ für $n \in \mathbb{N}$.
Rekursionseigenschaft	Für $x > 0$ gilt $\Gamma(x+1) = x\Gamma(x)$

Bemerkungen

- Für Beweise der Eigenschaften wird auf die einschlägige Literatur verwiesen.
- Die Gammafunktion ist im Kontext von χ^2 -, t - und F -Verteilung zentral.

Graph der Gammafunktion auf $]0, 5[$



Zur Stetigkeit von Funktionen

- Auf eine tiefergehende Diskussion der Stetigkeit von Funktion wird verzichtet.
- Intuitiv bedeutet Stetigkeit, dass beliebig kleine Änderungen des Funktionswertes durch genügend kleine Änderungen des Funktionsargumentes erreicht werden können. Anderes ausgedrückt, heißt dies, dass die Funktion keine Sprünge hat, oder dass man “beim Zeichnen der Funktion den Bleistift nicht absetzen muss”.
- Eine Möglichkeit diese Intuition zu formalisieren, ist es zu fordern, dass für jede Folge $(x_n)_{n \in \mathbb{N}}$ die gegen ein x in der Definitionsmenge der Funktion konvergiert, die korrespondierende Folge von Funktionswerten $(f(x_n))_{n \in \mathbb{N}}$ in der Zielmenge der Funktion gegen den Funktionswert $f(x)$ konvergiert, und zwar für alle x der Definitionsmenge:

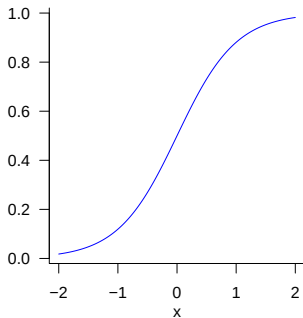
$f : D \rightarrow Z$ ist stetig $\Leftrightarrow$

Für alle $x \in D$ und $(x_n)_{n \rightarrow \infty} \subset D : \lim_{n \rightarrow \infty} x_n = x \Rightarrow \lim_{n \rightarrow \infty} (f(x_n)) = f(x)$.

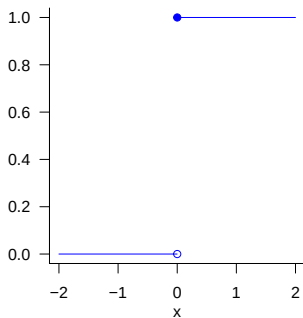
- Die Polynom-, Exponential-, Logarithmusfunktion sind stetige Funktionen.
- Nichtstetige Funktionen haben oft Fallunterscheidungen, z.B. $f(x) := \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$.

Zur Stetigkeit von Funktionen

Stetige Funktion



Nichtstetige Funktion



Mathematische Grundlagen

- Mengen
- Summen, Produkte, Potenzen
- Funktionen
- **Differentialrechnung**
- Analytische Optimierung
- Integralrechnung
- Selbstkontrollfragen

Definition (Ableitung)

Es sei $I \subseteq \mathbb{R}$ ein Intervall und

$$f : I \rightarrow \mathbb{R}, x \mapsto f(x) \quad (49)$$

eine univariate reellwertige Funktion. f heißt in $a \in I$ *differenzierbar*, wenn der Grenzwert

$$f'(a) := \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h} \quad (50)$$

existiert. $f'(a)$ heißt dann die *Ableitung von f an der Stelle a* . Ist f differenzierbar für alle $x \in I$, so heißt f *differenzierbar* und die Funktion

$$f' : I \rightarrow \mathbb{R}, x \mapsto f'(x) \quad (51)$$

heißt *Ableitung von f*

Bemerkungen

- Für $h > 0$ heißt $\frac{f(a+h) - f(a)}{h}$ *Differenzquotient*.
- Der Differenzquotient misst die Änderung $f(a+h) - f(a)$ von f pro Strecke h .
- Für $h \rightarrow 0$ misst der Differenzquotient die Änderungsrate von f in a .
- $f'(a)$ ist eine Zahl, f' ist eine Funktion.
- Wir werden keine Grenzwertbildung zur Berechnung von Ableitungen benötigen.

Definition (Notation für Ableitungen univariater reellwertiger Funktionen)

Es sei f eine univariate reellwertige Funktion. Äquivalente Schreibweisen für die Ableitung von f und die Ableitung von f an einer Stelle x sind

- (1) Die *Lagrange-Notation* f' und $f'(x)$
- (2) Die *Newton-Notation* $\dot{f}$ und $\dot{f}(x)$.
- (3) Die *Leibniz-Notation* $\frac{df}{dx}$ und $\frac{df(x)}{dx}$.
- (4) Die *Euler-Notation* Df und $Df(x)$.

Bemerkungen

- Für univariate reellwertige Funktionen benutzen wir f' und $f'(x)$ als Bezeichner.
- In Berechnungen benutzen wir auch die “Operator-Schreibweise” $\frac{d}{dx}f(x)$.
- Wir verstehen $\frac{d}{dx}f(x)$ als den Auftrag, die Ableitung von f zu berechnen.

Definition (Höhere Ableitungen)

Es sei f eine univariate reellwertige Funktion und

$$f^{(1)} := f' \quad (52)$$

sei die Ableitung von f . Die k -te Ableitung von f ist rekursiv definiert durch

$$f^{(k)} := \left(f^{(k-1)}\right)' \text{ für } k \geq 0, \quad (53)$$

unter der Annahme, dass $f^{(k-1)}$ differenzierbar ist. Insbesondere ist die *zweite Ableitung* von f definiert durch die Ableitung von f' , also

$$f'' := (f')' \quad (54)$$

Bemerkung

- Wir schreiben auch $\frac{d^2}{dx^2} f(x)$ für den Auftrag, die zweite Ableitung von f zu bestimmen.
- Die nullte Ableitung $f^{(0)}$ von f ist f selbst.
- Üblicherweise schreibt man für $k < 4$ f' , f'' , f''' statt $f^{(1)}$, $f^{(2)}$, $f^{(3)}$.
- Im Allgemeinen benötigen wir nur f' und f'' .

Theorem (Rechenregeln für Ableitungen)

Für $i = 1, \dots, n$ seien g_i reellwertige univariate differenzierbare Funktionen. Dann gelten folgende Rechenregeln für Ableitungen

(1) Summenregel

$$\text{Für } f(x) := \sum_{i=1}^n g_i(x) \text{ gilt } f'(x) = \sum_{i=1}^n g'_i(x). \quad (55)$$

(2) Produktregel

$$\text{Für } f(x) := g_1(x)g_2(x) \text{ gilt } f'(x) = g'_1(x)g_2(x) + g_1(x)g'_2(x). \quad (56)$$

(3) Quotientenregel

$$\text{Für } f(x) := \frac{g_1(x)}{g_2(x)} \text{ gilt } f'(x) = \frac{g'_1(x)g_2(x) - g_1(x)g'_2(x)}{g_2^2(x)} \quad (57)$$

(4) Kettenregel

$$\text{Für } f(x) := g_1(g_2(x)) \text{ gilt } f'(x) = g'_1(g_2(x))g'_2(x) \quad (58)$$

Bemerkung

- Für Herleitung und Beweise der Rechenregeln wird auf die einschlägige Literatur verwiesen.

Theorem (Ableitungen elementarer Funktionen)

Für elementare Funktionen der Statistik ergeben sich folgende Ableitungen

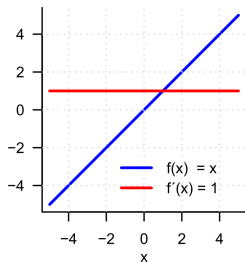
Name	Definition	Ableitung
Polynomfunktionen	$f(x) := \sum_{i=0}^n a_i x^i$	$f'(x) = \sum_{i=1}^n i a_i x^{i-1}$
Konstante Funktion	$f(x) := a$	$f'(x) = 0$
Identitätsfunktion	$f(x) := x$	$f'(x) = 1$
Lineare Funktion	$f(x) := ax + b$	$f'(x) = a$
Quadratfunktion	$f(x) := x^2$	$f'(x) = 2x$
Exponentialfunktion	$f(x) := \exp(x)$	$f'(x) = \exp(x)$
Logarithmusfunktion	$f(x) := \ln(x)$	$f'(x) = \frac{1}{x}$

Bemerkung

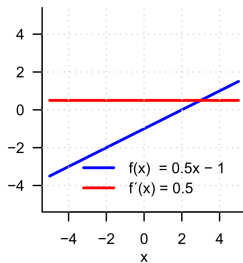
- Für Beweise wird auf die einschlägige Literatur verwiesen.

Beispiele

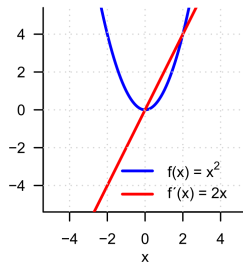
Identitätsfunktion



Lineare Funktion

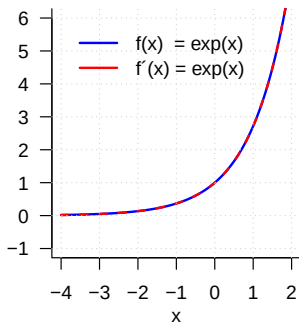


Quadratfunktion

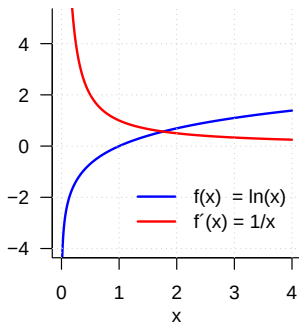


Beispiele

Exponentialfunktion



Logarithmusfunktion



Mathematische Grundlagen

- Mengen
- Summen, Produkte, Potenzen
- Funktionen
- Differentialrechnung
- **Analytische Optimierung**
- Integralrechnung
- Selbstkontrollfragen

Definition (Extremstellen und Extremwerte)

Es sei $U \subseteq \mathbb{R}$ und $f : U \rightarrow \mathbb{R}$ eine univariate reellwertige Funktion. f hat an der Stelle $x_0 \in U$

- ein *lokales Minimum*, wenn es ein Intervall $I :=]a, b[$ gibt mit $x_0 \in]a, b[$ und

$$f(x_0) \leq f(x) \text{ für alle } x \in I \cap U, \quad (59)$$

- ein *globales Minimum*, wenn gilt, dass

$$f(x_0) \leq f(x) \text{ für alle } x \in U, \quad (60)$$

- ein *lokales Maximum*, wenn es ein Intervall $I :=]a, b[$ gibt mit $x_0 \in]a, b[$ und

$$f(x_0) \geq f(x) \text{ für alle } x \in I \cap U, \quad (61)$$

- *lokales Maximum*, wenn gilt, dass

$$f(x_0) \geq f(x) \text{ für alle } x \in U. \quad (62)$$

Der Wert $x_0 \in U$ der Definitionsmenge von f heißt entsprechend *lokale* oder *globale Minimalstelle* oder *Maximalstelle*, der Funktionswert $f(x_0) \in \mathbb{R}$ heißt entsprechend *lokales* oder *globales Minimum* oder *Maximum*. Generell heißt der Wert $x_0 \in U$ *Extremstelle* und der Funktionswert $f(x_0) \in \mathbb{R}$ *Extremwert*.

Bemerkungen

- Extremstellen werden auch mit $\arg \min_{x \in I \cap U} f(x)$ oder $\arg \max_{x \in I \cap U} f(x)$ bezeichnet.
- Extremwerte werden auch mit $\min_{x \in I \cap U} f(x)$ oder $\max_{x \in I \cap U} f(x)$ bezeichnet.

Definition (Notwendige Bedingung für Extrema)

f sei eine univariate reellwertige Funktion. Dann gilt

$$x_0 \text{ ist Extremstelle von } f \Rightarrow f'(x_0) = 0. \quad (63)$$

Bemerkungen

- Wenn x_0 eine Extremstelle von f ist, dann ist die erste Ableitung von f in x_0 null.
- Sei zum Beispiel x_0 eine **lokale Maximalstelle** von f . Dann gilt
 - Links von x_0 steigt f an, rechts von x_0 fällt f ab.
 - In x_0 steigt f weder an, noch fällt f ab, also ist $f'(x_0) = 0$.

Definition (Hinreichende Bedingungen für lokale Extrema)

f sei eine zweimal differenzierbare univariate reellwertige Funktion.

- Wenn für $x_0 \in U \subseteq \mathbb{R}$

$$f'(x_0) = 0 \text{ und } f''(x_0) > 0 \quad (64)$$

gilt, dann hat f an der Stelle x_0 ein Minimum.

- Wenn für $x_0 \in U \subseteq \mathbb{R}$

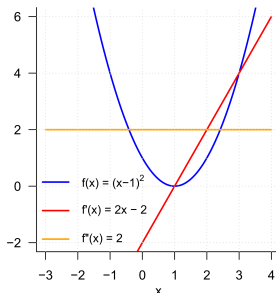
$$f'(x_0) = 0 \text{ und } f''(x_0) < 0 \quad (65)$$

gilt, dann hat f an der Stelle x_0 ein Maximum.

Bemerkung

- Eine Intuition vermittelt nachfolgende Abbildung.

Hinreichende Bedingungen für lokale Extrema



Hier ist offenbar $x_0 = 1$ eine lokale Minimalstelle von $f(x) = (x - 1)^2$. Man erkennt:

- Links von x_0 fällt f ab, rechts von x_0 steigt f an.
- In x_0 steigt f weder an, noch fällt f ab, also ist $f'(x_0) = 0$.
- Links und rechts von x_0 und in x_0 ist die Änderung f'' von f' positiv.
- Links von x_0 schwächt sich die Negativtaet von f' zu 0 ab.
- Rechts von x_0 verstärkt sich die Positivtaet von f' .

Definition (Standardverfahren der analytischen Optimierung)

f sei eine univariate reellwertige Funktion. Lokale Extremstellen von f können mit folgendem *Standardverfahren der analytischen Optimierung* identifiziert werden:

- (1) Berechnen der ersten und zweiten Ableitung von f .
- (2) Bestimmen von Nullstellen x^* von f' durch Auflösen von $f'(x^*) = 0$ nach x^* .
⇒ Nullstellen von f' sind Kandidaten für Extremstellen von f .
- (3) Evaluation von $f''(x^*)$.
⇒ Wenn $f''(x^*) > 0$, dann ist x^* lokale Minimumstelle von f .
⇒ Wenn $f''(x^*) < 0$, dann ist x^* lokale Maximumstelle von f .
⇒ Wenn $f''(x^*) = 0$, dann ist x^* keine Extremstelle von f .

Beispiel

- Wir betrachten die Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := (x - 1)^2. \quad (66)$$

- Die erste Ableitung von f ergibt sich mit der Kettenregel zu

$$f'(x) = \frac{d}{dx} ((x - 1)^2) = 2(x - 1) \cdot \frac{d}{dx} (x - 1) = 2x - 2. \quad (67)$$

- Die zweite Ableitung von f ergibt sich zu

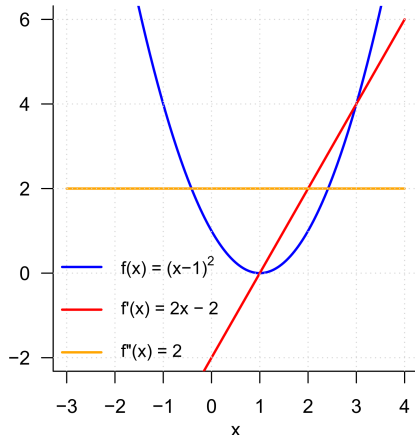
$$f''(x) = \frac{d}{dx} f'(x) = \frac{d}{dx} (2x - 2) = 2 > 0 \text{ für alle } x \in \mathbb{R}. \quad (68)$$

- Auflösen von $f'(x^*) = 0$ nach x^* ergibt

$$f'(x^*) = 0 \Leftrightarrow 2x^* - 2 = 0 \Leftrightarrow 2x^* = 2 \Leftrightarrow x^* = 1. \quad (69)$$

- $x^* = 1$ ist folglich eine Minimalstelle von f mit zugehörigen Minimalwert $f(1) = 0$.

Beispiel



Mathematische Grundlagen

- Mengen
- Summen, Produkte, Potenzen
- Funktionen
- Differentialrechnung
- Analytische Optimierung
- **Integralrechnung**
- Selbstkontrollfragen

Übersicht

- Integrale kommen in der W-Theorie und Statistik an vielen Stellen vor:
 - WDF Definition und Berechnen von Wahrscheinlichkeiten aus WDF
 - Beziehung von WDF und KVF
 - Erwartungswerte inklusive Varianz und Kovarianz
 - u.v.a.m.
- Ziel hier ist es, zentrale Begriffe aus der Schulmathematik zu wiederholen.
- In der Schulmathematik wird üblicherweise das *Riemannintegral* betrachtet.
- Zentrale Begriffe sind *Stammfunktion*, *unbestimmtes Integral*, *bestimmtes Integral*, *Hauptsätze der Differential- und Integralrechnung*, *uneigentliches Integral*, und *mehrdimensionales Integral*.
- Der Fokus liegt auf der Klärung von Begriffen und Symbolen, nicht dem Rechnen.
- Auf die Einführung des modernen Integralbegriffs nach Lebesgue wird verzichtet.

Definition (Stammfunktion, Unbestimmtes Integral)

Für ein Intervall $I \subset \mathbb{R}$ sei $f : I \rightarrow \mathbb{R}$ eine univariate reellwertige Funktion. Dann heißt eine differenzierbare Funktion $F : I \rightarrow \mathbb{R}$ mit der Eigenschaft

$$F' = f \quad (70)$$

Stammfunktion von f . Ist F eine Stammfunktion von f , dann heißt

$$\int f(x) dx := F(x) + c \text{ mit } c \in \mathbb{R} \quad (71)$$

unbestimmtes Integral der Funktion f .

Bemerkungen

- Die Ableitung der Stammfunktion F von f ist f .
- Das unbestimmte Integral ist die Gesamtheit aller Stammfunktionen von f .
- Die Konstante $c \in \mathbb{R}$ heißt *Integrationskonstante*, es gilt $\frac{d}{dx} c = 0$.
- Der Ausdruck $\int f(x) dx$ ist als $F(x) + c$ definiert
- In $\int f(x) dx$ haben $\int$ und dx keine eigentliche Bedeutung, $f(x)$ heißt *Integrand*.

Theorem (Stammfunktionen elementarer Funktionen)

Für elementare Funktionen der Statistik ergeben sich folgende Stammfunktionen

Name	Definition	Stammfunktion
Polynomfunktionen	$f(x) := \sum_{i=0}^n a_i x^i$	$F(x) = \sum_{i=0}^n \frac{a_i}{i+1} x^{i+1}$
Konstante Funktion	$f(x) := a$	$F(x) = ax$
Identitätsfunktion	$f(x) := x$	$F(x) = \frac{1}{2}x^2$
Lineare Funktion	$f(x) := ax + b$	$F(x) = \frac{1}{2}ax^2 + bx$
Quadratfunktion	$f(x) := x^2$	$F(x) = \frac{1}{3}x^3$
Exponentialfunktion	$f(x) := \exp(x)$	$F(x) = \exp(x)$
Logarithmusfunktion	$f(x) := \ln(x)$	$F(x) = x \ln x - x$

Bemerkung

- Beweise ergeben sich direkt durch Ableiten der Stammfunktionen.

Theorem (Rechenregeln für Stammfunktionen)

f und g seien univariate reellwertige Funktion, die Stammfunktionen besitzen, und g sei invertierbar. Dann gelten folgende Rechenregeln für die Bestimmung von Integralen

(1) Summenregel

$$\int a f(x) + b g(x) dx = a \int f(x) dx + b \int g(x) dx \text{ für } a, b \in \mathbb{R} \quad (72)$$

(2) Partielle Integration

$$\int f'(x) g(x) dx = f(x) g(x) - \int f(x) g'(x) dx \quad (73)$$

(3) Substitutionsregeln

$$\int f(g(x)) g'(x) dx = \int f(t) dt \text{ mit } t = g(x) \quad (74)$$

Bemerkung

- Für die Herleitung der Summenregel wird auf die einschlägige Literatur verwiesen.

Integralrechnung

Beweis

Die Rechenregel der partiellen Integration ergibt sich durch Integration der Produktregel der Differentiation. Wir erinnern uns, dass gilt

$$(f(x)g(x))' = f'(x)g(x) + f(x)g'(x). \quad (75)$$

Integration beider Seiten der Gleichung und Berücksichtigung der Summenregel für Stammfunktionen ergibt dann

$$\begin{aligned} \int (f(x)g(x))' dx &= \int f'(x)g(x) + f(x)g'(x) dx \\ \Leftrightarrow f(x)g(x) &= \int f'(x)g(x) dx + \int f(x)g'(x) dx \\ \Leftrightarrow \int f'(x)g(x) dx &= f(x)g(x) - \int f(x)g'(x) dx \end{aligned} \quad (76)$$

Die Substitutionsregel ergibt sich für $F' = f$ durch Anwendung der Kettenregel der Differentiation auf die verkettete Funktion $F(g)$. Speziell gilt zunächst

$$(F(g(x)))' = F'(g(x))g'(x) = f(g(x))g'(x) \quad (77)$$

Integration beider Seiten der Gleichung

$$(F(g(x)))' = f(g(x))g'(x) \quad (78)$$

ergibt dann

$$\begin{aligned} \int (F(g(x)))' dx &= \int f(g(x))g'(x) dx \\ \Leftrightarrow F(g(x)) + c &= \int f(g(x))g'(x) dx \\ \Leftrightarrow \int f(g(x))g'(x) dx &= \int f(t) dt \text{ mit } t := g(x). \end{aligned} \quad (79)$$

Dabei ist die rechte Seite der letzten obigen Gleichung zu verstehen als $F(g(x)) + c$, also als Stammfunktion von f evaluiert an der Stelle $t := g(x)$. Das dt ist nicht durch $dg(x)$ zu ersetzen, sondern rein notationstechnischer Natur. $\square$

Bestimmtes Integral von $f : [a, b] \rightarrow \mathbb{R}$

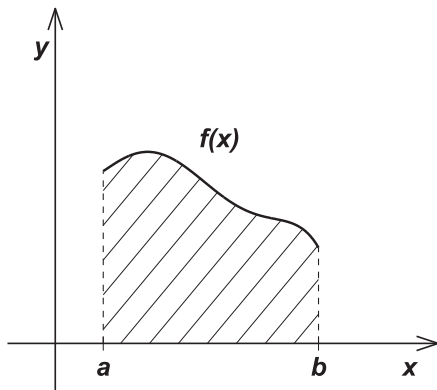


Abbildung 2.54 aus Bärwolff (2017)

Definition (Zerlegung eines Intervalls, Feinheit)

Es sei $[a, b] \subset \mathbb{R}$ ein Intervall und $x_0, x_1, x_2, \dots, x_n \in [a, b]$ eine Menge von Punkten mit

$$a =: x_0 < x_1 < x_2 \cdots < x_n := b \quad (80)$$

mit

$$\Delta x_i := x_i - x_{i-1} \text{ für } i = 1, \dots, n. \quad (81)$$

Dann heißt die Menge

$$Z := \{[x_0, x_1], [x_1, x_2], \dots, [x_{n-1}, x_n]\} \quad (82)$$

der durch $x_0, x_1, x_2, \dots, x_n$ definierten Teilintervalle von $[a, b]$ *Zerlegung von $[a, b]$* .
Weiterhin heißt

$$Z_{\max} := \max_{i \in n} \Delta x_i, \quad (83)$$

also die größte der Teilintervalllängen Δx_i , die *Feinheit* von Z .

Bemerkung

- Δx_i ist die Breite der der Streifen in untenstehender Abbildung.

Definition (Riemannsche Summe)

$f : [a, b] \rightarrow \mathbb{R}$ sei eine beschränkte Funktion auf $[a, b]$, d.h. $|f(x)| < c$ für $0 < c < \infty$ und alle $x \in [a, b]$, Z sei eine Zerlegung von $[a, b]$ mit Teilintervalllängen Δx_i für $i = 1, \dots, n$. Weiterhin sei für $i = 1, \dots, n$ ξ_i ein beliebiger Punkt im Teilintervall $[x_{i-1}, x_i]$ der Zerlegung Z . Dann heißt

$$R(Z) := \sum_{i=1}^n f(\xi_i) \Delta x_i \quad (84)$$

Riemannsche Summe von f auf $[a, b]$ bezüglich der Zerlegung Z .

Bemerkungen

- Wählt man in jedem $[x_{i-1}, x_i]$ das Maximum von f , ergibt sich die *Riemannsche Obersumme*.
- Wählt man in jedem $[x_{i-1}, x_i]$ das Minimum von f , ergibt sich die *Riemannsche Untersumme*.
- Für $\Delta x_i \rightarrow 0, i = 1, \dots, n$ geht der Unterschied zwischen Ober- und Untersumme gegen 0.

Zerlegung und Riemann Ober- und Untersummen für $f : [a, b] \rightarrow \mathbb{R}$

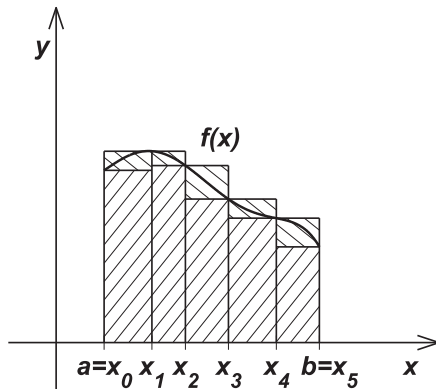


Abbildung 2.55 aus Bärwolff (2017)

Theorem (Riemannsches Integral)

$f : [a, b] \rightarrow \mathbb{R}$ sei eine beschränkte reellwertige Funktion auf $[a, b]$. Weiterhin sei für $Z_k, k = 1, 2, 3, \dots$ eine Folge von Zerlegungen von $[a, b]$ mit zugehörigen Feinheit $|Z_k|$. Wenn für jede Folge von Zerlegungen $Z_1, Z_2, \dots$ mit $|Z_k| \rightarrow 0$ für $k \rightarrow \infty$ und für beliebig gewählte Punkte $\xi_{ki}, i = 1, \dots, n$ im Teilintervall $[x_{k,i-1}, x_{k,i}]$ der Zerlegung Z_k gilt, dass die Folge der zugehörigen Riemannschen Summen $R(Z_1), R(Z_2), \dots$ gegen den gleichen Grenzwert strebt, dann heißt f auf $[a, b]$ *integrierbar*. Der entsprechende Grenzwert der Folge von Riemannschen Summen wird *bestimmtes Riemannsches Integral* genannt und mit

$$\int_a^b f(x) dx := \lim_{k \rightarrow \infty} R(Z_k) \text{ für } |Z_k| \rightarrow 0 \quad (85)$$

bezeichnet.

Bemerkungen

- Für $f > 0$ ist $\int_a^b f(x) dx$ der Flächeninhalt zwischen den $f(x)$ und der x -Achse
- Generell ist $\int_a^b f(x) dx$ der vorzeichenbehaftete Flächeninhalt den $f(x)$ und der x -Achse.
- Positive und negative Flächeninhalt gleichen einander aus.
- $\int_a^b f(x) dx$ ist als Mittelwert von f auf $[a, b]$ zu verstehen.

Theorem (Mittelwertsatz der Integralrechnung)

Für eine stetige Funktion $f : [a, b] \rightarrow \mathbb{R}$ existiert ein $\xi \in]a, b[$ mit

$$\int_a^b f(x) dx = f(\xi)(b - a) \quad (86)$$

- Statt eines Beweises verweisen wir auf untenstehende graphische Darstellung

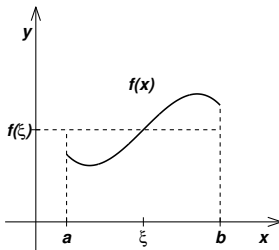


Abbildung 2.59 aus Bärwolff (2017)

Theorem (Rechenregeln für bestimmte Integrale)

Es seien f und g integrierbare Funktionen auf $[a, b]$. Dann gelten folgende Rechenregeln.

(1) Linearität. Für $c_1, c_2 \in \mathbb{R}$ gilt

$$\int_a^b (c_1 f(x) + c_2 g(x)) dx = c_1 \int_a^b f(x) dx + c_2 \int_a^b f(x) dx. \quad (87)$$

(2) Additivität. Für $a < c < b$ gilt

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx. \quad (88)$$

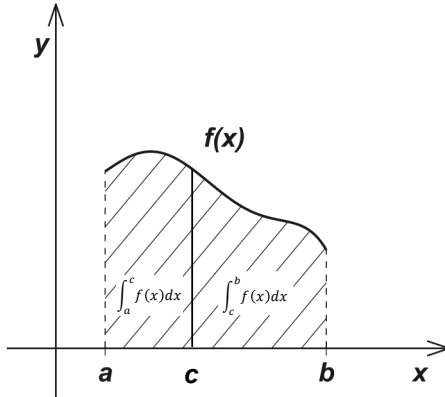
(2) Vorzeichenwechsel bei Umkehrung der Integralgrenzen

$$\int_a^b f(x) dx = - \int_b^a f(x) dx. \quad (89)$$

Bemerkung

- Für eine formale Herleitung verweisen wir auf die einschlägige Literatur.
- Eine graphische Darstellung der Additivität findet sich untenstehend.

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx$$



Nach Abbildung 2.55 aus Bärwolff (2017)

Theorem (Erster Hauptsatz der Differential- und Integralrechnung)

Ist $f : I \rightarrow \mathbb{R}$ eine auf dem Intervall $I \subset \mathbb{R}$ stetige Funktion, dann ist die Funktion

$$F : I \rightarrow \mathbb{R}, x \mapsto F(x) := \int_a^x f(t) dt \text{ mit } x, a \in I \quad (90)$$

eine Stammfunktion von f .

Beweis

Wir betrachten den Differenzquotienten

$$\frac{1}{h} (F(x+h) - F(x)) \quad (91)$$

Mit der Definition $F(x) := \int_a^x f(t) dt$ und der Additivität des bestimmten Integrals gilt dann

$$\frac{1}{h} (F(x+h) - F(x)) = \frac{1}{h} \left(\int_a^{x+h} f(t) dt - \int_a^x f(t) dt \right) = \frac{1}{h} \int_x^{x+h} f(t) dt \quad (92)$$

Mit dem Mittelwertsatz der Integralrechnung gibt es also ein $\xi \in]x, x+h[$, so dass

$$\frac{1}{h} (F(x+h) - F(x)) = f(\xi) \quad (93)$$

Grenzwertbildung ergibt dann

$$\lim_{h \rightarrow 0} \frac{1}{h} (F(x+h) - F(x)) = \lim_{h \rightarrow 0} f(\xi) \text{ für } \xi \in]x, x+h[\Leftrightarrow F'(x) = f(x). \quad (94)$$

□

Theorem (Zweiter Hauptsatz der Differential- und Integralrechnung)

Ist F eine Stammfunktion einer stetigen Funktion $f : I \rightarrow \mathbb{R}$ auf einem Intervall I , so gilt für $a, b \in I$ mit $a \leq b$

$$\int_a^b f(x) dx = F(b) - F(a) =: F(x) \Big|_a^b \quad (95)$$

Beweis

Mit den Rechenregeln für bestimmte Integrale und dem ersten Hauptsatz der Differential- und Integralrechnung ergibt sich

$$F(b) - F(a) = \int_a^b f(t) dt - \int_a^a f(t) dt = \int_a^b f(x) dx \quad (96)$$

□

Definition (Uneigentliche Integrale)

$f : \mathbb{R} \rightarrow \mathbb{R}$ sei eine univariate reellwertige Funktion. Mit den Definitionen

$$\int_{-\infty}^b f(x) dx := \lim_{a \rightarrow -\infty} \int_a^b f(x) dx \text{ und } \int_a^{\infty} f(x) dx := \lim_{b \rightarrow \infty} \int_a^b f(x) dx \quad (97)$$

und der Additivität von Integralen

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^b f(x) dx + \int_b^{\infty} f(x) dx \quad (98)$$

wird der Begriff des bestimmten Integrals auf die unbeschränkten Integrationsintervalle $]-\infty, b]$, $[a, \infty[$ und $]-\infty, \infty[$ erweitert. Integrale mit unbeschränkten Integrationsintervallen heißen *uneigentliche Integrale*. Wenn die entsprechenden Grenzwerte existieren, sagt man, dass die uneigentlichen Integrale *konvergieren*.

Bemerkung

- Für die WDF f einer Zufallsvariable ist die Forderung $\int_{-\infty}^{\infty} f(x) dx = 1$ zentral.

Beispiel (Uneigentliches Integral)

Wir betrachten das uneigentliche Integral $\int_1^\infty \frac{1}{x^2} dx$. Nach den Festlegungen in der Definition uneigentlicher Integrale gilt

$$\int_1^\infty \frac{1}{x^2} dx = \lim_{b \rightarrow \infty} \int_1^b \frac{1}{x^2} dx. \quad (99)$$

Mit der Stammfunktion $F(x) = -x^{-1}$ von $f(x) = x^{-2}$ ergibt sich für das bestimmte Integral in obiger Gleichung

$$\int_1^b \frac{1}{x^2} dx = F(b) - F(1) = -\frac{1}{b} - \left(-\frac{1}{1}\right) = -\frac{1}{b} + 1. \quad (100)$$

Es ergibt sich also

$$\int_1^\infty \frac{1}{x^2} dx = \lim_{b \rightarrow \infty} \int_1^b \frac{1}{x^2} dx = \lim_{b \rightarrow \infty} \left(-\frac{1}{b} + 1\right) = -\lim_{b \rightarrow \infty} \frac{1}{b} + \lim_{b \rightarrow \infty} 1 = 0 + 1 = 1. \quad (101)$$

Definition (Mehrdimensionale Integrale)

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ sei eine multivariate reellwertige Funktion. Dann heißen Integrale der Form

$$\int_{[a_1, b_1] \times \dots \times [a_n, b_n]} f(x) dx = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} f(x_1, \dots, x_n) dx_1 \dots dx_n \quad (102)$$

mehrdimensionale bestimmte Integrale auf Hyperrechtecken. Weiterhin heißen Integrale der Form

$$\int_{\mathbb{R}^n} f(x) dx = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_1 \dots dx_n \quad (103)$$

mehrdimensionale uneigentliche Integrale.

Bemerkungen

- Man kann multivariate reellwertige Funktion nicht nur auf Hyperrechtecken, sondern im Prinzip auf beliebigen Hyperflächen integrieren. Dies kann sich jedoch oft schwierig gestalten.
- Der *Satz von Fubini* besagt, dass man mehrdimensionale Integrale in beliebiger Koordinatenfolge auswerten kann, also dass zum Beispiel

$$\int_{a_1}^{b_1} \left(\int_{a_2}^{b_2} f(x_1, x_2) dx_2 \right) dx_1 = \int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x_1, x_2) dx_1 \right) dx_2. \quad (104)$$

- Für die WDF eines Zufallsvektors ist die Forderung $\int_{\mathbb{R}^n} f(x) dx = 1$ zentral.

Beispiel (Zweidimensionales Integral)

Wir betrachten das zweidimensionale bestimmte Integral der Funktion

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, (x_1, x_2) \mapsto f(x_1, x_2) := x_1^2 + 4x_2 \quad (105)$$

auf dem Rechteck $[0, 1] \times [0, 1]$. In

$$\int_0^1 \int_0^1 x_1^2 + 4x_2 \, dx_1 \, dx_2 = \int_0^1 \left(\int_0^1 x_1^2 + 4x_2 \, dx_1 \right) dx_2 \quad (106)$$

betrachten wir zunächst das innere Integral. x_2 nimmt hier die Rolle einer Konstanten ein. Eine Stammfunktion von $g(x_1) := x_1^2 + 4x_2$ ist $G(x_1) = \frac{1}{3}x_1^3 + 4x_2x_1$, wie man sich durch Ableiten von G überzeugt. Es ergibt sich also für das innere Integral

$$\int_0^1 x_1^2 + 4x_2 \, dx_1 = G(1) - G(0) = \frac{1}{3} \cdot 1^3 + 4x_2 \cdot 1 - \frac{1}{3} \cdot 0^3 - 4x_2 \cdot 0 = \frac{1}{3} + 4x_2 \quad (107)$$

Betrachten des äußeren Integrals in Gleichung (106) ergibt dann mit der Stammfunktion $H(x_2) = \frac{1}{3}x_2 + 2x_2^2$ von $h(x_2) := \frac{1}{3} + 4x_2$, dass

$$\int_0^1 \int_0^1 x_1^2 + 4x_2 \, dx_1 \, dx_2 = \int_0^1 \frac{1}{3} + 4x_2 \, dx_2 = H(1) - H(0) = \frac{1}{3} \cdot 1 + 4 \cdot 1^2 - \frac{1}{3} \cdot 0 + 4 \cdot 0^2 = \frac{13}{3}. \quad (108)$$

Mathematische Grundlagen

- Mengen
- Summen, Produkte, Potenzen
- Funktionen
- Differentialrechnung
- Analytische Optimierung
- Integralrechnung
- **Selbstkontrollfragen**

Selbstkontrollfragen

1. Diskutieren Sie den Unterschiede zwischen Definitionen, Theoremen, und Beweisen.
2. Geben Sie die Definition einer Menge nach Cantor (1895) wieder.
3. Nennen Sie drei Möglichkeiten zur Definition einer Menge.
4. Für Mengen M, N erläutern Sie die Ausdrücke $m \in M, m \notin N, M \subseteq N, M \subset N$.
5. Definieren Sie den Begriff der Kardinalität einer Menge.
6. Definieren Sie den Begriff der Potenzmenge einer Menge.
7. Es sei $M := \{a, b\}$. Bestimmen Sie $\mathcal{P}(M)$.
8. Es seien $M := \{a, b\}, N := \{a, c, d\}$. Bestimmen Sie $M \cup N, M \cap N, M \setminus N, M \Delta N$.
9. Erläutern Sie die Symbole $\mathbb{N}, \mathbb{N}_n$, und $\mathbb{N}^0$.
10. Erläutern Sie den Unterschied zwischen $\mathbb{N}$ und $\mathbb{Z}$.
11. Erläutern Sie den Unterschied zwischen $\mathbb{R}$ und $\mathbb{Q}$.
12. Definieren Sie die Begriffe des abgeschlossenen, offenen, und halboffenen Intervalls.
13. Es seien M und N Mengen. Erläutern Sie die Notation $M \times N$.
14. Definieren sie die Menge $\mathbb{R}^n$.
15. Berechnen Sie die Summen $\sum_{i=1}^3 2$, $\sum_{i=1}^3 i^2$, und $\sum_{i=1}^3 \frac{2}{3}i$.
16. Schreiben Sie die Summe $1 + 3 + 5 + 7 + 9 + 11$ mithilfe des Summenzeichens.

Selbstkontrollfragen

17. Schreiben Sie die Summe $0 + 2 + 4 + 6 + 8 + 10$ mithilfe des Summenzeichens.
18. Definieren Sie das Produktzeichen.
19. Für $a \in \mathbb{R}$, definieren Sie die n te (negative) Potenz von a .
20. Berechnen Sie $2^2 \cdot 2^3$ und 2^5 . Geben Sie die zugehörige Potenzregel an.
21. Berechnen Sie 6^2 und $2^2 \cdot 3^2$. Geben Sie die zugehörige Potenzregel an.
22. Warum kann die n -te Wurzel von a als $a^{\frac{1}{n}}$ geschrieben werden?
23. Berechnen Sie $(\sqrt{2})^{\frac{2}{3}}$, $9^{\frac{1}{2}}$, und $4^{-\frac{1}{2}}$.
24. Erläutern Sie die Komponenten der Funktionsschreibweise $f : D \rightarrow Z, x \mapsto f(x)$.
25. Definieren Sie die Begriffe Bildmenge, Wertebereich, und Urbildmenge einer Funktion.
26. Definieren Sie die Begriffe Surjektivität, Injektivität, und Bijektivität einer Funktion.
27. Erläutern Sie, warum $f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := x^2$ weder injektiv noch surjektiv ist.
28. Erläutern Sie, warum $f : [0, \infty[\rightarrow [0, \infty[, x \mapsto f(x) := x^2$ bijektiv ist.
29. Erläutern Sie die Komponenten der Schreibweise $g \circ f : D \rightarrow S, x \mapsto (g \circ f)(x)$.
30. Definieren Sie den Begriff der inversen Funktion.
31. Geben Sie die inverse Funktion von x^2 auf $[0, \infty[$ an.
32. Definieren Sie den Begriff der linearen Abbildung.

Selbstkontrollfragen

33. Definieren Sie die Begriffe der univariat-und multivariat-reellwertigen Funktion.
34. Definieren Sie Begriff der multivariaten vektorwertigen Funktion.
35. Skizzieren Sie die konstante Funktion für $a := 1$ und die Identitätsfunktion.
36. Für $a = 2$ und $b = 3$, skizzieren Sie die linear Funktion $f(x) = ax + b$.
37. Skizzieren Sie die Funktionen $f(x) := (x - 1)^2$ und $g(x) := (x + 3)^2$
38. Skizzieren Sie die Exponential- und Logarithmusfunktionen.
39. Geben Sie Exponentialeigenschaften der Exponentialfunktion an.
40. Geben Sie die Logarithmeneigenschaften der Logarithmusfunktion an.
41. Definieren Sie den Begriff der Ableitung $f'(a)$ einer Funktion f an einer Stelle a .
42. Definieren den Begriff der Ableitung f' einer Funktion f .
43. Erläutern Sie die Symbole $f'(x)$, $\dot{f}(x)$, $\frac{df(x)}{dx}$, und $\frac{d}{dx} f(x)$.
44. Definieren Sie den Begriff der zweiten Ableitung f'' einer Funktion f .
45. Geben Sie die Summenregel für Ableitungen wieder.
46. Geben Sie die Produktregel für Ableitungen wieder.
47. Geben Sie die Quotientenregel für Ableitungen wieder.
48. Geben Sie die Kettenregel für Ableitungen wieder.

50. Bestimmen Sie die Ableitung der Funktion $f(x) := 3x^2 + \exp(-x^2) - x \ln(x)$.
51. Bestimmen Sie die Ableitung der Funktion $f(x) := \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2$ für $\mu \in \mathbb{R}$.
52. Definieren Sie die Begriffe des globalen und lokalen Maximums/Minimums einer Funktion.
53. Geben Sie die notwendige Bedingung für ein Extremum einer Funktion wieder.
54. Geben Sie die hinreichende Bedingung für ein lokales Extremum einer Funktion wieder.
55. Geben Sie das Standardverfahren der analytischen Optimierung wieder.
56. Bestimmen Sie einen Extremwert von $f(x) := \exp(-\frac{1}{2}(x - \mu)^2)$ für $\mu \in \mathbb{R}$.
57. Definieren Sie den Begriff der Stammfunktion einer univariaten reellwertigen Funktion.
58. Definieren Sie den Begriff des unbestimmten Integrals einer univariaten reellwertigen Funktion.
59. Erläutern Sie den Begriff des Riemanschen Integrals.
60. Geben Sie den ersten Hauptsatz der Differential- und Integralrechnung wieder.
61. Geben Sie den zweiten Hauptsatz der Differential- und Integralrechnung wieder.
62. Erläutern Sie den Begriff des uneigentlichen Integrals.
63. Erläutern Sie den Begriff des mehrdimensionalen Integrals.
64. Berechnen Sie das bestimmte Integral $\int_0^1 2x \, dx$.

Literatur

- Bärwolff, G. (2017). *Höhere Mathematik für Naturwissenschaftler und Ingenieure*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Cantor, G. (1895). Beiträge zur Begründung der transfiniten Mengenlehre. *Mathematische Annalen*, 46(4):481–512.



Inferenzstatistik

BSc Psychologie SoSe 2021

Prof. Dr. Dirk Ostwald

Prof. Dr. Dirk Ostwald (dirk.ostwald@ovgu.de)



Seit 2021	W2 Professur Methodenlehre I
2014 - 2020	W1 Professur Freie Universität Berlin
2010 - 2014	Postdoc BCCN & MPIB Berlin
2007 - 2010	PhD Psychologie Birmingham
2004 - 2006	MSc Neurowissenschaften Tübingen
2005 - 2012	BSc Mathematik Hagen
2000 - 2003	BSc Medizin Hamburg

Forschung	Komputationale Kognitive Neurowissenschaften
Lehre	Datenwissenschaft

(1) Einführung

Einführung

- Datenwissenschaft
- Statistik
- Modul B2 Inferenzstatistik
- Selbstkontrollfragen

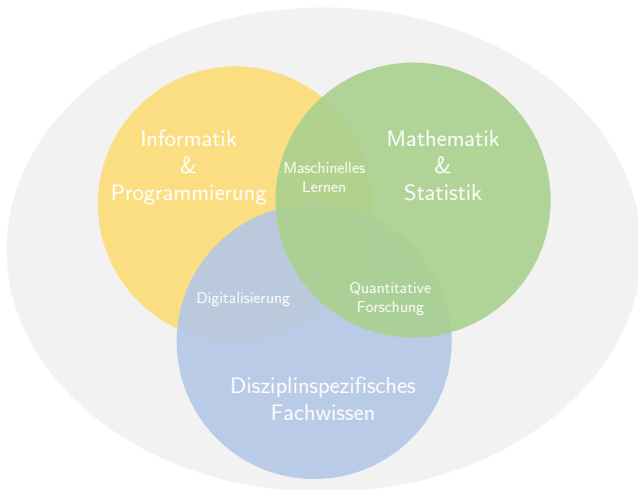
Einführung

- **Datenwissenschaft**
- Statistik
- Modul B2 Inferenzstatistik
- Selbstkontrollfragen

Datenwissenschaft

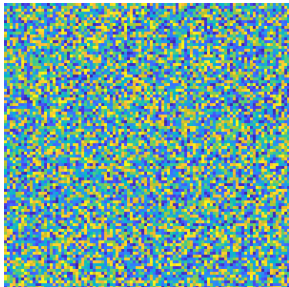
Die Kunst, aus Daten Sinn zu generieren

Datenwissenschaft



Datenanalyse ist Datenreduktion

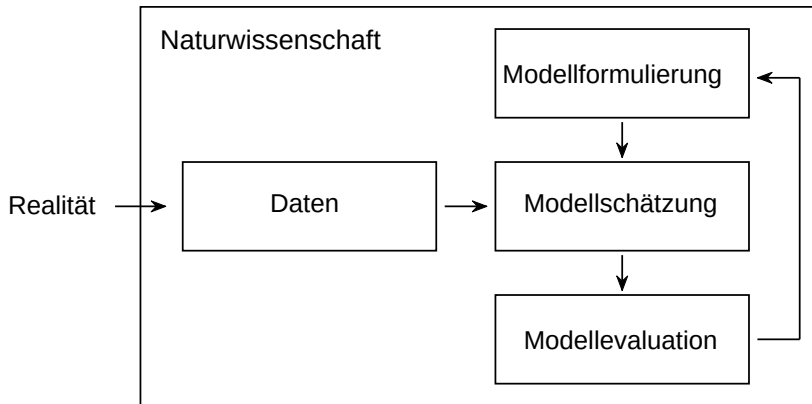
Rohdaten



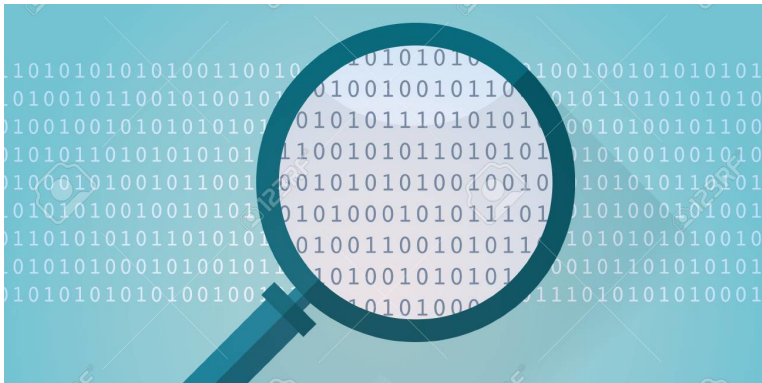
Reduzierte Daten



Datenanalyse ist modellbasiert



Datenanalyse ist Dateninterpretation



Datenwissenschaft

Statistik = Maschinelles Lernen = Künstliche Intelligenz

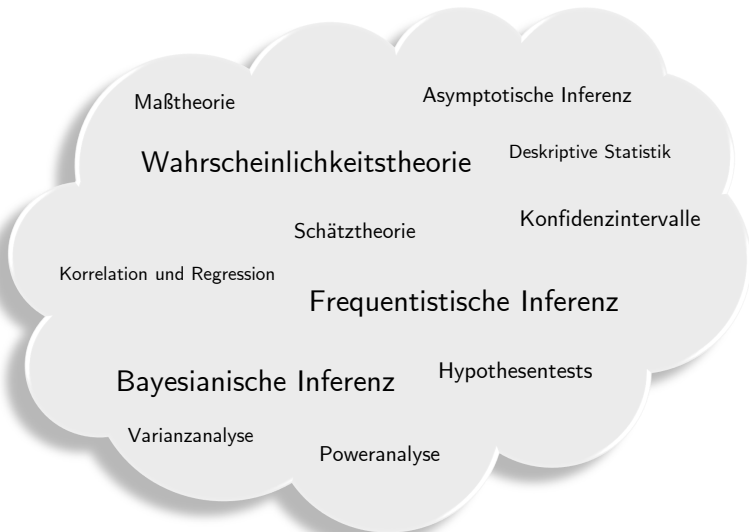
Statistik	Maschinelles Lernen	Künstliche Intelligenz
Probabilistische Modelle	Deterministische Modelle	Agenten-basierte Modelle
Theoretische Analyse	Klassifikation	Reinforcement learning
Optimalitätstheorie	Bayesianische Modelle	Symbolik
Asymptotische Theorie	Anwendung	Anwendung
Wissenschaftsphilosophie	Benchmarking	Hype

Einführung

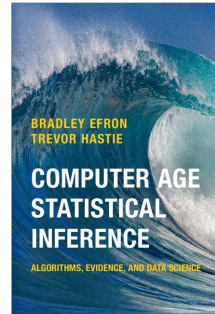
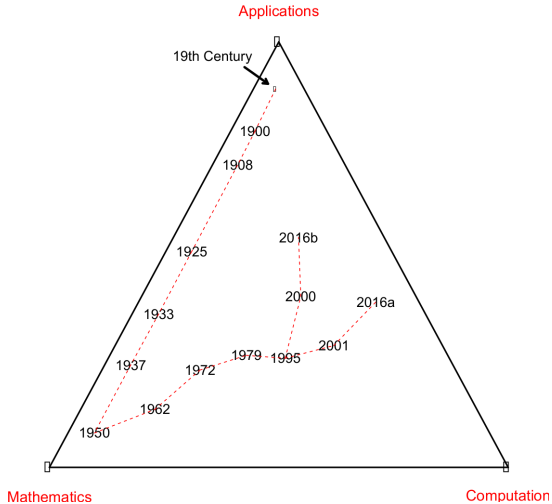
- Datenwissenschaft
- **Statistik**
- Modul B2 Inferenzstatistik
- Selbstkontrollfragen

Statistik

Die Kunst, aus Daten Sinn zu generieren
und seine Unsicherheit zu quantifizieren



Geschichte



Geschichte

1900	Pearsons Chi-Quadrat Test
1908	Studentsche t-Verteilung
1925	Fishers "Statistical Methods for Research Workers"
1933	Neyman und Pearsons Hypothesentesttheorie
1937	Neymans Konfidenzintervalle
1950	Walds Statistische Entscheidungstheorie
1950	Savages & de Finettis Bayesianische Entscheidungstheorie
1961	Raiffa & Schlaifers Statistische Entscheidungstheorie
1962	Tukeys "The future of data analysis"
1971	Lindleys Bayesianische Statistik
1979	Efrons Bootstrap
1995	Benjamini und Hochbergs Falscherkennungsrate
1996	Vapniks Statistische Lerntheorie
2000	Multiples Testen bei Microarray and Neuroimaginganalysen
2003	Beals Variational inference
2010	Renaissance neuronaler Netze als Deep Learning
2015	Datenwissenschaft

Fundamentale Annahmen der Wahrscheinlichkeitstheorie

- Zufallsprozesse können mathematisch modelliert werden.
- Mathematik kann zur Vorhersage zufälliger Ereignisse genutzt werden.
- Die Wahrscheinlichkeitstheorie ist mengentheoretisch begründet.

Fundamentale Annahmen der Frequentistischen Statistik

- Wahrscheinlichkeiten spiegeln die relative Frequenz des Auftretens eines zufälligen Ereignisses und beschreiben objektive Eigenschaften der realen Welt.
- Die Parameter probabilistischer Modelle sind feste, unbekannte Konstanten, die als *wahre, aber unbekannte, Parameterwerte* bezeichnet werden. Über Parameterwerte und Modelle werden keine probabilistischen Aussagen getroffen.
- Statistische Methoden werden so gestaltet, dass sie gute langfristige relative Frequenzeigenschaften besitzen und werden typischerweise anhand ihrer Stichprobenverteilungen bewertet.

Im BSc Psychologie lernen Sie nur Methoden der Frequentistischen Statistik kennen.

Fundamentale Annahmen der Bayesianischen Statistik

- Wahrscheinlichkeiten werden als Grade der Sicherheit, nicht als langfristige relative Häufigkeiten interpretiert. Aussagen der Form “Die Wahrscheinlichkeit, dass das Wintersemester 2021/22 in Präsenzlehre stattfindet, ist 0.7” haben eine Bedeutung.
- Die Parameter probabilistischer Modelle sind feste, unbekannte Konstanten, die als *wahre, aber unbekannte, Parameterwerte* bezeichnet werden. Über Parameterwerte und Modelle werden probabilistische Aussagen getroffen, die unseren Grad an Sicherheit hinsichtlich ihrer quantitativen Ausprägung und Validität widerspiegeln.
- Probabilistische Aussagen über Parameter werden mithilfe von Wahrscheinlichkeitsverteilungen getroffen, auf deren Grundlage optimale Entscheidungen im Sinne von Kosten-Nutzenfunktionen getroffen werden können.

Im MSc Psychologie lernen Sie auch Methoden der Bayesianischen Statistik kennen.

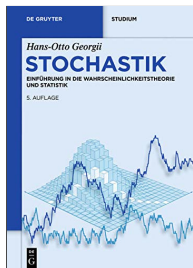
Einführung

- Datenwissenschaft
- Statistik
- **Modul B2 Inferenzstatistik**
- Selbstkontrollfragen

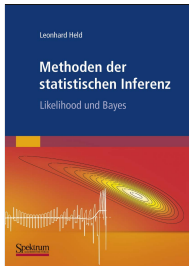
Modul B2 Inferenzstatistik Sommersemester 2021

Einheit	Datum	Thema
(0) Mathematische Grundlagen	Ad libitum	Allgemeine Grundlagen
(1) Einführung	07.04.2021	Allgemeine Grundlagen
(2) Wahrscheinlichkeitsräume und Zufallsvariablen	14.04.2021	Wahrscheinlichkeitstheorie
(3) Multivariate Verteilungen und Stichproben	21.04.2021	Wahrscheinlichkeitstheorie
(4) Erwartungswert und Kovarianz	28.04.2021	Wahrscheinlichkeitstheorie
(5) Ungleichungen und Grenzwerte	05.05.2021	Wahrscheinlichkeitstheorie
(6) Transformationen der Normalverteilung	12.05.2021	Wahrscheinlichkeitstheorie
(7) Parameterschätzung	19.05.2021	Wahrscheinlichkeitstheorie
(8) Schätzeigenschaften	26.05.2021	Grundlagen der Statistik
(9) Konfidenzintervalle	02.06.2021	Grundlagen der Statistik
(10) Hypothesentests	09.06.2021	Grundlagen der Statistik
(11) T-Tests	16.06.2021	Spezielle Verfahren
(12) Poweranalyse	23.06.2021	Spezielle Verfahren
(13) Einfaktorielle Varianzanalyse	30.06.2021	Spezielle Verfahren
(14) Zweifaktorielle Varianzanalyse	07.07.2021	Spezielle Verfahren
Klausur	Juli	
Wiederholungstermin	September	

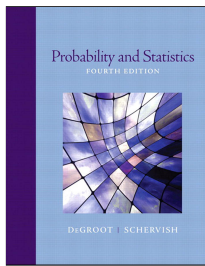
Weiterführende Literatur



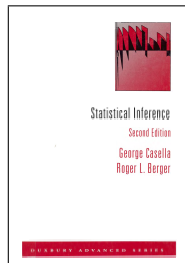
Georgii (2009)



Held and Sabanés Bové
(2014)



DeGroot and Schervish
(2012)



Casella and Berger (2012)

Ziele des Moduls

- Erwerben von grundlegenden Kenntnisse der Inferenzstatistik
- Tieferes Verständnis statistisch gesicherter wissenschaftlicher Erkenntnis
- Förderung des logischen, quantitativen, und methodenkritischen Denkens
- Theoretischer Hintergrund zu Modul C2 Computergestützte Datenanalyse

Keine Ziele des Moduls

- Wecken des Interesses an Wissenschaft oder psychologische Datenanalyse
⇒ Beides bringen Sie schon mit, sonst wären Sie nicht hier
- Extrinsische Lernmotivation
⇒ Wir setzen intrinsische Motivation voraus
- Durch Mathematikverzicht unverständliche “Statistik für Psycholog:innen”

Organisation des Moduls

- Synchrone Onlinevorlesung mittwochs 9 - 12 Uhr (7.4.2021 - 7.7.2021)
- Vorlesungsfolien und Vorlesungsvideolinks auf Abteilungswebseite
- Interaktive Q&A Session donnerstags 11 - 13 Uhr (8.4.2021 - 8.7.2021)
- Selbstkontrollfragen zur selbständigen Bearbeitung

Leistungsnachweis

- Klausur am Ende des Semesters
- Prüfungsnote entspricht der Klausurnote
- Klausurfragen angelehnt an Selbstkontrollfragen

Einige Ratschläge

- Bearbeiten Sie die Inhalte anhand der Selbstkontrollfragen wöchentlich aktiv nach.
⇒ Schauen Sie sich das Vorlesungsvideo ruhig nochmal an.
- Lernen Sie ihre Antworten auf die Selbstkontrollfragen auswendig.
⇒ Der Kurs ist sehr hierarchisch, Wissen aus Woche k wird in Woche $k + 1$ benötigt.
⇒ Mathematik ist auch nur Detailwissen!
- Lesen Sie alternative Darstellungen aus den Literaturhinweisen oder auf Wikipedia.
⇒ Prüfungsrelevant ist allerdings allein die Darstellung im Kurs.
- Bilden Sie Lerngruppen und/oder besuchen Sie das Tutorium.
⇒ Üben Sie, das Gelernte aktiv und laut wiederzugeben.
- Bereiten Sie Fragen für die Q&A Session am Donnerstagvormittag vor.

Fragen?

Einführung

- Datenwissenschaft
- Statistik
- Modul B2 Inferenzstatistik
- **Selbstkontrollfragen**

1. Diskutieren Sie den Begriff "Datenwissenschaft".
2. Nennen und erläutern Sie drei zentrale Charakteristika der Datenanalyse.
3. Diskutieren Sie Gemeinsamkeiten und Unterschiede von Statistik, Maschinellem Lernen und künstlicher Intelligenz.
4. Nennen und erläutern Sie drei fundamentale Annahmen der Wahrscheinlichkeitstheorie.
5. Nennen und erläutern Sie drei fundamentale Annahmen der Frequentistischen Statistik.
6. Nennen und erläutern Sie drei fundamentale Annahmen der Bayesianischen Statistik.

Literatur

Casella, G. and Berger, R. (2012). *Statistical Inference*. Duxbury.

DeGroot, M. H. and Schervish, M. J. (2012). *Probability and Statistics*. Addison-Wesley, Boston, 4th ed edition.

Georgii, H.-O. (2009). *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. DeGruyter-Lehrbuch. de Gruyter, Berlin, 4., überarb. und erw. Aufl. edition.

Held, L. and Sabanés Bové, D. (2014). *Applied Statistical Inference*. Springer Berlin Heidelberg, Berlin, Heidelberg.



Inferenzstatistik

BSc Psychologie SoSe 2021

Prof. Dr. Dirk Ostwald

(2) Wahrscheinlichkeitsräume und Zufallsvariablen

Literaturhinweise

Die in diesem Abschnitt vorgestellten Konzepte gehören zum Standard der Wahrscheinlichkeitstheorie. Zu Wahrscheinlichkeitsräumen wird die Lektüre von Georgii (2009, Abschnitte 1.1, 1.2, 3.1 und 3.3) empfohlen. Zu Zufallsvariablen wird die Lektüre von Georgii (2009, Abschnitte 1.3, 2.1, 2.6) empfohlen. In englischer Sprache finden sich entsprechende Darstellungen in Wasserman (2004, Sections 2.1 - 2.8) und DeGroot and Schervish (2012, Chapter 3).

Wahrscheinlichkeitsräume und Zufallsvariablen

- Wahrscheinlichkeitsräume
- Elementare Wahrscheinlichkeiten
- Konstruktion, Definition, und Notation von Zufallsvariablen
- Wahrscheinlichkeitsmassenfunktionen
- Wahrscheinlichkeitsdichtefunktionen
- Kumulative Verteilungsfunktionen
- Selbstkontrollfragen

Wahrscheinlichkeitsräume und Zufallsvariablen

- **Wahrscheinlichkeitsräume**
- Elementare Wahrscheinlichkeiten
- Konstruktion, Definition, und Notation von Zufallsvariablen
- Wahrscheinlichkeitsmassefunktionen
- Wahrscheinlichkeitsdichtefunktionen
- Kumulative Verteilungsfunktionen
- Selbstkontrollfragen

Definition (Wahrscheinlichkeitsraum)

Ein *Wahrscheinlichkeitsraum* (*W-Raum*) ist ein Triple $(\Omega, \mathcal{A}, \mathbb{P})$, wobei

- Ω eine Menge *Ergebnismenge* ist, d.h. eine Menge von Elementarereignissen ω ,
- $\mathcal{A}$ eine σ -Algebra ist, d.h., $\mathcal{A}$ ist eine Menge mit den Eigenschaften
 - $\Omega \in \mathcal{A}$,
 - $\mathcal{A}$ ist abgeschlossen unter der Bildung von Komplementärereignissen, d.h., wenn $A \in \mathcal{A}$, dann ist auch $A^c := \Omega \setminus A \in \mathcal{A}$ für alle $A \in \mathcal{A}$,
 - $\mathcal{A}$ ist abgeschlossen unter der abzählbaren Vereinigung von Ereignissen, d.h., wenn $A_1, A_2, \dots \in \mathcal{A}$, dann folgt $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$, und
- $\mathbb{P}$ ein *Wahrscheinlichkeitsmaß* ist, d.h. eine Abbildung der Form $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ mit
 - $\mathbb{P}(A) \geq 0$ für alle $A \in \mathcal{A}$ (Nicht-Negativität)
 - $\mathbb{P}(\Omega) = 1$ (Normiertheit)
 - $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ für paarweise disjunkte $A_i \in \mathcal{A}$ (σ -Additivität).

Wahrscheinlichkeitsräume

ERGEBNISSE DER MATHEMATIK UND IHRER GRENZGEBIETE

HERAUSGEGEBEN VON DER SCHRIFTLEITUNG
DES
„ZENTRALBLATT FÜR MATHEMATIK“
ZWEITER BAND

GRUNDBEGRIFFE DER WAHRSCHEINLICHKEITS- RECHNUNG

VON
A. KOLMOGOROFF



BERLIN
VERLAG VON JULIUS SPRINGER
1933

306/ 1. Terminologische Vorbemerkungen

erniedrigt. Ist jedoch die Anzahl der Behauptungen sehr groß, so lassen sich aus der praktischen Sicherheit jeder einzelnen dieser Behauptungen in bezug auf die Richtigkeit der simultanen Behauptung überhaupt keine Schlüsse ziehen. Deshalb folgt aus dem Prinzip A nach folgendem Weg, daß bei einer sehr großen Anzahl von Serien von Versuchen, von denen jede Serie aus n Versuchen besteht, in jeder Serie der Quotient n wie sich von $P(A)$ wenig unterscheiden wird.

Bemerkung II. Dem unmöglichen Ereignis (des leeren Menge) entspricht kraft unserer Axiome die Wahrscheinlichkeit $P(\emptyset) = 0$ *, während umgekehrt aus $P(A) = 0$ die Unmöglichkeit des Ereignisses A durchaus nicht zu folgen braucht; nach dem Prinzip B folgt aus dem Nullwerden der Wahrscheinlichkeit nur, daß bei einer einmaligen Realisation der Bedingungen $\bar{G}$ das Ereignis A praktisch unmöglich ist. Das bedeutet jedoch keineswegs, daß auch bei einer genügend langen Reihe von Versuchen das Ereignis A nicht auftreten wird. Andererseits kann man nach dem Prinzip A nur behaupten, daß bei $P(A) = 0$ und sehr großen n der Quotient n sehr klein wird (er kann z. B. gleich 1 sein).

§ 3. Terminologische Vorbemerkungen.

Wir haben die eigentlichen Objekte unserer weiteren Betrachtungen — die zufälligen Ereignisse — als Mengen definiert. Mehrere mengentheoretische Begriffe bezeichnet man aber in der Wahrscheinlichkeitsrechnung mit anderen Namen. Wir wollen hier ein kurzes Verzeichnis solcher Begriffe geben.

Mengentheoretisch.

1. A und B sind disjunkt, d. h.

$AB = \emptyset$.

2. $AB \dots N = \emptyset$.

3. $AB \dots N = X$.

4. $A + B + \dots + N = X$.

5. Die Komplementärmenge $\bar{A}$.

6. $A = \emptyset$.

7. $A = E$.

* Vgl. § 1. Formel (10).

Im Falle der zufälligen Ereignisse.

1. Die Ereignisse A und B sind

unvereinbar.

2. Die Ereignisse $A, B, \dots, N$

sind unvereinbar.

3. Das Ereignis X besteht in der

gleichzeitigen Realisation aller

Ereignisse $A, B, \dots, N$.

4. Das Ereignis X besteht in der Er-

scheinung mindestens eines un-

ter den Ereignissen $A, B, \dots, N$.

5. Das entgegengesetzte Ereignis

$\bar{A}$ besteht in der Nichter-

scheinung des Ereignisses A .

6. A ist unmöglich.

7. A muß notwendig vorkommen.

6. 1. Die elementare Wahrscheinlichkeitsrechnung. /306

8. Ein System $\mathcal{H}$ der Mengen $A_1,$

$A_2, \dots, A_n$ bildet eine Zer-

legung der Menge E , wenn

$A_1 + A_2 + \dots + A_n = E$ ist

(das setzt bereits voraus, daß

die Mengen A_i paarweise dis-

junkt sind).

9. B ist eine Untermenge von A :

$B \subset A$.

8. Ein Versuch $\mathcal{H}$ besteht darin,

daß man feststellt, welches

unter den Ereignissen $A_1,$

$A_2, \dots, A_n$ vorkommt. $A_1,$

$A_2, \dots, A_n$ sind die möglichen

Ausgänge des Versuches $\mathcal{H}$.

9. Aus der Realisation des Er-

gebnisses B folgt notwendig

dieselbe von A .

§ 4. Unmittelbare Folgerungen aus den Axiomen, bedingte Wahrscheinlichkeiten, der Satz von BAYES.

Aus $A + A = E$ und den Axiomen IV und V folgt

(1) $P(A) + P(\bar{A}) = 1$.

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

Da $E = \emptyset$ ist, erhält man insbesondere

(3) $P(\emptyset) = 0$.

Wenn $A, B, \dots, N$ unvereinbar sind, so folgt aus dem Axiom IV die Formel

(4) $P(A + B + \dots + N) = P(A) + P(B) + \dots + P(N)$

(der Additionssatz).

Wenn $P(A) > 0$ ist, so nennt man den Quotienten

(5) $P_A(B) = \frac{P(AB)}{P(A)}$

die bedingte Wahrscheinlichkeit des Ereignisses B unter der Bedingung A .

Aus (5) folgt unmittelbar

(6) $P(AB) = P(A)P_A(B)$.

Ein Induktionsschluß ergibt sodann die allgemeine Formel

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

(der Multiplikationssatz).

Man beweist auch leicht folgende Formeln:

(8) $P_A(B) \geq 0$.

(9) $P_A(B) \leq 1$.

(10) $P_A(B + C) = P_A(B) + P_A(C)$.

Vergleicht man diese Formeln (8) bis (10) mit den Axiomen III bis V,

so ergibt sich, daß das Mengensystem $\mathcal{G}$ mit der Mengenfunktion $P_A(B)$

“Zweck des vorliegenden Heftes ist eine axiomatische Begründung der Wahrscheinlichkeitsrechnung. Der leitende Gedanke des Verfassers war dabei, die Grundbegriffe der Wahrscheinlichkeitsrechnung, welche noch unlängst für ganz eigenartig galten, natürlicherweise in die Reihe der allgemeinen Begriffsbildungen der modernen Mathematik* einzuordnen.” (*z.B. Mengenlehre, Abbildungen, Lebesgue Integral)

Kolmogorov, A.N. (1933) Grundbegriffe der Wahrscheinlichkeitsrechnung

Ausgewählte Bemerkungen

- W-Räume sind abstrakte Modelle von Zufallsexperimenten.
- W-Räume sind spezielle *Maßräume* $(\Omega, \mathcal{A}, \mu)$.
- Maßräume sind mathematische Modelle zur Beschreibung der Volumina von Mengen.
- Tuple von Ereignismengen und σ -Algebren heißen *Messräume*.
- W-Räume sind Messräume mit Wahrscheinlichkeitsmaßen.
- Elementarereignisse werden anhand von $\mathbb{P}(\{\omega\})$ “gezogen”.
- W-Räume vereinheitlichen endliche, abzählbare, und überabzählbare Ergebnismengen.
- W-Räume vereinheitlichen diskrete und kontinuierliche Wahrscheinlichkeit.
- σ -Algebren sind “vollständige Mengen von Ereignissen”.
- Für endliche Ereignismengen ist die typische σ -Algebra die Potenzmenge $\mathcal{P}(\Omega)$.
- Für die Ereignismenge $\mathbb{R}$ erfüllt die *Borel σ -algebra* die gleiche Funktion.
- Die W-theoretischen Charakteristiken von $(\Omega, \mathcal{A}, \mathbb{P})$ werden durch $\mathbb{P}$ festgelegt.
- W-Räume verschwinden oft hinter Zufallsvariablen und Verteilungen.

Beispiel (Würfel)

- Es ist sinnvoll, die Ereignismenge als $\Omega := \{1, 2, 3, 4, 5, 6\}$ zu definieren.
- Auch die Definition $\Omega := \{\cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot\}$ wäre möglich.
- Die Potenzmenge $\mathcal{P}(\{1, 2, 3, 4, 5, 6\})$ enthält alle möglichen Ereignisse A_i , z.B.

Es fällt eine beliebige Augenzahl	$A_1 = \Omega$
Es fällt eine Augenzahl größer als 4	$A_2 = \{5, 6\}$
Es fällt eine gerade Augenzahl	$A_3 = \{2, 4, 6\}$
Es fällt eine Sechs	$A_4 = \{6\}$
Eins, drei, oder sechs fällt	$A_5 = \{1, 3, 6\}$
Es fällt keine Augenzahl	$A_6 = \emptyset$

- $\mathbb{P}$ kann durch Festlegung von $\mathbb{P}(\{\omega\})$ für alle $\omega \in \Omega$ definiert $\rightarrow$, die Wahrscheinlichkeiten aller anderen Ereignisse $A \in \mathcal{A}$ können dann mit der σ -Additivität $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$ errechnet werden, weil die $\omega \in \mathcal{A}$ mit $\omega \in \Omega$ paarweise disjunkt sind.
- Im Modell eines unverfälschten Würfels wäre $\mathbb{P}(\{\omega\}) := 1/6$ für alle $\omega \in \Omega$.
- Für das Modell eines verfälschten Würfels könnte zum Beispiel definiert werden, dass $\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \mathbb{P}(\{6\}) := 1/9$, und $\mathbb{P}(\{3\}) = \mathbb{P}(\{4\}) = \mathbb{P}(\{5\}) := 2/9$.

Wahrscheinlichkeitsräume und Zufallsvariablen

- Wahrscheinlichkeitsräume
- **Elementare Wahrscheinlichkeiten**
- Konstruktion, Definition, und Notation von Zufallsvariablen
- Wahrscheinlichkeitsmassefunktionen
- Wahrscheinlichkeitsdichtefunktionen
- Kumulative Verteilungsfunktionen
- Selbstkontrollfragen

Definition (Wahrscheinlichkeitsmaß)

Es seien Ω eine Ereignismenge und $\mathcal{A}$ eine σ -Algebra auf Ω . Eine Abbildung

$$\mathbb{P} : \mathcal{A} \rightarrow [0, 1], A \mapsto \mathbb{P}(A) \quad (1)$$

mit den Eigenschaften

- (1) $\mathbb{P}(A) \geq 0$ für alle $A \in \mathcal{A}$,
- (2) $\mathbb{P}(\Omega) = 1$, und
- (3) wenn $A_1, A_2, \dots \in \mathcal{A}$ paarweise disjunkt sind, dann gilt $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

heißt ein *Wahrscheinlichkeitsmaß* oder einfach *Wahrscheinlichkeit*.

Bemerkungen

- Für $A \in \mathcal{A}$ heißt $\mathbb{P}(A)$ die *(totale) Wahrscheinlichkeit von A*.
- Für $A, B \in \mathcal{A}$ heißt $\mathbb{P}(A \cap B)$ die *gemeinsame Wahrscheinlichkeit von A und B*.

Zur Interpretation des Wahrscheinlichkeitsbegriffs

- Die Interpretation des Wahrscheinlichkeitsbegriffs $\mathbb{P}(A)$ ist nicht eindeutig.
- Nach der *frequentistischen Interpretation* ist $\mathbb{P}(A)$ die idealisierte relative Häufigkeit, mit der das Ereignis A unter den gleichen äußeren Bedingungen eintreten pflegt.
- Nach der *Bayesianischen Interpretation* ist $\mathbb{P}(A)$ ist der Grad der Sicherheit, den ein Beobachter aufgrund seiner subjektiven Einschätzung der Lage dem Eintreten des Ereignisses A zumisst.
- Frequentistische und Bayesianische Interpretationen basieren mit der Wahrscheinlichkeitstheorie auf dem gleichen mathematischen Bezugssystem.

Theorem (Eigenschaften von Wahrscheinlichkeiten)

Es sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein W-Raum und $A, B \in \mathcal{A}$. Dann gelten unter anderem

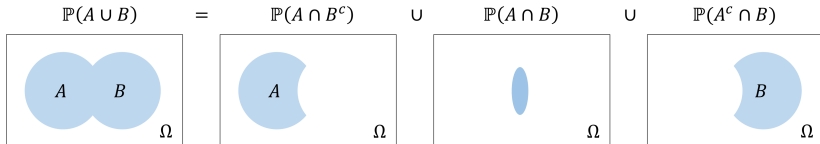
- (1) $\mathbb{P}(\emptyset) = 0$ und $0 \leq \mathbb{P}(A) \leq 1$.
- (2) $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$.
- (3) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
- (4) $A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
- (5) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Beweis von (5)

Mit der Additivität von $\mathbb{P}$ für disjunkte Ereignisse gilt

$$\begin{aligned}\mathbb{P}(A \cup B) &= \mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B) \\ &= \mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B) + \mathbb{P}(A \cap B) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}((A \cap B^c) \cup (A \cap B)) + \mathbb{P}((A^c \cap B) \cup (A \cap B)) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)\end{aligned}\tag{2}$$

Zur Veranschaulichung der ersten Gleichung



□

Definition (Unabhängige Ereignisse)

Zwei Ereignisse $A \in \mathcal{A}$ und $B \in \mathcal{A}$ heißen (*stochastisch*) *unabhängig*, wenn

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \quad (3)$$

Eine Menge von Ereignissen $\{A_i | i \in I\} \subset \mathcal{A}$ mit beliebiger Indexmenge I heißt (*stochastisch*) *unabhängig*, wenn für jede endliche Untermenge $J \subseteq I$ gilt, dass

$$\mathbb{P}(\cap_{j \in J} A_j) = \prod_{j \in J} \mathbb{P}(A_j). \quad (4)$$

Bemerkungen

- Stochastische Unabhängigkeit modelliert die Abwesenheit von gegenseitigen Einflüssen.
- Die Unabhängigkeit bestimmter Ereignissen kann in der Definition eines probabilistischen Modells vorausgesetzt werden, so zum Beispiel die Unabhängigkeit von Fehlervariablen.
- Die Unabhängigkeit von Ereignissen kann aus der Definition eines probabilistischen Modells folgen.
- Disjunkte Ereignisse mit von Null verschiedener Wahrscheinlichkeit sind nie unabhängig:

$$\mathbb{P}(A)\mathbb{P}(B) > 0, \text{ aber } \mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0, \text{ also } \mathbb{P}(A \cap B) \neq \mathbb{P}(A)\mathbb{P}(B).$$

- Die Bedingung der beliebigen Untermengen von I sichert die paarweise unabhängig der $A_i, i \in I$.

Definition (Bedingte Wahrscheinlichkeit)

Es sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein W-Raum und $A, B \in \mathcal{A}$ mit $\mathbb{P}(B) > 0$. Die *bedingte Wahrscheinlichkeit des Ereignisses $A \in \mathcal{A}$ gegeben Ereignis $B \in \mathcal{A}$* ist definiert als

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (5)$$

Für jedes feste B , ist $\mathbb{P}(\cdot|B)$ ein Wahrscheinlichkeitsmaß, d.h. $\mathbb{P}(\cdot|B) \geq 0$, $\mathbb{P}(\Omega|B) = 1$, und für paarweise disjunkte $A_1, A_2, \dots \in \mathcal{A}$ gilt $\mathbb{P}(\cup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} \mathbb{P}(A_i|B)$.

Anmerkungen

- $\mathbb{P}(A|B)$ ist die normalisierte Version der gemeinsamen Wahrscheinlichkeit $\mathbb{P}(A \cap B)$.
- Die Festlegung der gemeinsamen Wahrscheinlichkeit $\mathbb{P}(A \cap B)$ legt $\mathbb{P}(A|B)$ fest.
- Die Rechengeln der Wahrscheinlichkeit gelten für die Ereignisse links des Strichs.
- Üblicherweise gilt $\mathbb{P}(A|B) \neq \mathbb{P}(B|A)$, z.B. $\mathbb{P}(\text{Tod}|\text{Erhängen}) \neq \mathbb{P}(\text{Erhängen}|\text{Tod})$.
- Eine Verallgemeinerung für $\mathbb{P}(B) = 0$ ist möglich, aber technisch aufwändig.

Beispiel (Bedingte Wahrscheinlichkeit)

Wir betrachten das Wahrscheinlichkeitsraum Modell des fairen Würfels und wollen die bedingte Wahrscheinlichkeit für das Ereignis “Es fällt eine gerade Augenzahl” gegeben das Ereignis “Es fällt eine Zahl größer als 3” berechnen. Wir halten zunächst fest, dass das Ereignis “Es fällt eine gerade Augenzahl” der Teilmenge $A := \{2, 4, 6\}$ von Ω entspricht, und dass das Ereignis “Es fällt eine Zahl größer als 3” der Teilmenge $B := \{4, 5, 6\}$ von Ω entspricht. Wir halten dann fest, dass unter der Annahme des Modell des fairen Würfels gilt, dass

$$\mathbb{P}(\{2, 4, 6\}) = \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = 1/6 + 1/6 + 1/6 = 3/6 \quad (6)$$

$$\mathbb{P}(\{4, 5, 6\}) = \mathbb{P}(\{4\}) + \mathbb{P}(\{5\}) + \mathbb{P}(\{6\}) = 1/6 + 1/6 + 1/6 = 3/6. \quad (7)$$

Schließlich halten wir fest, dass das Ereignis $A \cap B$, also das Ereignis “Es fällt eine gerade Augenzahl, die größer als 3 ist”, die Wahrscheinlichkeit

$$\mathbb{P}(A \cap B) = \mathbb{P}(\{2, 4, 6\} \cap \{4, 5, 6\}) = \mathbb{P}(\{4, 6\}) = \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = 1/6 + 1/6 = 2/6 \quad (8)$$

hat. Nach Definition der bedingten Wahrscheinlichkeit ergibt sich dann

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\{4, 6\})}{\mathbb{P}(\{4, 5, 6\})} = \frac{2/6}{3/6} = 2/6 \cdot 6/3 = 2/3. \quad (9)$$

Wenn man also weiß, dass eine Augenzahl größer als 3 gefallen ist, ist die Wahrscheinlichkeit, dass es sich dabei um eine gerade Augenzahl handelt $2/3$. Wenn man ersteres nicht weiß, ist die Wahrscheinlichkeit für das Fallen einer geraden Augenzahl (nur) $1/2$.

Theorem (Bedingte Wahrscheinlichkeit unter Unabhängigkeit)

Es seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein W-Raum und $A, B \in \mathcal{A}$ unabhängige Ereignisse mit $\mathbb{P}(\cdot|B) \geq 0$. Dann gilt

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A). \quad (10)$$

Bemerkungen

- Bei Unabhängigkeit ändert das Wissen um das Eintreten von B die Wahrscheinlichkeit des Eintretens von A nicht.
- Unabhängigkeit wird als $\mathbb{P}(A)\mathbb{P}(B)$ modelliert, damit $\mathbb{P}(A|B) = \mathbb{P}(A)$ gilt.

Wahrscheinlichkeitsräume und Zufallsvariablen

- Wahrscheinlichkeitsräume
- Elementare Wahrscheinlichkeiten
- **Konstruktion, Definition, und Notation von Zufallsvariablen**
- Wahrscheinlichkeitsmassefunktionen
- Wahrscheinlichkeitsdichtefunktionen
- Kumulative Verteilungsfunktionen
- Selbstkontrollfragen

Konstruktion von Zufallsvariablen und Verteilungen

- Es seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein W-Raum und $X : \Omega \rightarrow \mathcal{X}$ eine Abbildung.
- Es sei $\mathcal{S}$ eine σ -Algebra auf $\mathcal{X}$.
- Für jedes $S \in \mathcal{S}$ sei das *Urbild von S* definiert als

$$X^{-1}(S) := \{\omega \in \Omega \mid X(\omega) \in S\}. \quad (11)$$

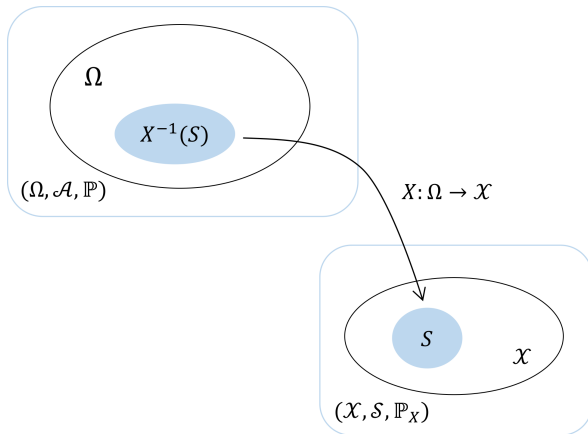
- Wenn $X^{-1}(S) \in \mathcal{A}$ für alle $S \in \mathcal{S}$ gilt, dann heißt X *messbar*.
- $X : \Omega \rightarrow \mathcal{X}$ sei messbar. Allen $S \in \mathcal{S}$ kann die Wahrscheinlichkeit

$$\mathbb{P}_X : \mathcal{S} \rightarrow [0, 1], S \mapsto \mathbb{P}_X(S) := \mathbb{P}(X^{-1}(S)) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in S\}) \quad (12)$$

zugeordnet werden.

- X heißt nun *Zufallsvariable* und $\mathbb{P}_X$ heißt *Bildmaß* oder *Verteilung von X* .
- $(\mathcal{X}, \mathcal{S}, \mathbb{P}_X)$ ist ein W-Raum.
- Mit $\mathcal{X} := \mathbb{R}$ und $\mathcal{S} := \mathcal{B}$ rückt der W-Raum $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$ ins Zentrum der Betrachtung.

Konstruktion von Zufallsvariablen und Verteilungen



$$\mathbb{P}(X^{-1}(S)) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \in S\}) =: \mathbb{P}_X(S)$$

Definition (Zufallsvariable)

Es sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein W-Raum und $(\mathcal{X}, \mathcal{S})$ ein *Messraum*. Eine *Zufallsvariable* (ZV) ist dann definiert als eine Abbildung $X : \Omega \rightarrow \mathcal{X}$ mit der *Messbarkeitseigenschaft*

$$\{\omega \in \Omega | X(\omega) \in S\} \in \mathcal{A} \text{ für alle } S \in \mathcal{S}. \quad (13)$$

Bemerkungen

- ZVen sind weder “zufällig” noch “Variablen”.
- Intuitiv wird $\omega \in \Omega$ “zufällig” anhand von $\mathbb{P}$ gezogen und $X(\omega)$ realisiert.
- Wir nennen $\mathcal{X}$ den *Ergebnisraum der ZV* X .
- Die Verteilungen (Bildmaße) von ZVen sind in der Statistik zentral.
- Der Begriff der Verteilung wird oft auch für W-Maße und Dichten verwendet.

Beispiel (Summe zweier Würfel)

- Für das Werfen zweier Würfel ist ein sinnvolles Wahrscheinlichkeitsraum-Modell
 - $\Omega := \{(d_1, d_2) | d_1 \in \mathbb{N}_6, d_2 \in \mathbb{N}_6\}$
 - $\mathcal{A} := \mathcal{P}(\Omega)$.
 - $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ mit $\mathbb{P}(\{(d_1, d_2)\}) = 1/36$ für alle $(d_1, d_2) \in \Omega$.
- Die Augenzahl-Summenbildung wird dann sinnvoller Weise durch die Zufallsvariable

$$X : \Omega \mapsto \mathcal{X}, (d_1, d_2) \mapsto X((d_1, d_2)) := d_1 + d_2. \quad (14)$$

beschrieben, wobei $\mathcal{X} := \{2, 3, \dots, 12\}$.

- $\mathcal{S} := \mathcal{P}(\mathcal{X})$ ist eine sinnvolle σ -Algebra auf $\mathcal{X}$.
- Mithilfe der σ -Additivität von $\mathbb{P}$ können wir die Verteilung $\mathbb{P}_X$ von X für alle Elementarereignisse $\{x\} \in \mathcal{S}$ berechnen, wie auf der nächsten Folie gezeigt.
- Wir haben damit ein weiteres Wahrscheinlichkeitsraum-Modell $(\mathcal{X}, \mathcal{S}, \mathbb{P}_X)$ konstruiert.

Beispiel (Summe zweier Würfel)

Bestimmung der Verteilung $\mathbb{P}_X$ von X

$$\mathbb{P}_X(\{2\}) = \mathbb{P}(X^{-1}(\{2\})) = \mathbb{P}(\{(1, 1)\}) = \frac{1}{36}$$

$$\mathbb{P}_X(\{3\}) = \mathbb{P}(X^{-1}(\{3\})) = \mathbb{P}(\{(1, 2), (2, 1)\}) = \frac{2}{36}$$

$$\mathbb{P}_X(\{4\}) = \mathbb{P}(X^{-1}(\{4\})) = \mathbb{P}(\{(1, 3), (3, 1), (2, 2)\}) = \frac{3}{36}$$

$$\mathbb{P}_X(\{5\}) = \mathbb{P}(X^{-1}(\{5\})) = \mathbb{P}(\{(1, 4), (4, 1), (2, 3), (3, 2)\}) = \frac{4}{36}$$

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

$$\mathbb{P}_X(\{7\}) = \mathbb{P}(X^{-1}(\{7\})) = \mathbb{P}(\{(1, 6), (6, 1), (2, 5), (5, 2), (3, 4), (4, 3)\}) = \frac{6}{36}$$

$$\mathbb{P}_X(\{8\}) = \mathbb{P}(X^{-1}(\{8\})) = \mathbb{P}(\{(2, 6), (6, 2), (3, 5), (5, 3), (4, 4)\}) = \frac{5}{36}$$

$$\mathbb{P}_X(\{9\}) = \mathbb{P}(X^{-1}(\{9\})) = \mathbb{P}(\{(3, 6), (6, 3), (4, 5), (5, 4)\}) = \frac{4}{36}$$

$$\mathbb{P}_X(\{10\}) = \mathbb{P}(X^{-1}(\{10\})) = \mathbb{P}(\{(4, 6), (6, 4), (5, 5)\}) = \frac{3}{36}$$

$$\mathbb{P}_X(\{11\}) = \mathbb{P}(X^{-1}(\{11\})) = \mathbb{P}(\{(5, 6), (6, 5)\}) = \frac{2}{36}$$

$$\mathbb{P}_X(\{12\}) = \mathbb{P}(X^{-1}(\{12\})) = \mathbb{P}(\{(6, 6)\}) = \frac{1}{36}$$

Definition (Notation für ZVen)

Es seien $(\Omega, \mathcal{A}, \mathbb{P})$ und $(\mathcal{X}, \mathcal{S}, \mathbb{P}_X)$ W-Räume und $X : \Omega \rightarrow \mathcal{X}$ sei eine ZV. Dann gelten folgende Konventionen:

$$\{X \in S\} := \{\omega \in \Omega | X(\omega) \in S\}, S \subset \mathcal{X},$$

$$\{X = x\} := \{\omega \in \Omega | X(\omega) = x\}, x \in \mathcal{X},$$

$$\{X \leq x\} := \{\omega \in \Omega | X(\omega) \leq x\}, x \in \mathcal{X},$$

$$\{X < x\} := \{\omega \in \Omega | X(\omega) < x\}, x \in \mathcal{X}.$$

Aus diesen Konventionen folgen exemplarisch die folgenden Konventionen für Verteilungen:

$$\mathbb{P}_X(X \in S) = \mathbb{P}(\{X \in S\}) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \in S\}), S \subset \mathcal{X}$$

$$\mathbb{P}_X(X \leq x) = \mathbb{P}(\{X \leq x\}) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \leq x\}), x \in \mathcal{X}.$$

Oft wird zudem auf das ZV Subskript bei Verteilungssymbolen verzichtet, z.B.

$$\mathbb{P}(X \in S) = \mathbb{P}_X(X \in S), S \subset \mathcal{X},$$

$$\mathbb{P}(X \leq x) = \mathbb{P}_X(X \leq S), x \in \mathcal{X}.$$

Zum Zusammenhang von W-Maßen, ZVen, und Dichten

- Verteilungen von Zufallsvariablen sind in der Statistik zentral.
- Verteilungen von Zufallsvariablen werden üblicherweise mithilfe von WMFen und WDFen definiert.
- Welche Rolle spielen dann Wahrscheinlichkeitsräume und Wahrscheinlichkeitsmaße?
- Wahrscheinlichkeitsmaße $\Leftrightarrow$ Verteilungen von Zufallsvariable

$\mathbb{P}$ sei ein Wahrscheinlichkeitsmaß auf der Borel σ -Algebra.

$\Leftrightarrow (\Omega, \mathcal{F}, \mathbb{P}) := (\mathbb{R}, \mathcal{B}, \mathbb{P})$ ist ein Wahrscheinlichkeitsraum

$\Leftrightarrow X : \Omega \rightarrow \mathbb{R}, \omega \mapsto X(\omega) := \omega$ ist eine Zufallsvariable

$\Leftrightarrow \mathbb{P}_X = \mathbb{P}$.

$\Leftrightarrow \mathbb{P}$ ist die Verteilung einer Zufallsvariable.

Wahrscheinlichkeitsräume und Zufallsvariablen

- Wahrscheinlichkeitsräume
- Elementare Wahrscheinlichkeiten
- Konstruktion, Definition, und Notation von Zufallsvariablen
- **Wahrscheinlichkeitsmassefunktionen**
- Wahrscheinlichkeitsdichtefunktionen
- Kumulative Verteilungsfunktionen
- Selbstkontrollfragen

Definition (Diskrete ZV, Wahrscheinlichkeitsmassefunktion)

Eine ZV X heißt *diskret*, wenn ihr Ergebnisraum $\mathcal{X}$ endlich oder abzählbar ist. Die *Wahrscheinlichkeitsmassefunktion (WMF)* einer diskreten ZV ist definiert als

$$p : \mathcal{X} \rightarrow [0, 1], x \mapsto p(x) := \mathbb{P}(X = x). \quad (15)$$

Bemerkungen

- Eine Menge heißt abzählbar, wenn sie bijektiv auf $\mathbb{N}$ abgebildet werden kann.
- WMFen sind nicht-negativ, d.h. $p(x) \geq 0$ für alle $x \in \mathcal{X}$.
- WMFen sind normiert, d.h. $\sum_{x \in \mathcal{X}} p(x) = 1$.
- WMFen heißen im Deutschen üblicherweise *W-Funktionen* oder *Zähldichten*.
- WMFen heißen auf Englisch *probability mass functions (PMFs)*.
- Zur Parallelität mit PMFs und WDFs bevorzugt wird den WMF Begriff.

Definition (Bernoulli-Zufallsvariable)

Es sei X eine Zufallsvariable mit Ergebnisraum $\mathcal{X} = \{0, 1\}$ und WMF

$$p : \mathcal{X} \rightarrow [0, 1], x \mapsto p(x) := \mu^x(1 - \mu)^{1-x} \text{ for } \mu \in [0, 1]. \quad (16)$$

Dann sagen wir, dass X einer *Bernoulli-Verteilung mit Parameter* $\mu \in [0, 1]$ unterliegt und nennen X eine *Bernoulli-Zufallsvariable*. Wir kürzen dies mit $X \sim \text{Bern}(\mu)$ ab. Die WMF einer Bernoulli-Zufallsvariable bezeichnen wir mit

$$\text{Bern}(x; \mu) := \mu^x(1 - \mu)^{1-x}. \quad (17)$$

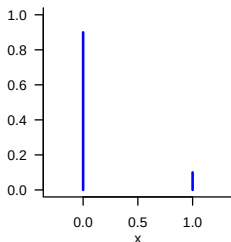
Bemerkungen

- Eine Bernoulli-Zufallsvariable kann als Modell eines Münzwurfs dienen.
- μ ist die Wahrscheinlichkeit dafür, dass X den Wert 1 annimmt,

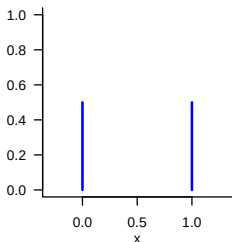
$$\mathbb{P}(X = 1) = \mu^1(1 - \mu)^{1-1} = \mu. \quad (18)$$

Beispiel (Bernoulli-Verteilungen)

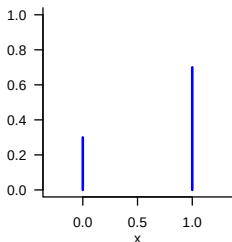
Bern($x;0.1$)



Bern($x;0.5$)



Bern($x;0.7$)



Definition (Binomial-Zufallsvariable)

Es sei X eine Zufallsvariable mit Ergebnisraum $\mathcal{X} := \mathbb{N}_n^0$ und WMF

$$p : \mathcal{X} \rightarrow [0, 1], x \mapsto p(x) := \binom{n}{x} \mu^x (1 - \mu)^{n-x} \text{ for } \mu \in [0, 1]. \quad (19)$$

Dann sagen wir, dass X einer *Binomialverteilung mit Parametern* $\mu \in [0, 1]$ und $n \in \mathbb{N}$ unterliegt und nennen X eine Binomial-Zufallsvariable. Wir kürzen dies mit $X \sim \text{Bin}(\mu, n)$ ab. Die WMF einer Binomial-Zufallsvariable bezeichnen wir mit

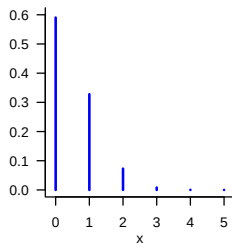
$$\text{Bin}(x; \mu, n) := \binom{n}{x} \mu^x (1 - \mu)^{n-x}. \quad (20)$$

Bemerkung

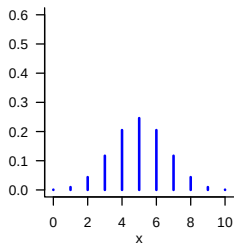
- Es gilt $\text{Bin}(x; \mu, 1) = \text{Bern}(x; \mu)$.

Beispiel (Binomial-Verteilungen)

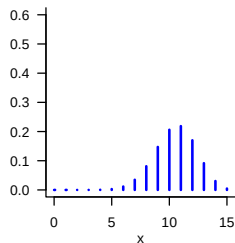
$\text{Bin}(x;5,0.1)$



$\text{Bin}(x;10,0.5)$



$\text{Bin}(x;15,0.7)$



Definition (Diskret-gleichverteilte Zufallsvariable)

Es sei X eine diskrete Zufallsvariable mit Ergebnisraum $\mathcal{X}$ und WMF

$$p : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}, x \mapsto p(x) := \frac{1}{|\mathcal{X}|}. \quad (21)$$

Dann sagen wir, dass X einer *diskreten Gleichverteilung* unterliegt und nennen X eine *diskret-gleichverteilte Zufallsvariable*. Wir kürzen dies mit $X \sim U(|\mathcal{X}|)$ ab. Die WMF einer diskret-gleichverteilten Zufallsvariable bezeichnen wir mit

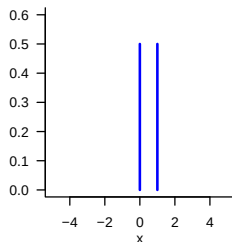
$$U(x; |\mathcal{X}|) := \frac{1}{|\mathcal{X}|}. \quad (22)$$

Bemerkungen

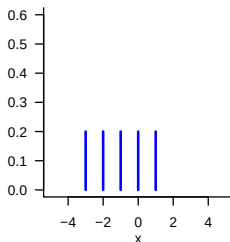
- $\text{Bern}(x; 0.5) = U(x; |\mathcal{X}|)$ für $\mathcal{X} = \{0, 1\}$.
- Es gilt zum Beispiel $\text{Bin}(x; 0.5) = U(x; |\mathcal{X}|)$ für $\mathcal{X} = \{0, 1\}$.

Beispiel (Diskrete Gleichverteilungen)

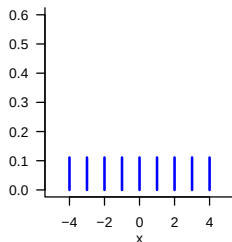
$U(x; X = \{0,1\})$



$U(x; X = \{-3,-2,-1,0,1\})$



$U(x; X = \{-4,-3,-2,-1,0,1,2,3,4\})$



Wahrscheinlichkeitsräume und Zufallsvariablen

- Wahrscheinlichkeitsräume
- Elementare Wahrscheinlichkeiten
- Konstruktion, Definition, und Notation von Zufallsvariablen
- Wahrscheinlichkeitsmassefunktionen
- **Wahrscheinlichkeitsdichtefunktionen**
- Kumulative Verteilungsfunktionen
- Selbstkontrollfragen

Definition (Kontinuierliche ZV, Wahrscheinlichkeitsdichtefunktion)

Eine Zufallsvariable X heißt *kontinuierlich*, wenn eine Funktion

$$p : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, x \mapsto p(x) \quad (23)$$

existiert, so dass gilt

- $p(x) \geq 0$ für alle $x \in \mathbb{R}$,
- $\int_{-\infty}^{\infty} p(x) dx = 1$,
- $\mathbb{P}(a \leq X \leq b) = \int_a^b p(x) dx$ für alle $a, b \in \mathbb{R}, a \leq b$.

Eine entsprechende Funktion p heißt *Wahrscheinlichkeitsdichtefunktion (WDF)* von X .

Bemerkungen

- WDFen können Werte größer als 1 annehmen.
- Es gilt $\mathbb{P}(X = a) = \int_a^a p(x) dx = 0$.
- Wahrscheinlichkeiten werden aus WDFen durch Integration berechnet.
- (Wahrscheinlichkeits)Masse = (Wahrscheinlichkeits)Dichte $\times$ (Mengen)Volumen.

Definition (Normalverteilte und standardnormalverteilte Zufallsvariablen)

Es sei X eine Zufallsvariable mit Ergebnisraum $\mathbb{R}$ und WDF

$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, x \mapsto p(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right). \quad (24)$$

Dann sagen wir, dass X einer *Normalverteilung (oder Gauß-Verteilung)* mit Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ unterliegt und nennen X eine *normalverteilte Zufallsvariable*. Wir kürzen dies mit $X \sim N(\mu, \sigma^2)$ ab. Die WDF einer normalverteilten Zufallsvariable bezeichnen wir mit

$$N(x; \mu, \sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right). \quad (25)$$

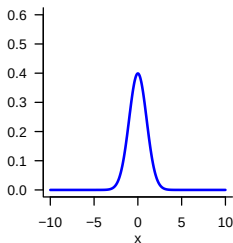
Eine normalverteilte Zufallsvariable mit $\mu = 0$ und $\sigma^2 = 1$ heißt *standardnormalverteilte Zufallsvariable* und wird oft als *Z-Zufallsvariable* bezeichnet.

Bemerkungen

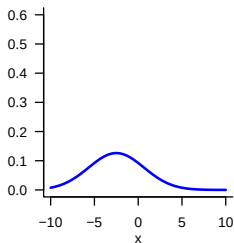
- Der Parameter μ entspricht dem Wert höchster Wahrscheinlichkeitsdichte.
- Der Parameter σ^2 spezifiziert die Breite der WDF.

Beispiel (Normalverteilungen)

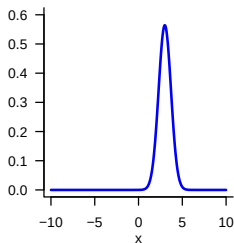
$N(x; 0,1)$



$N(x; -2.5,10)$



$N(x; 3,0.5)$



Definition (Gamma Zufallsvariable)

Es sei X eine Zufallsvariable mit Ergebnisraum $\mathcal{X} := \mathbb{R}_{>0}$ und WDF

$$p : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}, x \mapsto p(x) := \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right), \quad (26)$$

wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass X einer *Gammaverteilung mit Formparameter $\alpha > 0$ und Skalenparameter $\beta > 0$* unterliegt und nennen X eine *gammaverteilte Zufallsvariable*. Wir kürzen dies mit $X \sim G(\alpha, \beta)$ ab. Die WDF einer gammaverteilten Zufallsvariable bezeichnen wir mit

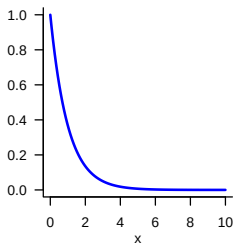
$$G(x; \alpha, \beta) := \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right). \quad (27)$$

Bemerkung

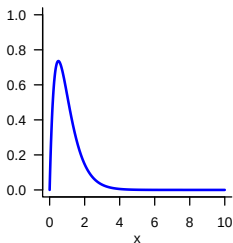
- $G\left(\frac{n}{2}, 2\right)$ heißt auch *Chi-Quadrat (χ^2) Verteilung mit n Freiheitsgraden*.

Beispiel (Gamma Verteilungen)

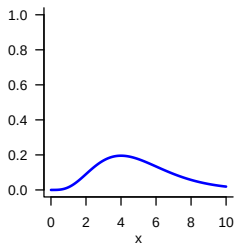
$G(x; 1,1)$



$G(x; 2,2)$



$G(x; 5,1)$



Definition (Beta-Zufallsvariable)

Es sei X eine Zufallsvariable mit Ergebnisraum $\mathcal{X} := [0, 1]$ und WDF

$$p : \mathcal{X} \rightarrow [0, 1], x \mapsto p(x) := \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \text{ for } \alpha, \beta \in \mathbb{R}_{>0}, \quad (28)$$

wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass X einer *Beta-Verteilung* mit Parametern $\alpha > 0$ und $\beta > 0$ unterliegt, und nennen X eine beta-verteilte Zufallsvariable. Wir kürzen dies mit $X \sim \text{Beta}(\alpha, \beta)$ ab. Die WDF einer beta-verteilten Zufallsvariable bezeichnen wir mit

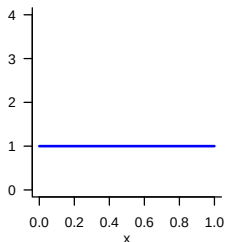
$$\text{Beta}(x; \alpha, \beta) := \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}. \quad (29)$$

Bemerkung

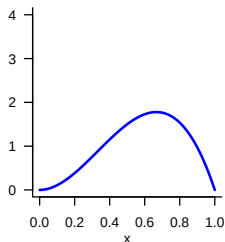
- Für $\alpha < 1, \beta < 1$ ist der Ergebnisraum $\mathcal{X} :=]0, 1[$.

Beispiel (Beta-Verteilungen)

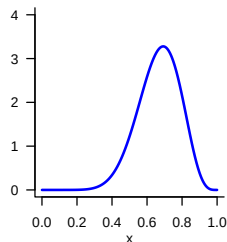
Beta(x; 1,1)



Beta(x; 3,2)



Beta(x; 10,5)



Definition (Gleichverteilte Zufallsvariable)

Es sei X eine kontinuierliche Zufallsvariable mit Ergebnisraum $\mathbb{R}$ und WDF

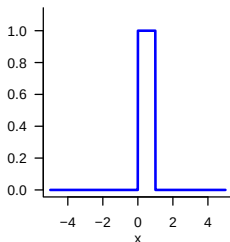
$$p : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, x \mapsto p(x) := \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & x \notin [a, b] \end{cases}. \quad (30)$$

Dann sagen wir, dass X einer *Gleichverteilung mit Parametern a und b* unterliegt und nennen X eine *gleichverteilte Zufallsvariable*. Wir kürzen dies mit $X \sim U(a, b)$ ab. Die WDF einer gleichverteilten Zufallsvariable bezeichnen wir mit

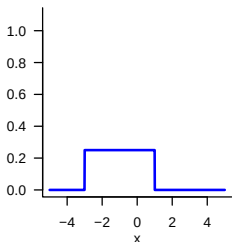
$$U(x; a, b) := \frac{1}{b-a}. \quad (31)$$

Beispiel (Gleichverteilung)

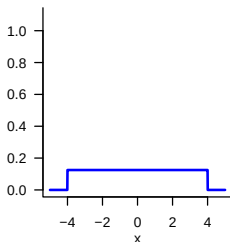
$U(x; 0,1)$



$U(x; -3,1)$



$U(x; -4,4)$



Wahrscheinlichkeitsräume und Zufallsvariablen

- Wahrscheinlichkeitsräume
- Elementare Wahrscheinlichkeiten
- Konstruktion, Definition, und Notation von Zufallsvariablen
- Wahrscheinlichkeitsmassefunktionen
- Wahrscheinlichkeitsdichtefunktionen
- **Kumulative Verteilungsfunktionen**
- Selbstkontrollfragen

Definition (Kumulative Verteilungsfunktion)

Die *kumulative Verteilungsfunktion (KVF)* einer Zufallsvariable X ist definiert als

$$P : \mathbb{R} \rightarrow [0, 1], x \mapsto P(x) := \mathbb{P}(X \leq x). \quad (32)$$

Bemerkungen

- KVFe sind sowohl für diskrete als auch kontinuierliche ZVen definiert.
- $P(x)$ ist für jedes $x \in \mathbb{R}$ definiert, auch wenn $x \notin \mathcal{X}$.
- Mithilfe von KVFe können Intervallwahrscheinlichkeiten angegeben werden

Theorem (Überschreitungswahrscheinlichkeit)

Es sei X eine Zufallsvariable mit Ereignisraum $\mathcal{X}$ und P ihre kumulative Verteilungsfunktion. Dann gilt für die *Überschreitungswahrscheinlichkeit* $\mathbb{P}(X > x)$, dass

$$\mathbb{P}(X > x) = 1 - P(x) \text{ für alle } x \in \mathcal{X}. \quad (33)$$

Beweis

Die Ereignisse $\{X > x\}$ und $\{X \leq x\}$ sind disjunkt und

$$\Omega = \{\omega \in \Omega | X(\omega) > x\} \cup \{\omega \in \Omega | X(\omega) \leq x\} = \{X > x\} \cup \{X \leq x\}. \quad (34)$$

Mit der σ -Additivität von $\mathbb{P}$ folgt dann

$$\begin{aligned} \mathbb{P}(\Omega) &= 1 \\ \Leftrightarrow \mathbb{P}(\{X > x\} \cup \{X \leq x\}) &= 1 \\ \Leftrightarrow \mathbb{P}(\{X > x\}) + \mathbb{P}(\{X \leq x\}) &= 1 \\ \Leftrightarrow \mathbb{P}(\{X > x\}) &= 1 - \mathbb{P}(\{X \leq x\}) \\ \Leftrightarrow \mathbb{P}(\{X > x\}) &= 1 - P(x). \end{aligned} \quad (35)$$

□

Theorem (Intervallwahrscheinlichkeiten)

Es sei X eine Zufallsvariable mit Ereignisraum $\mathcal{X}$ und P ihre kumulative Verteilungsfunktion. Dann gilt für die *Intervallwahrscheinlichkeit* $\mathbb{P}(X \in]x_1, x_2])$, dass

$$\mathbb{P}(X \in]x_1, x_2]) = P(x_2) - P(x_1) \text{ für alle } x_1, x_2 \in \mathcal{X} \text{ mit } x_1 < x_2. \quad (36)$$

Beweis

Wir betrachten die Ereignisse $\{X \leq x_1\}$, $\{x_1 < X \leq x_2\}$ und $\{X \leq x_2\}$, wobei

$$\{X \leq x_1\} \cap \{x_1 < X \leq x_2\} = \emptyset \text{ und } \{X \leq x_1\} \cup \{x_1 < X \leq x_2\} = \{X \leq x_2\}. \quad (37)$$

gelten. Mit der σ -Additivität von $\mathbb{P}$ gilt dann

$$\begin{aligned} \mathbb{P}(\{X \leq x_1\} \cup \{x_1 < X \leq x_2\}) &= \mathbb{P}(\{X \leq x_2\}) \\ \Leftrightarrow \mathbb{P}(\{X \leq x_1\}) + \mathbb{P}(\{x_1 < X \leq x_2\}) &= \mathbb{P}(\{X \leq x_2\}) \\ \Leftrightarrow \mathbb{P}(\{x_1 < X \leq x_2\}) &= \mathbb{P}(\{X \leq x_2\}) - \mathbb{P}(\{X \leq x_1\}) \\ \Leftrightarrow \mathbb{P}(\{x_1 < X \leq x_2\}) &= P(x_2) - P(x_1) \\ \Leftrightarrow \mathbb{P}(X \in]x_1, x_2]) &= P(x_2) - P(x_1). \end{aligned} \quad (38)$$

□

Theorem (Eigenschaften von kumulative Verteilungsfunktionen)

Es sei X eine Zufallsvariable und P ihre kumulative Verteilungsfunktion. Dann hat P die folgenden Eigenschaften

- (1) P ist *monoton steigend*, i.e., wenn $x_1 < x_2$, dann gilt $P(x_1) \leq P(x_2)$.
- (2) $\lim_{x \rightarrow -\infty} P(x) = 0$ und $\lim_{x \rightarrow \infty} P(x) = 1$.
- (3) P ist *rechtsseitig stetig*, d.h., $P(x) = P(x^+) = \lim_{y \rightarrow x, y > x} P(y)$ für alle $x \in \mathbb{R}$

Bemerkungen

- Die genannten Eigenschaften können auch zur Definition einer KVF genutzt werden.
- (3) $\Leftrightarrow$ Eine KVF hat keine Sprünge, wenn man sich Grenzpunkten von rechts nähert.

Beweis

(1) Wir halten zunächst fest, dass für Ereignisse $A \subset B$ gilt, dass $\mathbb{P}(A) \leq \mathbb{P}(B)$. Wie halten dann fest, dass für $x_1 < x_2$,

$$\{X \leq x_1\} = \{\omega \in \Omega | X(\omega) \leq x_1\} \subset \{\omega \in \Omega | X(\omega) \leq x_2\} = \{X \leq x_2\} \quad (39)$$

Also gilt

$$\mathbb{P}(\{X \leq x_1\}) \leq \mathbb{P}\{X \leq x_2\} \Rightarrow P(x_1) \leq P(x_2). \quad (40)$$

(2) Wir verzichten auf einen Beweis

(3) Wir definieren

$$P(x^+) = \lim_{y \rightarrow x, y > x} P(y). \quad (41)$$

Seien nun $y_1 > y_2 > \dots$ so, dass $\lim_{n \rightarrow \infty} y_n = x$. Dann gilt

$$\{X \leq x\} = \cap_{n=1}^{\infty} \{X \leq y_n\}. \quad (42)$$

Es gilt also

$$P(x) = \mathbb{P}(\{X \leq x\}) = \mathbb{P}(\cap_{n=1}^{\infty} \{X \leq y_n\}) = \lim_{n \rightarrow \infty} \mathbb{P}(\{X \leq y_n\}) = P(x^+), \quad (43)$$

wobei wir die dritte Gleichung unbegründet stehen lassen.

Kumulative Verteilungsfunktionen von diskreten Zufallsvariablen

- Wenn $a < b$ und $\mathbb{P}(a < X < b) = 0$, dann ist P konstant horizontal auf $]a, b[$.
- An jedem Punkt x mit $\mathbb{P}(X = x) > 0$ springt die KVF um den Betrag $\mathbb{P}(X = x)$.
- $\Leftrightarrow$ An jedem Punkt x mit $p(x) > 0$ springt die KVF um den Betrag $p(x) > 0$.
- Generell ist die KVF einer diskreten Zufallsvariable mit Ergebnisraum $\mathbb{N}_0$ durch

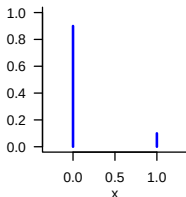
$$P : \mathbb{R} \rightarrow [0, 1], x \mapsto P(x) := \sum_{k=0}^{\lfloor x \rfloor} \mathbb{P}(X = k) \quad (44)$$

gegeben, wobei $\lfloor x \rfloor$ die Abrundungsfunktion bezeichnet.

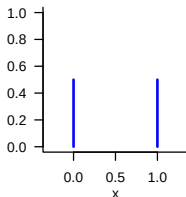
Kumulative Verteilungsfunktionen

Beispiel (Bernoulliverteilungen)

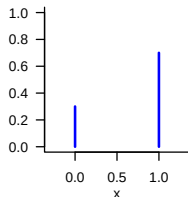
WMF von $\text{Bern}(x;0.1)$



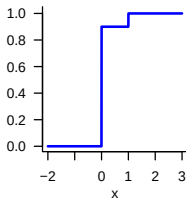
WMF von $\text{Bern}(x;0.5)$



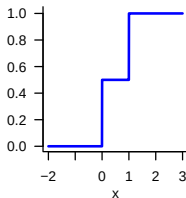
WMF von $\text{Bern}(x;0.7)$



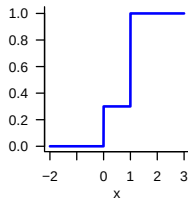
KVF von $\text{Bern}(x;0.1)$



KVF von $\text{Bern}(x;0.5)$



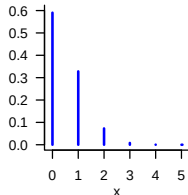
KVF von $\text{Bern}(x;0.7)$



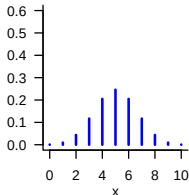
Kumulative Verteilungsfunktionen

Beispiel (Binomialverteilungen)

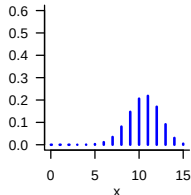
WMF von $\text{Bin}(x;5,0.1)$



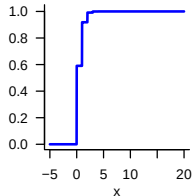
WMF von $\text{Bin}(x;10,0.5)$



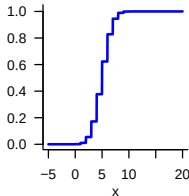
WMF von $\text{Bin}(x;15,0.7)$



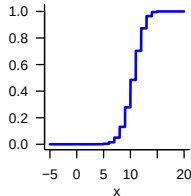
KVF von $\text{Bin}(x;5,0.1)$



KVF von $\text{Bin}(x;10,0.5)$



KVF von $\text{Bin}(x;15,0.7)$



Theorem (Kumulative Verteilungsfunktionen von kontinuierlichen ZVen)

X sei eine kontinuierliche Zufallsvariable mit WDF p und KVF P . Dann gilt

$$P(x) = \int_{-\infty}^x p(t) dt \text{ und } p(x) = \frac{d}{dx} P(x). \quad (45)$$

Beweis

Wir halten zunächst fest, dass weil $\mathbb{P}(X = x) = 0$ für alle $x \in \mathbb{R}$ gilt, die KVF von X keine Sprünge hat, d.h. P ist stetig. Mit der Definitionen von WDF und KVF, folgt, dass P die Form einer Stammfunktion von p hat. Dass p die Ableitung von P ist, folgt dann unmittelbar aus dem Fundamentalsatz der Analysis.

□

Bemerkungen

- Die KVF ist **eine** Stammfunktion der WDF, die WDF ist die Ableitung der KVF.
- Das *Theorem von Radon-Nikodym* ist eine generalisierte Variante dieser Einsicht.
- KVFe von kontinuierlichen ZV heißen auch kumulative Dichtefunktionen (KDFen).

Beispiel (Normalverteilung)

Es sei $X \sim N(\mu, \sigma^2)$.

- Die WDF von X ist

$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, x \mapsto p(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right).$$

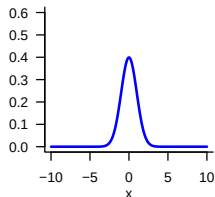
- Die KVF von X ist

$$P : \mathbb{R} \rightarrow]0, 1[, x \mapsto P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x \exp\left(-\frac{1}{2\sigma^2}(\xi - \mu)^2\right) d\xi.$$

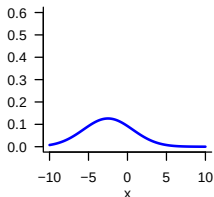
- Die KVF von $X \sim N(\mu, \sigma^2)$ kann nur numerisch, nicht analytisch, berechnet werden.
- Für $\mu = 1, \sigma^2 = 1$, gilt zum Beispiel $p(2) = 0.24$ und $P(2) = 0.84$.
- Die WDF und KVF von $Z \sim N(0, 1)$ werden oft mit ϕ und Φ , respektive, bezeichnet.

Beispiel (Normalverteilung)

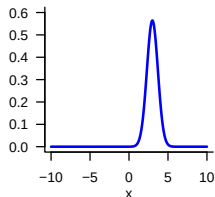
WDF von $N(0,1)$



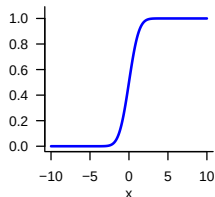
WDF von $N(-2.5,10)$



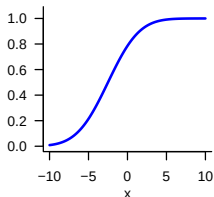
WDF von $N(3,0.5)$



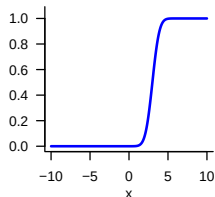
KDF von $N(0,1)$



KDF von $N(-2.5,10)$



KDF von $N(3,0.5)$



Definition (Inverse Kumulative Verteilungsfunktion)

X sei eine kontinuierliche Zufallsvariable mit KVF P . Dann heißt die Funktion

$$P^{-1} :]0, 1[\rightarrow \mathbb{R}, q \mapsto P^{-1}(q) := \{x \in \mathbb{R} | P(x) = q\} \quad (46)$$

die *inverse kumulative Verteilungsfunktion* von X .

Bemerkungen

- P^{-1} ist die Inverse von P , d.h. $P^{-1}(P(x)) = x$.
- Offenbar gilt $P(x) = q \Leftrightarrow \mathbb{P}(X \leq x) = q$.
- Für $q \in]0, 1[$ ist also $P^{-1}(q)$ der Wert x von X , so dass $\mathbb{P}(X \leq x) = q$ gilt.
- Wenn $Z \sim N(0, 1)$ mit KVF Φ ist, dann gilt zum Beispiel $\Phi^{-1}(0.975) = 1.960$.

Beispiel (Normalverteilung)

Es sei $X \sim N(\mu, \sigma^2)$.

- Die KVF von X ist

$$P : \mathbb{R} \rightarrow]0, 1[, x \mapsto \mathbb{P}(X \leq x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x \exp\left(-\frac{1}{2\sigma^2}(\xi - \mu)^2\right) d\xi \quad (47)$$

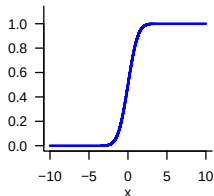
- Die inverse KVF von X ist

$$P^{-1} :]0, 1[\rightarrow \mathbb{R}, q \mapsto P^{-1}(q) = \{x \in \mathbb{R} | P(x) = q\}. \quad (48)$$

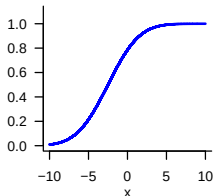
- Für $\mu = 1, \sigma^2 = 1$ gilt z.B., dass $P(2) = 0.84$ und $P^{-1}(0.84) = 2$.
- Die inverse KVF von $X \sim N(0, 1)$ wird oft mit Φ^{-1} bezeichnet.
- Typische Beispielwerte für die KVF und inverse KVF von $N(0, 1)$ sind
 - $\Phi(1.645) = 0.950$, $\Phi^{-1}(0.950) = \Phi^{-1}(1 - 0.050) = 1.640$.
 - $\Phi(1.960) = 0.975$, $\Phi^{-1}(0.975) = \Phi^{-1}(1 - \frac{0.050}{2})$.

Beispiel (Normalverteilung)

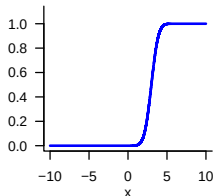
KDF von $N(0,1)$



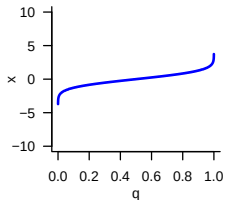
KDF von $N(-2.5,10)$



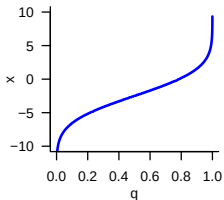
KDF von $N(3,0.5)$



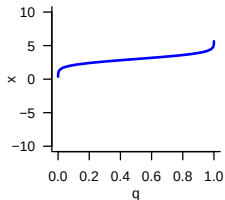
Inverse KDF von $N(0,1)$



Inverse KDF von $N(-2.5,10)$



Inverse KDF von $N(3,0.5)$



Wahrscheinlichkeitsräume und Zufallsvariablen

- Wahrscheinlichkeitsräume
- Elementare Wahrscheinlichkeiten
- Konstruktion, Definition, und Notation von Zufallsvariablen
- Wahrscheinlichkeitsmassefunktionen
- Wahrscheinlichkeitsdichtefunktionen
- Kumulative Verteilungsfunktionen
- **Selbstkontrollfragen**

1. Definieren Sie den Begriff der σ -Algebra
2. Definieren Sie den Begriff des Wahrscheinlichkeitsmaßes.
3. Definieren Sie den Begriff des Wahrscheinlichkeitsraumes
4. Skizzieren Sie das Wahrscheinlichkeitsraummodell eines fairen Würfels.
5. Geben sie zwei Interpretationen der Wahrscheinlichkeit eines Ereignisses.
6. Definieren Sie den Begriff der Unabhängigkeit zweier Ereignisse.
7. Definieren Sie die bedingte Wahrscheinlichkeit eines Ereignisses.
8. Geben Sie das Gesetz von der totalen Wahrscheinlichkeit wieder.
9. Geben Sie das Theorem zur bedingten Wahrscheinlichkeit unter Unabhängigkeit wieder.
10. Definieren Sie den Begriff der Zufallsvariable.

11. Definieren Sie den Begriff der Wahrscheinlichkeitsmassefunktion.
12. Definieren Sie die Begriffe der Wahrscheinlichkeitsdichtefunktion.
13. Definieren Sie den Begriff der kumulativen Verteilungsfunktion.
14. Schreiben sie die Intervallwahrscheinlichkeit einer Zufallsvariable mithilfe ihrer KVF.
15. Definieren Sie die WDF und KVF einer normalverteilten Zufallsvariable.
16. Schreiben Sie den Wert $P(x)$ der KVF einer Zufallsvariable mithilfe ihrer WDF.
17. Schreiben Sie den Wert $p(x)$ der WDF einer Zufallsvariable mithilfe ihrer KVF.
18. Nennen sie drei Eigenschaften von kumulativen Verteilungsfunktionen.
19. Definieren Sie den Begriff der inversen Verteilungsfunktion.

Literatur

- DeGroot, M. H. and Schervish, M. J. (2012). *Probability and Statistics*. Addison-Wesley, Boston, 4th ed edition.
- Georgii, H.-O. (2009). *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. DeGruyter-Lehrbuch. de Gruyter, Berlin, 4., überarb. und erw. aufl edition.
- Wasserman, L. (2004). *All of Statistics*.



Inferenzstatistik

BSc Psychologie SoSe 2021

Prof. Dr. Dirk Ostwald

(3) Multivariate Verteilungen

Literaturhinweise

Die Lektüre von Georgii (2009, Abschnitte 3.1 - 3.4) wird empfohlen. Entsprechende englischsprachige Darstellungen finden sich in DeGroot and Schervish (2012, Sections 3.4 - 3.7) und Wasserman (2004, Sections 2.9 - 2.10).

Multivariate Verteilungen

- Zufallsvektoren
- Multivariate Verteilungen
- Marginalverteilungen
- Bedingte Verteilungen
- Unabhängige Zufallsvariablen
- Selbstkontrollfragen

Multivariate Verteilungen

- **Zufallsvektoren**
- Multivariate Verteilungen
- Marginalverteilungen
- Bedingte Verteilungen
- Unabhängige Zufallsvariablen
- Selbstkontrollfragen

Definition (Zufallsvektor)

Es sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein W-Raum und $(\mathcal{X}, \mathcal{S})$ ein n -dimensionaler Messraum. Ein n -dimensionaler *Zufallsvektor* ist dann definiert als eine Abbildung

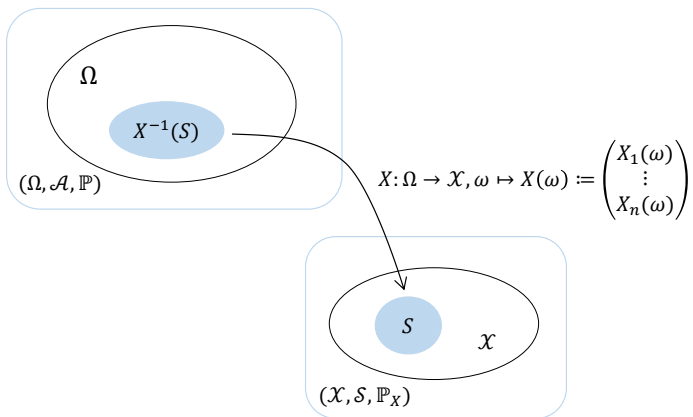
$$X : \Omega \rightarrow \mathcal{X}, \omega \mapsto X(\omega) := \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_n(\omega) \end{pmatrix} \quad (1)$$

mit der *Messbarkeitseigenschaft*

$$\{\omega \in \Omega \mid X(\omega) \in S\} \in \mathcal{A} \text{ für alle } S \in \mathcal{S}. \quad (2)$$

Bemerkungen

- X ist messbar, wenn die Komponentenfunktionen $X_1, \dots, X_n$ messbar sind.
- Die Komponentenfunktionen eines Zufallsvektors sind also Zufallsvariablen.
- Ein n -dimensionaler Zufallsvektor ist die Konkatenation von n Zufallsvariablen.
- Für einen Zufallsvektor schreiben wir auch häufig $X := (X_1, \dots, X_n)$.
- Für $n := 1$ ist ein Zufallsvektor eine Zufallsvariable.



$$\mathbb{P}(X^{-1}(S)) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \in S\}) =: \mathbb{P}_X(S)$$

Multivariate Verteilungen

- Zufallsvektoren
- **Multivariate Verteilungen**
- Marginalverteilungen
- Bedingte Verteilungen
- Unabhängige Zufallsvariablen
- Selbstkontrollfragen

Definition (Multivariate Verteilung)

Es sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $(\mathcal{X}, \mathcal{S})$ ein n -dimensionaler Messraum und

$$X : \Omega \rightarrow \mathcal{X}, \omega \mapsto X(\omega) \quad (3)$$

ein Zufallsvektor. Dann heißt das Wahrscheinlichkeitsmaß $\mathbb{P}_X$, definiert durch

$$\mathbb{P}_X : \mathcal{S} \rightarrow [0, 1], S \mapsto \mathbb{P}_X(S) := \mathbb{P}(X^{-1}(S)) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \in S\}) \quad (4)$$

die *multivariate Verteilung des Zufallsvektor* X .

Bemerkungen

- Der Einfachheit halber spricht man oft auch nur von “der Verteilung des Zufallsvektors X ”.
- Die Notationskonventionen für Zufallsvariablen gelten für Zufallsvektoren analog, z.B.

$$\begin{aligned} \mathbb{P}_X(X \in S) &:= \mathbb{P}(\{X \in S\}) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \in S\}) \\ \mathbb{P}_X(X = x) &:= \mathbb{P}(\{X = x\}) = \mathbb{P}(\{\omega \in \Omega | X(\omega) = x\}) \\ \mathbb{P}_X(X \leq x) &:= \mathbb{P}(\{X \leq x\}) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \leq x\}) \end{aligned} \quad (5)$$

$$\mathbb{P}_X(x_1 \leq X \leq x_2) := \mathbb{P}(\{x_1 \leq X \leq x_2\}) = \mathbb{P}(\{\omega \in \Omega | x_1 \leq X(\omega) \leq x_2\})$$

- Relationsoperatoren wie $\leq$ werden hier komponentenweise verstanden, d.h. zum Beispiel heißt $x \leq y$ für $x, y \in \mathbb{R}^n$, dass $x_i \leq y_i$ für alle $i = 1, \dots, n$.

Definition (Multivariate kumulative Verteilungsfunktionen)

X sei ein Zufallsvektor mit Ergebnisraum $\mathcal{X}$. Dann heißt eine Funktion der Form

$$P_X : \mathcal{X} \rightarrow [0, 1], x \mapsto P_X(x) := \mathbb{P}_X(X \leq x) \quad (6)$$

multivariate kumulative Verteilungsfunktion von X .

Bemerkungen

- M.v. kumulative Verteilungsfunktionen können zur Definition m.v. Verteilungen genutzt werden.
- Häufiger ist die Definition m.v. Verteilungen durch multivariate WMFen oder WDFen.

Definition (Diskreter Zufallsvektor, multivariate WMF)

Es sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathcal{X}$ ein Zufallsvektor. X heißt *diskreter Zufallsvektor*, wenn der Ergebnisraum $\mathcal{X}$ endlich viele oder höchstens abzählbar viele Elemente $x_i, i = 1, 2, \dots$ enthält. Die *multivariate Wahrscheinlichkeitsmassefunktion (WMF)* eines diskreten Zufallsvektors X wird mit p_X bezeichnet und ist definiert durch

$$p_X : \mathcal{X} \rightarrow [0, 1], x \mapsto p_X(x) := \mathbb{P}_X(X = x). \quad (7)$$

Bemerkungen

- Der Begriff der multivariaten WMF ist analog zum Begriff der WMF.
- Wie univariate WMFen sind multivariate WMFen nicht-negativ und normiert.

Beispiel (Multivariate Wahrscheinlichkeitsmassefunktion)

Wir betrachten einen zweidimensionalen Zufallsvektor $X := (X_1, X_2)$ der Werte in $\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2$ annimmt, wobei $\mathcal{X}_1 := \{1, 2, 3\}$ und $\mathcal{X}_2 = \{1, 2, 3, 4\}$ seien. Eine exemplarische bivariate WMF der Form

$$p_X : \{1, 2, 3\} \times \{1, 2, 3, 4\} \rightarrow [0, 1], (x_1, x_2) \mapsto p_X(x_1, x_2) \quad (8)$$

ist dann durch nachfolgende Tabelle definiert.

$p_X(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$
$x_1 = 1$	0.1	0.0	0.2	0.1
$x_1 = 2$	0.1	0.2	0.0	0.0
$x_1 = 3$	0.0	0.1	0.1	0.1

Man beachte, dass $\sum_{x_1=1}^3 \sum_{x_2=1}^4 p_X(x_1, x_2) = 1$.

Definition (Kontinuierlicher Zufallsvektor, multivariate WDF)

Es sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Ein Zufallsvektor der Form $X : \Omega \rightarrow \mathbb{R}^n$ heißt *kontinuierlicher Zufallsvektor*. Die *multivariate Wahrscheinlichkeitsdichtefunktion (WDF)* eines kontinuierlichen Zufallsvektors X ist eine Funktion

$$p_X : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}, x \mapsto p_X(x), \quad (9)$$

mit den Eigenschaften

$$(1) \int_{\mathbb{R}^n} p_X(x) dx = 1$$

$$(2) \mathbb{P}_X(x_1 \leq X \leq x_2) = \int_{x_{11}}^{x_{21}} \cdots \int_{x_{1n}}^{x_{2n}} p_X(s_1, \dots, s_n) ds_1 \cdots ds_n$$

Bemerkungen

- Der Begriff der multivariaten WDF ist analog zum Begriff der WDF.
- Wie univariate WDFen sind multivariate WDFen nicht-negativ und normiert.
- Wie für kontinuierliche Zufallsvariablen gilt für kontinuierliche Zufallsvektoren

$$\mathbb{P}_X(X = x) = \mathbb{P}_X(x \leq X \leq x) = \int_{x_1}^{x_1} \cdots \int_{x_n}^{x_n} p_X(s_1, \dots, s_n) ds_1 \cdots ds_n \quad (10)$$

Multivariate Verteilungen

- Zufallsvektoren
- Multivariate Verteilungen
- **Marginalverteilungen**
- Bedingte Verteilungen
- Unabhängige Zufallsvariablen
- Selbstkontrollfragen

Definition (Univariate Marginalverteilung)

$(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum, $(\mathcal{X}, \mathcal{S})$ sei ein n -dimensionaler Messraum, $X : \Omega \rightarrow \mathcal{X}$ sei ein Zufallsvektor, $\mathbb{P}_X$ sei die Verteilung von X , $\mathcal{X}_i \subset \mathcal{X}$ sei der Ergebnisraum der i ten Komponente X_i von X , und $\mathcal{S}_i$ sei eine σ -Algebra auf $\mathcal{X}_i$. Dann heit die durch

$$\mathbb{P}_{X_i} : \mathcal{S}_i \rightarrow [0, 1], S \mapsto \mathbb{P}_X (\mathcal{X}_1 \times \cdots \times \mathcal{X}_{i-1} \times S \times \mathcal{X}_{i+1} \times \cdots \times \mathcal{X}_n) \text{ fr } S \in \mathcal{S}_i \quad (11)$$

definierte Verteilung die *ite univariate Marginalverteilung von X* .

Bemerkungen

- Univariate Marginalverteilungen sind die Verteilungen der Komponenten eines Zufallsvektors.
- Die Komponenten eines Zufallsvektors sind Zufallsvariablen, univariate Marginalverteilungen sind also Verteilungen von Zufallsvariablen.
- Die Festlegung der multivariaten Verteilung von X legt auch die Verteilungen der X_i fest.

Theorem (Marginale Wahrscheinlichkeitsmasse/dichtefunktionen)

(1) $X = (X_1, \dots, X_n)$ sei ein n -dimensionaler diskreter Zufallsvektor mit Wahrscheinlichkeitsmassefunktion p_X und Komponentenergebnisräumen $\mathcal{X}_1, \dots, \mathcal{X}_n$. Dann ergibt sich die Wahrscheinlichkeitsmassefunktion der i ten Komponente X_i von X als

$$p_{X_i} : \mathcal{X}_i \rightarrow [0, 1], x_i \mapsto p_{X_i}(x_i) := \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} p_X(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) \quad (12)$$

(2) $X = (X_1, \dots, X_n)$ sei ein n -dimensionaler kontinuierlicher Zufallsvektor mit Wahrscheinlichkeitsdichtefunktion p_X und Komponentenergebnisraum $\mathbb{R}$. Dann ergibt sich die Wahrscheinlichkeitsdichtefunktion der i ten Komponente X_i von X als

$$p_{X_i} : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, x_i \mapsto p_{X_i}(x_i) := \int_{x_1} \cdots \int_{x_{i-1}} \int_{x_{i+1}} \cdots \int_{x_n} p_X(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n. \quad (13)$$

Bemerkungen

- Wir verzichten auf einen Beweis
- Die WMFen der univariaten Marginalverteilungen eines diskreten Zufallsvektors werden durch Summation bestimmt, die WDFen der univariaten Marginalverteilungen eines kontinuierlichen Zufallsvektors werden durch Integration bestimmt.

Beispiel (Marginale Wahrscheinlichkeitsmassefunktionen)

Wir betrachten erneut den zweidimensionalen Zufallsvektor $X := (X_1, X_2)$ der Werte in $\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2$ annimmt, wobei $\mathcal{X}_1 := \{1, 2, 3\}$ und $\mathcal{X}_2 = \{1, 2, 3, 4\}$ seien. Basierend auf der oben definierten WMF ergeben sich folgende marginalen WMFen p_{X_1} und p_{X_2}

$p_X(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$	$p_{X_1}(x_1)$
$x_1 = 1$	0.1	0.0	0.2	0.1	0.4
$x_1 = 2$	0.1	0.2	0.0	0.0	0.3
$x_1 = 3$	0.0	0.1	0.1	0.1	0.3
$p_{X_2}(x_2)$	0.2	0.3	0.3	0.2	

Man beachte, dass $\sum_{x_1=1}^3 p_{X_1}(x_1) = 1$ und $\sum_{x_2=1}^4 p_{X_2}(x_2) = 1$ gilt.



Multivariate Verteilungen

- Zufallsvektoren
- Multivariate Verteilungen
- Marginalverteilungen
- **Bedingte Verteilungen**
- Unabhängige Zufallsvariablen
- Selbstkontrollfragen

Vorbemerkungen

Wir erinnern uns, dass für einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ und zwei Ereignisse $A, B \in \mathcal{A}$ mit $\mathbb{P}(B) > 0$ die bedingte Wahrscheinlichkeit von A gegeben B definiert ist als

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (14)$$

Analog wird für zwei Zufallsvariablen X_1, X_2 mit Ereignisräumen $\mathcal{X}_1, \mathcal{X}_2$ und (messbaren) Mengen $S_1 \in \mathcal{X}_1, S_2 \in \mathcal{X}_2$ die bedingte Verteilung von X_1 gegeben X_2 mithilfe der Ereignisse $A := \{X_1 \in S_1\}$ und $B := \{X_2 \in S_2\}$ definiert.

So ergibt sich zum Beispiel die bedingte Wahrscheinlichkeit, dass $X_1 \in S_1$ gegeben dass $X_2 \in S_2$ unter der Annahme, dass $\mathbb{P}(\{X_2 \in S_2\}) > 0$, zu

$$\mathbb{P}(\{X_1 \in S_1\}|\{X_2 \in S_2\}) = \frac{\mathbb{P}(\{X_1 \in S_1\} \cap \{X_2 \in S_2\})}{\mathbb{P}(\{X_2 \in S_2\})}. \quad (15)$$

In der Folge betrachten wir zunächst durch die WMFen/WDFen zweidimensionaler Zufallsvektoren definierte bedingte Verteilungen.

Definition (Bedingte WMF, diskrete bedingte Verteilung)

$X := (X_1, X_2)$ sei ein diskreter Zufallsvektor mit Ergebnisraum $\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2$, WMF $p_X = p_{X_1, X_2}$ und marginalen WMFen p_{X_1} und p_{X_2} . Die bedingte WMF von X_1 gegeben $X_2 = x_2$ ist dann für $p_{X_2}(x_2) > 0$ definiert als

$$p_{X_1|X_2=x_2} : \mathcal{X}_1 \rightarrow [0, 1], x_1 \mapsto p_{X_1|X_2=x_2}(x_1|x_2) := \frac{p_{X_1, X_2}(x_1, x_2)}{p_{X_2}(x_2)} \quad (16)$$

Analog ist für $p_{X_1}(x_1) > 0$ die bedingte WMF von X_2 gegeben $X_1 = x_1$ definiert als

$$p_{X_2|X_1=x_1} : \mathcal{X}_2 \rightarrow [0, 1], x_2 \mapsto p_{X_2|X_1=x_1}(x_2|x_1) := \frac{p_{X_1, X_2}(x_1, x_2)}{p_{X_1}(x_1)} \quad (17)$$

Die bedingten Verteilungen mit WMFen $p_{X_1|X_2=x_2}$ und $p_{X_2|X_1=x_1}$ heißen dann die *diskreten bedingten Verteilungen* von X_1 gegeben $X_2 = x_2$ und X_2 gegeben $X_1 = x_1$, respektive.

Bemerkungen

- In Analogie zur Definition der bedingten Wahrscheinlichkeit von Ereignissen gilt also

$$p_{X_1|X_2}(x_1|x_2) = \frac{p_{X_1, X_2}(x_1, x_2)}{p_{X_2}(x_2)} = \frac{\mathbb{P}(\{X_1 = x_1\} \cap \{X_2 = x_2\})}{\mathbb{P}(\{X_2 = x_2\})}. \quad (18)$$

- **Bedingte Verteilungen sind lediglich normalisierte gemeinsame Verteilungen!**

Beispiel (Bedingte Wahrscheinlichkeitsmassenfunktionen)

Wir betrachten erneut den zweidimensionalen Zufallsvektor $X := (X_1, X_2)$ der Werte in $\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2$ annimmt, wobei $\mathcal{X}_1 := \{1, 2, 3\}$ und $\mathcal{X}_2 = \{1, 2, 3, 4\}$ seien. Basierend auf der oben definierten WMF und den entsprechenden oben evaluierten marginalen WMFen ergeben sich folgende bedingte WMFen für $p_{X_2|X_1=x_1}$

$p_{X_1 X_2}(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$
$p_{X_2 X_1=1}(x_2 x_1 = 1)$	$\frac{0.1}{0.4} = 0.25$	$\frac{0.0}{0.4} = 0.00$	$\frac{0.2}{0.4} = 0.50$	$\frac{0.1}{0.4} = 0.25$
$p_{X_2 X_1=2}(x_2 x_1 = 2)$	$\frac{0.1}{0.3} = 0.3\bar{3}$	$\frac{0.2}{0.3} = 0.6\bar{6}$	$\frac{0.0}{0.3} = 0.00$	$\frac{0.0}{0.3} = 0.00$
$p_{X_2 X_1=3}(x_2 x_1 = 3)$	$\frac{0.0}{0.3} = 0.00$	$\frac{0.1}{0.3} = 0.3\bar{3}$	$\frac{0.1}{0.3} = 0.3\bar{3}$	$\frac{0.1}{0.3} = 0.3\bar{3}$

Bemerkungen

- Man beachte, dass $\sum_{x_2=1}^4 p_{X_2|X_1=x_1}(x_2|x_1) = 1$ für alle $x_1 \in \mathcal{X}_1$.
- Man beachte die qualitative Ähnlichkeit der WMFen $p_{X_1, X_2}(x_1, x_2)$ und $p_{X_2|X_1}(x_2|x_1)$.
- **Bedingte Verteilungen sind lediglich normalisierte gemeinsame Verteilungen!**

Definition (Bedingte WDF, kontinuierliche bedingte Verteilungen)

$X := (X_1, X_2)$ sei ein kontinuierlicher Zufallsvektor mit Ergebnisraum $\mathbb{R}^2$, WDF $p_X = p_{X_1, X_2}$ und marginalen WDFen p_{X_1} und p_{X_2} . Die bedingte WDF von X_1 gegeben $X_2 = x_2$ ist dann für $p_{X_2}(x_2) > 0$ definiert als

$$p_{X_1|X_2=x_2} : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, x_1 \mapsto p_{X_1|X_2=x_2}(x_1|x_2) := \frac{p_{X_1, X_2}(x_1, x_2)}{p_{X_2}(x_2)} \quad (19)$$

Analog ist für $p_{X_1}(x_1) > 0$ die bedingte WMF von X_2 gegeben $X_1 = x_1$ definiert als

$$p_{X_2|X_1=x_1} : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, x_2 \mapsto p_{X_2|X_1=x_1}(x_2|x_1) := \frac{p_{X_1, X_2}(x_1, x_2)}{p_{X_1}(x_1)} \quad (20)$$

Die Verteilungen mit WDFen $p_{X_1|X_2=x_2}$ und $p_{X_2|X_1=x_1}$ heißen dann die *kontinuierlichen bedingten Verteilungen* von X_1 gegeben $X_2 = x_2$ und X_2 gegeben $X_1 = x_1$, respektive.

Bemerkungen

- Im kontinuierlichen Fall gilt zwar $\mathbb{P}(X = x) = 0$, aber nicht notwendig auch $p_X(x) = 0$!

Multivariate Verteilungen

- Zufallsvektoren
- Multivariate Verteilungen
- Marginalverteilungen
- Bedingte Verteilungen
- **Unabhängige Zufallsvariablen**
- Selbstkontrollfragen

Definition (Unabhängige Zufallsvariablen)

$(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum und $X := (X_1, X_2)$ ein zweidimensionaler Zufallsvektor. Die Zufallsvariablen X_1, X_2 mit Ergebnisräumen $\mathcal{X}_1, \mathcal{X}_2$ heißen *unabhängig*, wenn für alle $S_1 \subseteq \mathcal{X}_1$ und $S_2 \subseteq \mathcal{X}_2$ gilt, dass

$$\mathbb{P}_X(X_1 \in S_1, X_2 \in S_2) = \mathbb{P}_{X_1}(X_1 \in S_1)\mathbb{P}_{X_2}(X_2 \in S_2). \quad (21)$$

Bemerkungen

- Die Definition besagt, dass die Ereignisse $\{X_1 \in S_1\}$ und $\{X_2 \in S_2\}$ unabhängig sind.
- Es gilt also auch, dass $\mathbb{P}(\{X_1 \in S_1\}|\{X_2 \in S_2\}) = \mathbb{P}(\{X_1 \in S_1\})$.
- Wissen um das Ereignis $\{X_2 \in S_2\}$ verändert die Wahrscheinlichkeit von $\{X_1 \in S_1\}$ also nicht.
- Einen formaleren Zugang bietet das Konzept der *Produktwahrscheinlichkeitsräume*.

Theorem (Unabhängigkeit und Faktorisierung der WMF/WDF)

(1) $X := (X_1, X_2)$ sei ein diskreter Zufallsvektor mit Ergebnisraum $\mathcal{X}_1 \times \mathcal{X}_2$, WMF p_X und marginalen WMFen p_{X_1}, p_{X_2} . Dann gilt

X_1 und X_2 sind unabhängige Zufallsvariablen $\Leftrightarrow$

$$p_X(x_1, x_2) = p_{X_1}(x_1)p_{X_2}(x_2) \text{ für alle } (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2. \quad (22)$$

(2) $X := (X_1, X_2)$ sei ein kontinuierlicher Zufallsvektor mit Ergebnisraum $\mathbb{R}^2$, WDF p_X und marginalen WMFen p_{X_1}, p_{X_2} . Dann gilt

X_1 und X_2 sind unabhängige Zufallsvariablen $\Leftrightarrow$

$$p_X(x_1, x_2) = p_{X_1}(x_1)p_{X_2}(x_2) \text{ für alle } (x_1, x_2) \in \mathbb{R}^2. \quad (23)$$

Bemerkungen

- Wir verzichten auf einen Beweis.
- Die Produkteigenschaft $p_X(x_1, x_2) = p_{X_1}(x_1)p_{X_2}(x_2)$ heißt auch *Faktorisierung*.
- Unabhängigkeit zweier ZVen entspricht der Faktorisierung ihrer gemeinsamen WMF/WDF.

Beispiel (Unabhängige diskrete Zufallsvariablen)

Wir betrachten erneut den zweidimensionalen Zufallsvektor $X := (X_1, X_2)$, der Werte in $\{1, 2, 3\} \times \{1, 2, 3, 4\}$ annimmt, und dessen gemeinsame und marginale WMFen die untenstehende Form haben

$p_X(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$	$p_{X_1}(x_1)$
$x_1 = 1$	0.10	0.00	0.20	0.10	0.40
$x_1 = 2$	0.10	0.20	0.00	0.00	0.30
$x_1 = 3$	0.00	0.10	0.10	0.10	0.30
$p_{X_2}(x_2)$	0.20	0.30	0.30	0.20	

Da hier gilt, dass

$$p_X(1, 1) = 0.10 \neq 0.08 = 0.40 \cdot 0.20 = p_{X_1}(1)p_{X_2}(1) \quad (24)$$

sind die Zufallsvariablen X_1 und X_2 nicht unabhängig.

Beispiel (Unabhängige diskrete Zufallsvariablen)

Die gemeinsame Verteilung von X_1 und X_2 unter der Annahme der Unabhängigkeit von X_1 und X_2 bei gleichen Marginalverteilungen ergibt sich zu

$p_X(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$	$p_{X_1}(x_1)$
$x_1 = 1$	0.08	0.12	0.12	0.08	0.40
$x_1 = 2$	0.06	0.09	0.09	0.06	0.30
$x_1 = 3$	0.06	0.09	0.09	0.06	0.30
$p_{X_2}(x_2)$	0.20	0.30	0.30	0.20	

(NEU) Weiterhin ergeben sich im Falle der Unabhängigkeit von X_1 und X_2 zum Beispiel die bedingten Wahrscheinlichkeitsmassfunktion $p_{X_2|X_1}$ zu

$p_{X_1 X_2}(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$
$p_{X_2 X_1=1}(x_2 x_1 = 1)$	$\frac{0.08}{0.40} = 0.2$	$\frac{0.12}{0.40} = 0.3$	$\frac{0.12}{0.40} = 0.3$	$\frac{0.08}{0.40} = 0.2$
$p_{X_2 X_1=2}(x_2 x_1 = 2)$	$\frac{0.06}{0.30} = 0.2$	$\frac{0.09}{0.30} = 0.3$	$\frac{0.09}{0.30} = 0.3$	$\frac{0.06}{0.30} = 0.2$
$p_{X_2 X_1=3}(x_2 x_1 = 3)$	$\frac{0.06}{0.30} = 0.2$	$\frac{0.09}{0.30} = 0.3$	$\frac{0.09}{0.30} = 0.3$	$\frac{0.06}{0.30} = 0.2$

Im Falle der Unabhängigkeit von X_1 und X_2 ändert sich die Verteilung von X_2 gegeben (oder im Wissen um) den Wert von X_1 also nicht und entspricht jeweils der Marginalverteilung von X_2 . Dies entspricht natürlich der Intuition der Unabhängigkeit von Ereignissen im Kontext elementarer Wahrscheinlichkeiten.

Definition (n unabhängige Zufallsvariablen)

$X := (X_1, \dots, X_n)$ sei ein n -dimensionaler Zufallsvektor mit Ergebnisraum $\mathcal{X} = \times_{i=1}^n \mathcal{X}_i$. Die n Zufallsvariablen $X_1, \dots, X_n$ heißen *unabhängig*, wenn für alle $S_i \in \mathcal{X}_i, i = 1, \dots, n$ gilt, dass

$$\mathbb{P}_X(X_1 \in S_1, \dots, X_n \in S_n) = \prod_{i=1}^n \mathbb{P}_{X_i}(X_i \in S_i). \quad (25)$$

Wenn der Zufallsvektor eine n -dimensionale WMF oder WDF p_X mit marginalen WMFen oder WDFen $p_{X_i}, i = 1, \dots, n$ besitzt, dann ist die Unabhängigkeit von $X_1, \dots, X_n$ gleichbedeutend mit der Faktorisierung der gemeinsamen WMF oder WDF, also mit

$$p_X(x_1, \dots, x_n) = \prod_{i=1}^n p_{X_i}(x_i). \quad (26)$$

Bemerkung

- Es handelt sich um eine direkte Generalisierung des zweidimensionalen Falls.

Definition (Unabhängig und identisch verteilte Zufallsvariablen)

n Zufallsvariablen $X_1, \dots, X_n$ heißen *unabhängig und identisch verteilt (u.i.v.)*, wenn

- (1) $X_1, \dots, X_n$ unabhängige Zufallsvariablen sind, und
- (2) die Marginalverteilungen der X_i übereinstimmen, also gilt, dass

$$p_{X_i} = p_{X_j} \text{ für alle } 1 \leq i, j \leq n. \quad (27)$$

Wenn die Zufallsvariablen $X_1, \dots, X_n$ unabhängig und identisch verteilt sind, und die i te Marginalverteilung $\mathbb{P}_X := \mathbb{P}_{X_i}$ ist, so schreibt man auch

$$X_1, \dots, X_n \sim \mathbb{P}_X. \quad (28)$$

Bemerkungen

- Man sagt kurz, dass $X_1, \dots, X_n$ u.i.v. sind.
- Im Englischen spricht man von *independent and identically distributed (i.i.d)* Zufallsvariablen.
- In der Statistik werden Stichproben meist mit u.i.v. Zufallsvariablen modelliert.
- n u.i.v. normalverteilte ZVen werden als $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ geschrieben.

Multivariate Verteilungen

- Zufallsvektoren
- Multivariate Verteilungen
- Marginalverteilungen
- Bedingte Verteilungen
- Unabhängige Zufallsvariablen
- **Selbstkontrollfragen**

1. Definieren Sie den Begriff des Zufallsvektors.
2. Definieren Sie den Begriff der multivariaten Verteilung eines Zufallsvektors.
3. Definieren Sie den Begriff der multivariaten WMF.
4. Definieren Sie den Begriff der multivariaten WDF.
5. Definieren Sie den Begriff der univariaten Marginalverteilung eines Zufallsvektors.
6. Wie berechnet man die WMF der i ten Komponente eines diskreten Zufallsvektors?
7. Wie berechnet man die WDF der i ten Komponente eines kontinuierlichen Zufallsvektors?
8. Definieren Sie den Begriff der Unabhängigkeit zweier Zufallsvariablen.
9. Wie erkennt man an der gemeinsamen WMF oder WDF eines zweidimensionalen Zufallsvektors, ob die Komponenten des Zufallsvektors unabhängig sind oder nicht?
10. Definieren Sie den Begriff der Unabhängigkeit von n Zufallsvariablen.
11. Definieren Sie den Begriff n unabhängig und identisch verteilter Zufallsvariablen.

Literatur

- DeGroot, M. H. and Schervish, M. J. (2012). *Probability and Statistics*. Addison-Wesley, Boston, 4th ed edition.
- Georgii, H.-O. (2009). *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. DeGruyter-Lehrbuch. de Gruyter, Berlin, 4., überarb. und erw. aufl edition.
- Wasserman, L. (2004). *All of Statistics*.



Inferenzstatistik

BSc Psychologie SoSe 2021

Prof. Dr. Dirk Ostwald

(4) Erwartungswert und Kovarianz

Literaturhinweise

Die in diesem Abschnitt vorgestellten Konzepte gehören zum Standard der Wahrscheinlichkeitstheorie. Die Lektüre von Georgii (2009, Abschnitte 4.1 und 4.3) wird empfohlen. In englischer Sprache finden sich entsprechende Darstellungen in Wasserman (2004, Sections 3.1 - 3.3) und DeGroot and Schervish (2012, Sections 4.1 - 4.3, 4.6). Die angegebenen Beweise für den Erwartungswert und die Varianz der univariaten Normalverteilung wenden sich an interessierte Leser:innen und sollten beim ersten Durcharbeiten übergangen werden.

Erwartungswert und Kovarianz

- Erwartungswert
- Varianz und Standardabweichung
- Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung
- Kovarianz und Korrelation
- Stichprobenkovarianz und Stichprobenkorrelation
- Selbstkontrollfragen

Erwartungswert und Kovarianz

- **Erwartungswert**
- Varianz und Standardabweichung
- Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung
- Kovarianz und Korrelation
- Stichprobenkovarianz und Stichprobenkorrelation
- Selbstkontrollfragen

Definition (Erwartungswert)

Es sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und X sei eine Zufallsvariable. Dann ist der *Erwartungswert von X* definiert als

- $\mathbb{E}(X) := \sum_{x \in \mathcal{X}} x p_X(x)$, wenn $X : \Omega \rightarrow \mathcal{X}$ diskret mit WMF p_X ist, und als
- $\mathbb{E}(X) := \int_{-\infty}^{\infty} x p_X(x) dx$, wenn $X : \Omega \rightarrow \mathbb{R}$ kontinuierlich mit WDF p_X ist.

Der Erwartungswert einer Zufallsvariable heißt *existent*, wenn er endlich ist.

Bemerkungen

- Der Erwartungswert ist eine skalare Zusammenfassung einer Verteilung.
- Intuitiv ist $\mathbb{E}(X) \approx \frac{1}{n} \sum_{i=1}^n X_i$ für eine große Zahl n von Kopien X_i von X .

Beispiel (Erwartungswert einer Bernoulli Zufallsvariable)

Es sei $X \sim \text{Bern}(\mu)$. Dann gilt $\mathbb{E}(X) = \mu$.

Beweis

X ist diskret mit $\mathcal{X} = \{0, 1\}$. Also gilt

$$\begin{aligned}\mathbb{E}(X) &= \sum_{x \in \{0,1\}} x \text{Bern}(x; \mu) \\ &= 0 \cdot \mu^0 (1 - \mu)^{1-0} + 1 \cdot \mu^1 (1 - \mu)^{1-1} \\ &= 1 \cdot \mu^1 (1 - \mu)^0 \\ &= \mu.\end{aligned}\tag{1}$$

□

Beispiel (Erwartungswert einer normalverteilten Zufallsvariable)

Es sei $X \sim N(\mu, \sigma^2)$. Dann gilt $\mathbb{E}(X) = \mu$.

Beweis

Wir halten zunächst ohne Beweis fest, dass

$$\int_{-\infty}^{\infty} \exp(-x^2) dx = \sqrt{\pi}. \quad (2)$$

Mit der Definition des Erwartungswerts für kontinuierliche Zufallsvariablen gilt

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right) dx. \quad (3)$$

Mit der Substitutionsregel

$$\int_{g(a)}^{g(b)} f(x) dx = \int_a^b f(g(x))g'(x) dx \quad (4)$$

und der Definition von

$$g : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto g(x) := \sqrt{2\sigma^2}x + \mu \text{ with } g'(x) = \sqrt{2\sigma^2}, \quad (5)$$

gilt dann

Beispiel (Erwartungswert einer normalverteilten Zufallsvariable)

$$\begin{aligned}\mathbb{E}(X) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx \\&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (\sqrt{2\sigma^2}x + \mu) \exp\left(-\frac{1}{2\sigma^2}\left((\sqrt{2\sigma^2}x + \mu) - \mu\right)^2\right) \sqrt{2\sigma^2} dx \\&= \frac{\sqrt{2\sigma^2}}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (\sqrt{2\sigma^2}x + \mu) \exp(-x^2) dx \\&= \frac{1}{\sqrt{\pi}} \left(\sqrt{2\sigma^2} \int_{-\infty}^{\infty} x \exp(-x^2) dx + \mu \int_{-\infty}^{\infty} \exp(-x^2) dx \right) \\&= \frac{1}{\sqrt{\pi}} \left(\sqrt{2\sigma^2} \int_{-\infty}^{\infty} x \exp(-x^2) dx + \mu\sqrt{\pi} \right)\end{aligned}\tag{6}$$

Eine Stammfunktion von $x \exp(-x^2)$ ist $-\frac{1}{2} \exp(-x^2)$. Mit

$$\lim_{x \rightarrow -\infty} \exp(-x^2) = 0 \text{ und } \lim_{x \rightarrow \infty} \exp(-x^2) = 0\tag{7}$$

verschwindet der Integralterm und wir erhalten

$$\mathbb{E}(X) = \frac{1}{\sqrt{\pi}} (\mu\sqrt{\pi}) = \mu.\tag{8}$$

□

Theorem (Eigenschaften des Erwartungswerts)

(1) (Linear-affine Transformation) Für eine Zufallsvariable X und $a, b \in \mathbb{R}$ gilt

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b. \quad (9)$$

(2) (Linearkombination) Für Zufallsvariablen $X_1, \dots, X_n$ und $a_1, \dots, a_n \in \mathbb{R}$ gilt

$$\mathbb{E}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i \mathbb{E}(X_i). \quad (10)$$

(3) (Faktorisierung bei Unabhängigkeit) Für unabhängige Zufallsvariablen $X_1, \dots, X_n$ gilt

$$\mathbb{E}\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n \mathbb{E}(X_i). \quad (11)$$

Bemerkung

- Die genannten Eigenschaften sind oft nützlich zur Berechnung von Erwartungswerten.

Beweis

Eigenschaft (1) folgt aus den Linearitätseigenschaften von Summen und Integralen. Wir betrachten den Fall einer kontinuierlichen Zufallsvariable X mit WDF p_X genauer. In diesem Fall gilt

$$\begin{aligned}\mathbb{E}(Y) &= \mathbb{E}(aX + b) \\ &= \int_{-\infty}^{\infty} (ax + b)p_X(x) dx \\ &= \int_{-\infty}^{\infty} axp_X(x) + bp_X(x) dx \\ &= a \int_{-\infty}^{\infty} xp_X(x) dx + b \int_{-\infty}^{\infty} p_X(x) dx \\ &= a\mathbb{E}(X) + b.\end{aligned}\tag{12}$$

Eigenschaft (2) folgt ebenfalls aus den Linearitätseigenschaften von Summen und Integralen. Wir betrachten den Fall von zwei kontinuierlichen Zufallsvariablen X_1 und X_2 mit bivariater WDF p_{X_1, X_2} genauer. In diesem Fall gilt

Beweis (fortgeführt)

$$\begin{aligned} & \mathbb{E} \left(\sum_{i=1}^2 a_i X_i \right) \\ &= \mathbb{E}(a_1 X_1 + a_2 X_2) \\ &= \iint_{\mathbb{R}^2} (a_1 x_1 + a_2 x_2) p_{X_1, X_2}(x_1, x_2) dx_1 dx_2 \\ &= \iint_{\mathbb{R}^2} a_1 x_1 p_{X_1, X_2}(x_1, x_2) + a_2 x_2 p_{X_1, X_2}(x_1, x_2) dx_1 dx_2 \\ &= a_1 \iint_{\mathbb{R}^2} x_1 p_{X_1, X_2}(x_1, x_2) dx_1 dx_2 + a_2 \iint_{\mathbb{R}^2} x_2 p_{X_1, X_2}(x_1, x_2) dx_1 dx_2 \\ &= a_1 \int_{-\infty}^{\infty} x_1 \left(\int_{-\infty}^{\infty} p_{X_1, X_2}(x_1, x_2) dx_2 \right) dx_1 + a_2 \int_{-\infty}^{\infty} x_2 \left(\int_{-\infty}^{\infty} p_{X_1, X_2}(x_1, x_2) dx_1 \right) dx_2 \\ &= a_1 \int_{-\infty}^{\infty} x_1 p_{X_1}(x_1) dx_1 + a_2 \int_{-\infty}^{\infty} x_2 p_{X_2}(x_2) dx_2 \\ &= a_1 \mathbb{E}(X_1) + a_2 \mathbb{E}(X_2) \\ &= \sum_{i=1}^2 a_i \mathbb{E}(X_i). \end{aligned} \tag{13}$$

Ein Induktionsargument erlaubt dann die Generalisierung vom bivariaten zum n -variaten Fall.

Erwartungswert

Beweis (fortgeführt)

Zu Eigenschaft (3) betrachten wir den Fall von n kontinuierlichen Zufallsvariablen mit gemeinsamer WDF $p_{X_1, \dots, X_n}$. Weil als $X_1, \dots, X_n$ unabhängig vorausgesetzt sind, gilt

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n p_{X_i}(x_i). \quad (14)$$

Weiterhin gilt also

$$\begin{aligned} \mathbb{E} \left(\prod_{i=1}^n x_i \right) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\prod_{i=1}^n x_i \right) p_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^n x_i \prod_{i=1}^n p_{X_i}(x_i) dx_1 \dots dx_n \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^n x_i p_{X_i}(x_i) dx_1 \dots dx_n \quad (15) \\ &= \prod_{i=1}^n \int_{-\infty}^{\infty} x_i p_{X_i}(x_i) dx_i \\ &= \prod_{i=1}^n \mathbb{E}(X_i). \end{aligned}$$

□

Erwartungswert und Kovarianz

- Erwartungswert
- **Varianz und Standardabweichung**
- Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung
- Kovarianz und Korrelation
- Stichprobenkovarianz und Stichprobenkorrelation
- Selbstkontrollfragen

Definition (Varianz und Standardabweichung)

Es sei X eine Zufallsvariable mit Erwartungswert $\mathbb{E}(X)$. Die *Varianz von X* ist definiert als

$$\mathbb{V}(X) := \mathbb{E}((X - \mathbb{E}(X))^2), \quad (16)$$

unter der Annahme, dass dieser Erwartungswert existiert. Die *Standardabweichung von X* ist definiert

$$\mathbb{S}(X) := \sqrt{\mathbb{V}(X)}. \quad (17)$$

Bemerkungen

- Die Varianz misst die Streuung (Breite) einer Verteilung.
- Quadration ist nötig wegen $\mathbb{E}(X - \mathbb{E}(X)) = \mathbb{E}(X) - \mathbb{E}(X) = 0$.
- Ein alternatives Maß für die Streuung einer Verteilung ist $\mathbb{E}(|X - \mathbb{E}(X)|)$.
- Ein weiteres Maß für die Streuung einer Verteilung ist die Entropie $-\mathbb{E}(\ln p_X(X))$.

Beispiel (Varianz einer Bernoulli Zufallsvariable)

Es sei $X \sim \text{Bern}(\mu)$. Dann ist die Varianz von X gegeben durch

$$\mathbb{V}(X) = \mu(1 - \mu). \quad (18)$$

Beweis

X ist eine diskrete Zufallsvariable und es gilt $\mathbb{E}(X) = \mu$. Also gilt

$$\begin{aligned} \mathbb{V}(X) &= \mathbb{E}\left((X - \mu)^2\right) \\ &= \sum_{x \in \{0,1\}} (x - \mu)^2 \text{Bern}(x; \mu) \\ &= (0 - \mu)^2 \mu^0 (1 - \mu)^{1-0} + (1 - \mu)^2 \mu^1 (1 - \mu)^{1-1} \\ &= \mu^2 (1 - \mu) + (1 - \mu)^2 \mu \\ &= \left(\mu^2 + (1 - \mu)\mu\right) (1 - \mu) \\ &= \left(\mu^2 + \mu - \mu^2\right) (1 - \mu) \\ &= \mu(1 - \mu). \end{aligned} \quad (19)$$

□

Theorem (Varianzverschiebungssatz)

Es sei X eine Zufallsvariable. Dann gilt

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \quad (20)$$

Beweis

Mit der Definition der Varianz und der Linearität des Erwartungswerts gilt

$$\begin{aligned} \mathbb{V}(X) &= \mathbb{E}\left((X - \mathbb{E}(X))^2\right) \\ &= \mathbb{E}\left(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2\right) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}(\mathbb{E}(X)^2) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2 \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2. \end{aligned} \quad (21)$$

□

Bemerkung

- Das Theorem ist nützlich, wenn $\mathbb{E}(X^2)$ und $\mathbb{E}(X)$ leicht zu berechnen sind.

Beispiel (Varianz einer normalverteilten Zufallsvariable)

Es sei $X \sim N(\mu, \sigma^2)$. Dann gilt $\mathbb{V}(X) = \sigma^2$.

Beweis

Wir halten zunächst fest, dass mit dem Varianzverschiebungssatz gilt, dass

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} x^2 \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right) dx - \mu^2 \quad (22)$$

Mit der Substitutionsregel

$$\int_a^b f(g(x))g'(x) dx = \int_{g(a)}^{g(b)} f(x) dx \quad (23)$$

und der Definition von

$$g : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \sqrt{2\sigma^2}x + \mu, g(-\infty) := -\infty, g(\infty) := \infty, \text{ with } g'(x) = \sqrt{2\sigma^2}, \quad (24)$$

kann das Integral auf der rechten Seite von Gleichung (22) dann als

Beweis (fortgeführt)

$$\begin{aligned} \int_{-\infty}^{\infty} x^2 \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx \\ &= \int_{-\infty}^{\infty} (\sqrt{2\sigma^2}x + \mu)^2 \exp\left(-\frac{1}{2\sigma^2}((\sqrt{2\sigma^2}x + \mu) - \mu)^2\right) \sqrt{2\sigma^2} dx \\ &= \sqrt{2\sigma^2} \int_{-\infty}^{\infty} (\sqrt{2\sigma^2}x + \mu)^2 \exp\left(-\frac{2\sigma^2 x^2}{2\sigma^2}\right) dx \\ &= \sqrt{2\sigma^2} \int_{-\infty}^{\infty} (\sqrt{2\sigma^2}x + \mu)^2 \exp(-x^2) dx. \end{aligned} \tag{25}$$

geschrieben werden. Also gilt

$$\begin{aligned} \mathbb{V}(X) &= \frac{\sqrt{2\sigma^2}}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (\sqrt{2\sigma^2}x + \mu)^2 \exp(-x^2) dx - \mu^2 \\ &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} (\sqrt{2\sigma^2}x)^2 + 2\sqrt{2\sigma^2}x\mu + \mu^2 \exp(-x^2) dx - \mu^2 \\ &= \frac{1}{\sqrt{\pi}} \left(2\sigma^2 \int_{-\infty}^{\infty} x^2 \exp(-x^2) dx + 2\sqrt{2\sigma^2}\mu \int_{-\infty}^{\infty} x \exp(-x^2) dx + \mu^2 \int_{-\infty}^{\infty} \exp(-x^2) dx \right) - \mu^2 \end{aligned} \tag{26}$$

Beweis (fortgeführt)

Wir halten weiterhin ohne Beweis fest, dass

$$\int_{-\infty}^{\infty} x \exp(-x^2) dx = 0 \text{ und } \int_{-\infty}^{\infty} \exp(-x^2) dx = \sqrt{\pi}. \quad (27)$$

Es ergibt sich dann

$$\begin{aligned} \mathbb{V}(X) &= \frac{1}{\sqrt{\pi}} \left(2\sigma^2 \int_{-\infty}^{\infty} x^2 \exp(-x^2) dx + \mu^2 \sqrt{\pi} \right) - \mu^2 \\ &= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} x^2 \exp(-x^2) dx + \mu^2 - \mu^2 \\ &= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} x^2 \exp(-x^2) dx \end{aligned} \quad (28)$$

Mit der partiellen Integrationsregel

$$\int_a^b f'(x)g(x) dx = f(x)g(x)|_a^b - \int_a^b f(x)g'(x) dx \quad (29)$$

und der Definition von

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := \exp(-x^2) \text{ with } f'(x) = -2 \exp(-x^2) \quad (30)$$

Beweis (fortgeführt)

und

$$g : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto g(x) := -\frac{1}{2}x \text{ with } g'(x) = -\frac{1}{2}, \quad (31)$$

so dass

$$f'(x)g(x) = -2 \exp(-x^2) \left(-\frac{1}{2}x\right) = x^2 \exp(-x^2), \quad (32)$$

gilt, ergibt sich dann

$$\begin{aligned} \mathbb{V}(X) &= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} x^2 \exp(-x^2) dx \\ &= \frac{2\sigma^2}{\sqrt{\pi}} \left(-\frac{1}{2}x \exp(-x^2) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \exp(-x^2) \left(-\frac{1}{2}\right) dx \right) \\ &= \frac{2\sigma^2}{\sqrt{\pi}} \left(-\frac{1}{2}x \exp(-x^2) \Big|_{-\infty}^{\infty} + \frac{1}{2} \int_{-\infty}^{\infty} \exp(-x^2) dx \right), \end{aligned} \quad (33)$$

Aus $\lim_{x \rightarrow \pm\infty} \exp(-x^2) = 0$ schließen wir, dass der erste Term in den Klammern auf der rechten Seite der obigen Gleichung gleich 0 ist. Schließlich ergibt sich

$$\mathbb{V}(X) = \frac{2\sigma^2}{\sqrt{\pi}} \left(\frac{1}{2} \int_{-\infty}^{\infty} \exp(-x^2) dx \right) = \frac{\sigma^2}{\sqrt{\pi}} \sqrt{\pi} = \sigma^2. \quad (34)$$

□

Theorem (Eigenschaften der Varianz)

(1) (Linear-affine Transformation) Für eine Zufallsvariable X und $a, b \in \mathbb{R}$ gelten

$$\mathbb{V}(aX + b) = a^2\mathbb{V}(X) \text{ und } \mathbb{S}(aX + b) = |a|\mathbb{S}(X). \quad (35)$$

(2) (Linearkombination bei Unabhängigkeit) Für unabhängige Zufallsvariablen $X_1, \dots, X_n$ und $a_1, \dots, a_n \in \mathbb{R}$ gilt

$$\mathbb{V}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \mathbb{V}(X_i). \quad (36)$$

Beweis

Um Eigenschaft (1) zu zeigen, definieren wir zunächst $Y := aX + b$ und halten fest, dass $\mathbb{E}(Y) = a\mathbb{E}(X) + b$. Für die Varianz von Y ergibt sich dann

$$\begin{aligned}\mathbb{V}(Y) &= \mathbb{E}\left((Y - \mathbb{E}(Y))^2\right) \\ &= \mathbb{E}\left((aX + b - a\mathbb{E}(X) - b)^2\right) \\ &= \mathbb{E}\left((aX - a\mathbb{E}(X))^2\right) \\ &= \mathbb{E}\left((a(X - \mathbb{E}(X)))^2\right) \\ &= \mathbb{E}\left(a^2(X - \mathbb{E}(X))^2\right) \\ &= a^2\mathbb{E}\left((X - \mathbb{E}(X))^2\right) \\ &= a^2\mathbb{V}(X)\end{aligned}\tag{37}$$

Wurzelziehen ergibt dann das Resultat für die Standardabweichung.

Für Eigenschaft (2) betrachten wir den Fall zweier unabhängiger Zufallsvariablen X_1 und X_2 genauer. Wir halten zunächst fest, dass in diesem Fall gilt, dass

$$\mathbb{E}(a_1X_1 + a_2X_2) = a_1\mathbb{E}(X_1) + a_2\mathbb{E}(X_2).\tag{38}$$

Beweis (fortgeführt)

Es ergibt sich also

$$\begin{aligned} & \mathbb{V} \left(\sum_{i=1}^2 a_i X_i \right) \\ &= \mathbb{V}(a_1 X_1 + a_2 X_2) \\ &= \mathbb{E} \left((a_1 X_1 + a_2 X_2 - \mathbb{E}(a_1 X_1 + a_2 X_2))^2 \right) \\ &= \mathbb{E} \left((a_1 X_1 + a_2 X_2 - a_1 \mathbb{E}(X_1) - a_2 \mathbb{E}(X_2))^2 \right) \\ &= \mathbb{E} \left((a_1 X_1 - a_1 \mathbb{E}(X_1) + a_2 X_2 - a_2 \mathbb{E}(X_2))^2 \right) \\ &= \mathbb{E} \left(((a_1 (X_1 - \mathbb{E}(X_1)) + (a_2 (X_2 - \mathbb{E}(X_2))))^2 \right) \\ &= \mathbb{E} \left((a_1 (X_1 - \mathbb{E}(X_1)))^2 - 2(a_1 (X_1 - \mathbb{E}(X_1))(a_2 (X_2 - \mathbb{E}(X_2))) + (a_2 (X_2 - \mathbb{E}(X_2)))^2 \right) \\ &= \mathbb{E} \left((a_1^2 (X_1 - \mathbb{E}(X_1))^2 - 2a_1 a_2 (X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2)) + a_2^2 (X_2 - \mathbb{E}(X_2))^2 \right) \\ &= a_1^2 \mathbb{E} \left((X_1 - \mathbb{E}(X_1))^2 \right) - 2a_1 a_2 \mathbb{E} \left((X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2)) \right) + a_2^2 \mathbb{E} \left((X_2 - \mathbb{E}(X_2))^2 \right) \\ &= a_1^2 \mathbb{V}(X_1) - 2a_1 a_2 \mathbb{E} \left((X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2)) \right) + a_2^2 \mathbb{V}(X_2) \\ &= \sum_{i=1}^2 a_i^2 \mathbb{V}(X_i) - 2a_1 a_2 \mathbb{E} \left((X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2)) \right) \end{aligned}$$

Beweis (fortgeführt)

Weil X_1 und X_2 unabhängig sind, ergibt sich mit den Eigenschaften des Erwartungswerts für unabhängige Zufallsvariablen, dass

$$\begin{aligned}\mathbb{E}((X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2))) &= \mathbb{E}((X_1 - \mathbb{E}(X_1))) \mathbb{E}((X_2 - \mathbb{E}(X_2))) \\ &= (\mathbb{E}(X_1) - \mathbb{E}(X_1))(\mathbb{E}(X_2) - \mathbb{E}(X_2)) \\ &= 0\end{aligned}\tag{39}$$

ist. Damit folgt also

$$\mathbb{V}\left(\sum_{i=1}^2 a_i X_i\right) = \sum_{i=1}^2 a_i^2 \mathbb{V}(X_i).\tag{40}$$

Ein Induktionsargument erlaubt dann die Generalisierung vom bivariaten zum n -variaten Fall.

□

Erwartungswert und Kovarianz

- Erwartungswert
- Varianz und Standardabweichung
- **Stichprobenmittel, -varianz, -standardabweichung**
- Kovarianz und Korrelation
- Stichprobenkovarianz und Stichprobenkorrelation
- Selbstkontrollfragen

Definition (Stichprobenmittel, -varianz, -standardabweichung)

$X_1, \dots, X_n$ seien Zufallsvariablen.

- Das *Stichprobenmittel* von $X_1, \dots, X_n$ ist definiert als der arithmetische Mittelwert

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i. \quad (41)$$

- Die *Stichprobenvarianz* von $X_1, \dots, X_n$ ist definiert als

$$S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2. \quad (42)$$

- Die *Stichprobenstandardabweichung* ist definiert als

$$S_n := \sqrt{S_n^2}. \quad (43)$$

Bemerkungen

- $\mathbb{E}(X)$, $\mathbb{V}(X)$, und $\mathbb{S}(X)$ sind Kennzahlen einer Zufallsvariable X .
- $\bar{X}_n$, S_n^2 , und S_n sind Kennzahlen einer Stichprobe $X_1, \dots, X_n$.
- $\bar{X}_n$, S_n^2 , und S_n sind Zufallsvariablen, ihre Realisationen werden mit $\bar{x}_n$, s_n^2 , und s_n bezeichnet.

Beispiel (Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung)

- Es seien $X_1, \dots, X_{10} \sim N(1, 2)$.
- Wir nehmen die folgenden Realisationen an

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
0.54	1.01	-3.28	0.35	2.75	-0.51	2.32	1.49	0.96	1.25

- Die Stichprobenmittelrealisation ist

$$\bar{x}_{10} = \frac{1}{10} \sum_{i=1}^{10} x_i = \frac{6.88}{10} = 0.68. \quad (44)$$

- Die Stichprobenvarianzrealisation ist

$$s_{10}^2 = \frac{1}{9} \sum_{i=1}^{10} (x_i - \bar{x}_{10})^2 = \frac{1}{9} \sum_{i=1}^{10} (x_i - 0.68)^2 = \frac{25.37}{9} = 2.82. \quad (45)$$

- Die Stichprobenstandardabweichungrealisation ist

$$s_{10} = \sqrt{s_{10}^2} = \sqrt{2.82} = 1.68. \quad (46)$$

Erwartungswert und Kovarianz

- Erwartungswert
- Varianz und Standardabweichung
- Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung
- **Kovarianz und Korrelation**
- Stichprobenkovarianz und Stichprobenkorrelation
- Selbstkontrollfragen

Definition (Kovarianz und Korrelation)

Die *Kovarianz* zweier Zufallsvariablen X und Y ist definiert als

$$\mathbb{C}(X, Y) := \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))). \quad (47)$$

Die *Korrelation* zweier Zufallsvariablen X und Y ist definiert als

$$\rho(X, Y) := \frac{\mathbb{C}(X, Y)}{\sqrt{\mathbb{V}(X)}\sqrt{\mathbb{V}(Y)}} = \frac{\mathbb{C}(X, Y)}{\mathbb{S}(X)\mathbb{S}(Y)}. \quad (48)$$

Bemerkungen

- Die Kovarianz von X mit sich selbst ist die Varianz von X ,

$$\mathbb{C}(X, X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{V}(X). \quad (49)$$

- $\rho(X, Y)$ wird auch *Korrelationskoeffizient* von X und Y genannt.
- Wenn $\rho(X, Y) = 0$ ist, werden X und Y *unkorreliert* genannt.
- Wir zeigen später mit der Cauchy-Schwarz Ungleichung, dass $-1 \leq \rho(X, Y) \leq 1$.

Beispiel (Kovarianz und Korrelation zweier diskreter Zufallsvariablen)

Es sei $X := (X_1, X_2)$ ein Zufallsvektor mit WMF p_{X_1, X_2} definiert durch (Lange and Mosler, 2017)

$p_{X_1, X_2}(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$p_{X_1}(x_1)$
$x_1 = 1$	0.10	0.05	0.15	0.30
$x_1 = 2$	0.60	0.05	0.05	0.70
$p_{X_2}(x_2)$	0.70	0.10	0.20	

X_1, X_2 sind also zwei Zufallsvariablen mit einer definierten bivariaten Verteilung. Um $\mathbb{C}(X_1, X_2)$ und $\rho(X_1, X_2)$ zu berechnen, halten wir zunächst fest, dass

$$\mathbb{E}(X_1) = \sum_{x_1=1}^2 x_1 p_{X_1}(x_1) = 1 \cdot 0.3 + 2 \cdot 0.7 = 1.7 \quad (50)$$

und

$$\mathbb{E}(X_2) = \sum_{x_2=1}^3 x_2 p_{X_2}(x_2) = 1 \cdot 0.7 + 2 \cdot 0.1 + 3 \cdot 0.2 = 1.5. \quad (51)$$

Mit der Definition der Kovarianz von X_1 und X_2 , gilt dann

$$\begin{aligned}C(X_1, X_2) &= \mathbb{E}((X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2))) \\&= \sum_{x_1=1}^2 \sum_{x_2=1}^3 (x_1 - \mathbb{E}(X_1))(x_2 - \mathbb{E}(X_2))p_{X_1, X_2}(x_1, x_2) \\&= \sum_{x_1=1}^2 \sum_{x_2=1}^3 (x_1 - 1.7)(x_2 - 1.5)p_{X_1, X_2}(x_1, x_2) \\&= \sum_{x_1=1}^2 (x_1 - 1.7)(1 - 1.5)p_{X_1, X_2}(x_1, 1) \\&\quad + (x_1 - 1.7)(2 - 1.5)p_{X_1, X_2}(x_1, 2) \\&\quad + (x_1 - 1.7)(3 - 1.5)p_{X_1, X_2}(x_1, 3) \\&= (1 - 1.7)(1 - 1.5)p_{X_1, X_2}(1, 1) + (1 - 1.7)(2 - 1.5)p_{X_1, X_2}(1, 2) + (1 - 1.7)(3 - 1.5)p_{X_1, X_2}(1, 3) \\&\quad + (2 - 1.7)(1 - 1.5)p_{X_1, X_2}(2, 1) + (2 - 1.7)(2 - 1.5)p_{X_1, X_2}(2, 2) + (2 - 1.7)(3 - 1.5)p_{X_1, X_2}(2, 3) \\&= (-0.7) \cdot (-0.5) \cdot 0.10 \quad + (-0.7) \cdot 0.5 \cdot 0.05 \quad + (-0.7) \cdot 1.5 \cdot 0.15 \\&\quad + 0.3 \cdot (-0.5) \cdot 0.60 \quad + 0.3 \cdot 0.5 \cdot 0.05 \quad + 0.3 \cdot 1.5 \cdot 0.05 \\&= 0.035 - 0.0175 - 0.1575 - 0.09 + 0.0075 + 0.0225 \\&= -0.2.\end{aligned}$$

Theorem (Kovarianzverschiebungssatz)

X und Y seien Zufallsvariablen. Dann gilt

$$\mathbb{C}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y). \quad (52)$$

Beweis

Mit der Definition der Kovarianz gilt

$$\begin{aligned} \mathbb{C}(X, Y) &= \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) \\ &= \mathbb{E}(XY - X\mathbb{E}(Y) - \mathbb{E}(X)Y + \mathbb{E}(X)\mathbb{E}(Y)) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y). \end{aligned} \quad (53)$$

□

Bemerkungen

- Das Theorem ist nützlich, wenn $\mathbb{E}(XY)$, $\mathbb{E}(X)$, und $\mathbb{E}(Y)$ leicht zu berechnen sind.
- Für $Y = X$ erhalten wir $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.

Theorem (Korrelation und Unabhängigkeit)

X und Y seien zwei Zufallsvariablen. Wenn X und Y unabhängig sind, dann ist $\mathbb{C}(X, Y) = 0$ und X und Y sind unkorreliert. Ist dagegen $\mathbb{C}(X, Y) = 0$ und sind X und Y somit unkorreliert, dann sind X und Y nicht notwendigerweise unabhängig.

Beweis

Wir zeigen zunächst, dass aus der Unabhängigkeit von X und Y $\mathbb{C}(X, Y) = 0$ folgt. Hierzu halten wir zunächst fest, dass für unabhängige Zufallsvariablen gilt, dass

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y). \quad (54)$$

Mit dem Kovarianzverschiebungssatz folgt dann

$$\mathbb{C}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) = 0. \quad (55)$$

Mit der Definition des Korrelationskoeffizienten folgt dann unmittelbar, dass $\rho(X, Y) = 0$ und X und Y somit unkorreliert sind.

Beweis (fortgeführt)

Wir zeigen nun durch Angabe eines Beispiels, dass die Kovarianz von abhängigen Zufallsvariablen X und Y null sein kann. Zu diesem Zweck betrachten wir den Fall zweier diskreter Zufallsvariablen X und Y mit Ergebnisräumen $\mathcal{X} = \{-1, 0, 1\}$ und $\mathcal{Y} = \{0, 1\}$, marginaler WMF von X gegeben durch $p_X(X = x) = 1/3$ für $x \in \mathcal{X}$ und der Definition $Y := X^2$. Wir halten dann zunächst fest, dass

$$\mathbb{E}(X) = \sum_{x \in \mathcal{X}} xp_X(X = x) = -1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = 0 \quad (56)$$

und

$$\mathbb{E}(XY) = \mathbb{E}(XX^2) = \mathbb{E}(X^3) = \sum_{x \in \mathcal{X}} x^3 p_X(X = x) = -1^3 \cdot \frac{1}{3} + 0^3 \cdot \frac{1}{3} + 1^3 \cdot \frac{1}{3} = 0. \quad (57)$$

Mit dem Kovarianzverschiebungssatz ergibt sich dann

$$\mathbb{C}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(X^3) - \mathbb{E}(X)\mathbb{E}(Y) = 0 - 0 \cdot \mathbb{E}(Y) = 0. \quad (58)$$

Die Kovarianz von X und Y ist also null. Wie unten gezeigt faktorisiert die gemeinsame WMF von X und Y jedoch nicht, und somit sind X und Y nicht unabhängig.

Beweis (fortgeführt)

Die Definition of $Y := X^2$ impliziert die folgende bedingte WMF

$p_{Y X}(y x)$	$x = -1$	$x = 0$	$x = 1$
$y = 0$	0	1	0
$y = 1$	1	0	1

Die marginale WMF p_X und die bedingte WMF $p_{Y|X}$ implizieren die gemeinsame WMF

$p_{X,Y}(x, y)$	$x = -1$	$x = 0$	$x = 1$	$p_Y(y)$
$y = 0$	0	1/3	0	1/3
$y = 1$	1/3	0	1/3	2/3
$p_X(x)$	1/3	1/3	1/3	

Es gilt also zum Beispiel

$$p_{X,Y}(x = -1, y = 0) = 0 \neq \frac{1}{9} = \frac{1}{3} \cdot \frac{1}{3} = p_X(x = -1)p_Y(y = 0) \quad (59)$$

und damit sind X und Y nicht unabhängig.

□

Theorem (Varianzen von Summen und Differenzen von Zufallsvariablen)

X und Y seien zwei Zufallsvariablen und es seien $a, b, c \in \mathbb{R}$. Dann gilt

$$\mathbb{V}(aX + bY + c) = a^2\mathbb{V}(X) + b^2\mathbb{V}(Y) + 2ab\mathbb{C}(X, Y). \quad (60)$$

Speziell gelten

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\mathbb{C}(X, Y) \quad (61)$$

und

$$\mathbb{V}(X - Y) = \mathbb{V}(X) + \mathbb{V}(Y) - 2\mathbb{C}(X, Y) \quad (62)$$

Bemerkungen

- Varianzen von Zufallsvariablen addieren sich nicht einfach.
- Die Varianz der Summe zweier Zufallsvariablen hängt von ihrer Kovarianz ab.

Beweis

Wir halten zunächst fest, dass

$$\mathbb{E}(aX + bY + c) = a\mathbb{E}(X) + b\mathbb{E}(Y) + c. \quad (63)$$

Es ergibt sich also

$$\begin{aligned} & \mathbb{V}(aX + bY + c) \\ &= \mathbb{E} \left((aX + bY + c - a\mathbb{E}(X) - b\mathbb{E}(Y) - c)^2 \right) \\ &= \mathbb{E} \left((a(X - \mathbb{E}(X)) + b(Y - \mathbb{E}(Y)))^2 \right) \\ &= \mathbb{E} \left(a^2(X - \mathbb{E}(X))^2 + b^2(Y - \mathbb{E}(Y))^2 + 2ab(X - \mathbb{E}(X))(Y - \mathbb{E}(Y)) \right) \\ &= a^2\mathbb{E} \left((X - \mathbb{E}(X))^2 \right) + b^2\mathbb{E} \left((Y - \mathbb{E}(Y))^2 \right) + 2ab\mathbb{E} \left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)) \right) \\ &= a^2\mathbb{V}(X) + b^2\mathbb{V}(Y) + 2ab\mathbb{C}(X, Y) \end{aligned} \quad (64)$$

Die Spezialfälle folgen dann direkt mit $a := b := 1$ und $a := 1, b := -1$, respektive.

□

Erwartungswert und Kovarianz

- Erwartungswert
- Varianz und Standardabweichung
- Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung
- Kovarianz und Korrelation
- **Stichprobenkovarianz und Stichprobenkorrelation**
- Selbstkontrollfragen

Definition (Stichprobenkovarianz und Stichprobenkorrelation)

$(X_1, Y_1), \dots, (X_n, Y_n)$ seien zweidimensionale Zufallsvektoren.

- Das Stichprobenmittel der $(X_1, Y_1), \dots, (X_n, Y_n)$ ist definiert als

$$\overline{(X, Y)}_n := (\bar{X}_n, \bar{Y}_n) = \left(\frac{1}{n} \sum_{i=1}^n X_i, \frac{1}{n} \sum_{i=1}^n Y_i \right). \quad (65)$$

- Die Stichprobenkovarianz der $(X_1, Y_1), \dots, (X_n, Y_n)$ ist definiert als

$$C_n := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n). \quad (66)$$

- Der Stichprobenkorrelationskoeffizient der $(X_1, Y_1), \dots, (X_n, Y_n)$ ist definiert als

$$R_n := \frac{C_n}{S_{X,n} S_{Y,n}}, \quad (67)$$

wobei $S_{X,n}$ und $S_{Y,n}$ die Stichprobenstandardabweichungen von $X_1, \dots, X_n$ und $Y_1, \dots, Y_n$, bezeichnen.

Bemerkungen

- $\overline{(X, Y)}_n$, C_n , und R_n sind Zufallsvariablen
- $\overline{(x, y)}_n$, c_n , und r_n bezeichnen ihre Realisierungen

Beispiel (Stichprobenkovarianz und Stichprobenkorrelation)

- Es seien $(X_1, Y_1), \dots, (X_{10}, Y_{10})$ zweidimensionale Zufallsvariablen.
- Wir nehmen folgende Realisierungen an

$$\begin{array}{cccccccccc} (x_1, y_1) & (x_2, y_2) & (x_3, y_3) & (x_4, y_4) & (x_5, y_5) & (x_6, y_6) & (x_7, y_7) & (x_8, y_8) & (x_9, y_9) & (x_{10}, y_{10}) \\ (0.8, -0.7) & (1.1, 1.6) & (-0.8, 1.1) & (-0.2, 0.1) & (1.1, 0.4) & (0.5, 1.5) & (1.3, -1.2) & (1.8, 0.6) & (0.4, 0.2) & (1.5, -1.0) \end{array}$$

- Die Stichprobenmittelrealisation ergibt sich zu

$$\overline{(x, y)}_{10} = \left(\frac{1}{10} \sum_{i=1}^{10} x_i, \frac{1}{10} \sum_{i=1}^{10} y_i \right) = (0.75, 0.26). \quad (68)$$

- Die Stichprobenstandardabweichungrealisationen ergeben sich zu

$$s_{X,n} = \sqrt{\frac{1}{9} \sum_{i=1}^{10} (x_i - \bar{x}_{10})^2} = 0.79 \text{ und } s_{Y,n} = \sqrt{\frac{1}{9} \sum_{i=1}^{10} (y_i - \bar{y}_{10})^2} = 0.99. \quad (69)$$

- Die Stichprobenkovarianz- und Stichprobenkorrelationsrealisationen ergeben sich zu

$$c_n = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n) = -0.26 \text{ und } r_n = \frac{c_n}{s_{x,n} s_{y,n}} = -0.33. \quad (70)$$

Erwartungswert und Kovarianz

- Erwartungswert
- Varianz und Standardabweichung
- Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung
- Kovarianz und Korrelation
- Stichprobenkovarianz und Stichprobenkorrelation
- **Selbstkontrollfragen**

1. Definieren und interpretieren Sie den Erwartungswert einer Zufallsvariable.
2. Berechnen Sie den Erwartungswert einer Bernoulli Zufallsvariable.
3. Nennen Sie die Linearitäts- und Multiplikationseigenschaften des Erwartungswerts.
4. Definieren und interpretieren Sie die Varianz einer Zufallsvariable.
5. Drücken Sie $\mathbb{E}(X^2)$ mithilfe der Varianz und des Erwartungswerts von X aus.
6. Was ist $\mathbb{V}(aX)$ für konstantes $a \in \mathbb{R}$?
7. Definieren Sie die Kovarianz und Korrelation zweier Zufallsvariablen X und Y .
8. Schreiben Sie die Kovarianz zweier Zufallsvariablen mithilfe ihrer Erwartungswerte.
9. Was ist die Varianz der Summe zweier Zufallsvariablen X und Y bei Unabhängigkeit?
10. Was ist die Varianz der Summe zweier Zufallsvariablen X und Y im Allgemeinen?

11. Gegeben seien folgende Realisationen von Zufallsvektoren $(X_1, Y_1), \dots, (X_5, Y_5)$:

(x_1, y_1)	(x_2, y_2)	(x_3, y_3)	(x_4, y_4)	(x_5, y_5)
(5.2, 4.1)	(-1.1, -0.8)	(2.7, 2.1)	(-3.2, -2.8)	(0.1, -0.2)

Berechnen Sie bitte mithilfe eines Taschenrechners das Stichprobenmittel, die Stichprobenvarianz, und die Stichprobenstandardabweichung der Realisationen von $X_1, \dots, X_5$, das Stichprobenmittel, die Stichprobenvarianz, und die Stichprobenstandardabweichung der Realisationen von $Y_1, \dots, Y_5$, und die Stichprobenkovarianz und die Stichprobenkorrelation der Realisationen von $(X_1, Y_1), \dots, (X_5, Y_5)$. Überprüfen Sie Ihre Ergebnisse mithilfe der R Funktionen `mean()`, `var()`, `sd()`, `cov()`, und `cor()`.

Literatur

- DeGroot, M. H. and Schervish, M. J. (2012). *Probability and Statistics*. Addison-Wesley, Boston, 4th ed edition.
- Georgii, H.-O. (2009). *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. DeGruyter-Lehrbuch. de Gruyter, Berlin, 4., überarb. und erw. aufl edition.
- Lange, T. and Mosler, K. (2017). *Statistik kompakt*. Springer-Lehrbuch. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Wasserman, L. (2004). *All of Statistics*.



Inferenzstatistik

BSc Psychologie SoSe 2021

Prof. Dr. Dirk Ostwald

(5) Ungleichungen und Grenzwerte

Literaturhinweise

Bei dem vorgestellten Konzepten dieses Abschnittes handelt es sich nicht um eine umfassende Einführung in die Theorie der Ungleichungen und Grenzwerte der Wahrscheinlichkeitstheorie. Es handelt sich lediglich um eine Zusammenstellung zentraler Resultate, auf die in späteren Abschnitten zurück gegriffen werden soll. Die Lektüre von (Georgii, 2009, Abschnitte 5.1 und 5.3) wird empfohlen. Englischsprachige Darstellungen finden sich z.B. in Wasserman (2004, Abschnitte 4.1 und 4.2) und Casella and Berger (2012, Abschnitte 3.6 und 4.7).

Ungleichungen und Grenzwerte

- Wahrscheinlichkeitsungleichungen
- Erwartungswertungleichungen
- Gesetze der großen Zahl
- Zentrale Grenzwertsätze
- Selbstkontrollfragen

Ungleichungen und Grenzwerte

- **Wahrscheinlichkeitsungleichungen**
- Erwartungswertungleichungen
- Gesetze der großen Zahl
- Zentrale Grenzwertsätze
- Selbstkontrollfragen

Theorem (Markov Ungleichung)

Es sei X eine Zufallsvariable mit $\mathbb{P}(X \geq 0) = 1$. Dann gilt für alle $x \in \mathbb{R}$, dass

$$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}(X)}{x}. \quad (1)$$

Bemerkungen

- Weil $\mathbb{P}(X \geq 0) = 1$ gilt, sagt man auch, dass X eine *nicht-negative* Zufallsvariable ist.
- Die Ungleichung setzt Überschreitungswahrscheinlichkeiten und Erwartungswerte in Bezug.
- Gilt z.B. für eine nichtnegative Zufallsvariable X , dass $\mathbb{E}(X) = 1$, dann ist $\mathbb{P}(X \geq 100) \leq 0.01$.

Beweis

Wir betrachten den Fall einer kontinuierlichen Zufallsvariable X mit WDF p . Wir halten zunächst fest, dass

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} \xi p(\xi) d\xi = \int_0^{\infty} \xi p(\xi) d\xi = \int_0^x \xi p(\xi) d\xi + \int_x^{\infty} \xi p(\xi) d\xi, \quad (2)$$

weil X nicht-negativ ist. Es folgt dann

$$\mathbb{E}(X) \geq \int_x^{\infty} \xi p(\xi) d\xi \geq \int_x^{\infty} x p(\xi) d\xi = x \int_x^{\infty} p(\xi) d\xi = x \mathbb{P}(X \geq x). \quad (3)$$

Dabei gilt die erste Ungleichung weil

$$\int_{-\infty}^x \xi p(\xi) d\xi = 0 \text{ für } x \leq 0 \text{ und } \int_{-\infty}^x \xi p(\xi) d\xi > 0 \text{ für } x > 0. \quad (4)$$

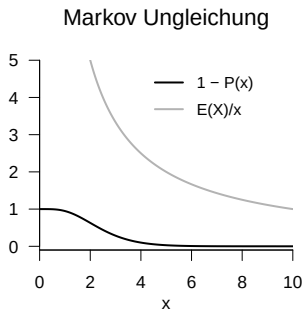
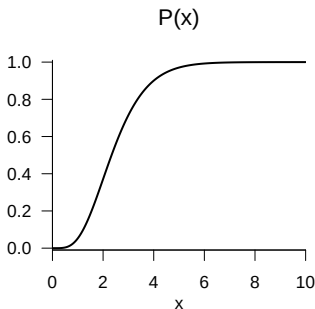
und die zweite Ungleichung gilt, weil $x \leq \xi$ für $\xi \in [x, \infty[$. Es folgt also, dass

$$\mathbb{E}(X) \geq x \mathbb{P}(X \geq x) \Leftrightarrow \mathbb{P}(X \geq x) \leq \frac{\mathbb{E}(X)}{x}. \quad (5)$$

□

Beispiel ($X \sim G(\alpha, \beta)$)

- Wir halten ohne Beweis fest, dass für $X \sim G(\alpha, \beta)$ gilt, dass $\mathbb{E}(X) = \alpha\beta$.
- Wir betrachten den Fall $\alpha := 5, \beta := 2$, so dass $G(x; 5, 2) = \chi^2(10)$



Theorem (Chebyshev Ungleichung)

Es sei X eine Zufallsvariable mit Varianz $\mathbb{V}(X)$. Dann gilt für alle $x \in \mathbb{R}$

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq x) \leq \frac{\mathbb{V}(X)}{x^2}. \quad (6)$$

Bemerkungen

- Die Chebyshev Ungleichung setzt Abweichungen vom Erwartungswert in Bezug zur Varianz.
- Zum Beispiel gilt

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geq 3\sqrt{\mathbb{V}(X)}\right) \leq \frac{\mathbb{V}(X)}{\left(3\sqrt{\mathbb{V}(X)}\right)^2} = \frac{1}{9}.$$

Beweis

Wir halten zunächst fest, dass für $a, b \in \mathbb{R}$ gilt

$$a^2 \geq b^2 \Leftrightarrow |a| \geq b. \quad (7)$$

Dazu betrachten wir die folgenden vier möglichen Fälle

$$\begin{array}{llllll} |a| \geq b \text{ für } a > 0, b > 0 & \Leftrightarrow & a \geq b & \Leftrightarrow & a^2 \geq b^2 \\ |a| \geq b \text{ für } a > 0, b < 0 & \Leftrightarrow & a \geq b & \Leftrightarrow & a^2 \geq b^2 \\ |a| \geq b \text{ für } a < 0, b > 0 & \Leftrightarrow & -a \geq b & \Leftrightarrow & (-a)^2 \geq b^2 = a^2 \geq b^2 \\ |a| \geq b \text{ für } a < 0, b < 0 & \Leftrightarrow & -a \geq b & \Leftrightarrow & (-a)^2 \geq b^2 = a^2 \geq b^2 \end{array}$$

Als nächstes definieren wir $Y := (X - \mathbb{E}(X))^2$. Dann folgt aus der Markov Ungleichung

$$\begin{aligned} \mathbb{P}(Y \geq x^2) &\leq \frac{\mathbb{E}(Y)}{x^2} \\ \Leftrightarrow \mathbb{P}((X - \mathbb{E}(X))^2 \geq x^2) &\leq \frac{\mathbb{E}((X - \mathbb{E}(X))^2)}{x^2} \\ \Leftrightarrow \mathbb{P}(|X - \mathbb{E}(X)| \geq x) &\leq \frac{\mathbb{V}(X)}{x^2}. \end{aligned} \quad (8)$$

□

Ungleichungen und Grenzwerte

- Wahrscheinlichkeitsungleichungen
- **Erwartungswertungleichungen**
- Gesetze der großen Zahl
- Zentrale Grenzwertsätze
- Selbstkontrollfragen

Theorem (Cauchy-Schwarz-Ungleichung)

Es seien X, Y zwei Zufallsvariablen und $\mathbb{E}(XY)$ endlich. Dann gilt

$$\mathbb{E}(XY)^2 \leq \mathbb{E}(X^2) \mathbb{E}(Y^2) \quad (9)$$

Bemerkungen

- Analog gilt für Vektoren $x, y \in \mathbb{R}^n$, dass $\langle x, y \rangle^2 \leq \|x\| \cdot \|y\|$.
- Die Korrelationsungleichung ist eine direkte Konsequenz der Cauchy-Schwarz-Ungleichung.

Erwartungswertungleichungen

Beweis

Wir betrachten den Fall, dass $0 < \mathbb{E}(X^2) < \infty$ und $0 < \mathbb{E}(Y^2) < \infty$. Für alle $a, b \in \mathbb{R}$ gilt dann

$$0 \leq \mathbb{E}((aX + bY)^2) \text{ und } 0 \leq \mathbb{E}((aX - bY)^2). \quad (10)$$

Also folgt

$$0 \leq a^2 \mathbb{E}(X^2) + b^2 \mathbb{E}(Y^2) + 2ab \mathbb{E}(XY) \text{ und } 0 \leq a^2 \mathbb{E}(X^2) + b^2 \mathbb{E}(Y^2) - 2ab \mathbb{E}(XY). \quad (11)$$

Definition von $a := \sqrt{\mathbb{E}(Y^2)}$ und $b := \sqrt{\mathbb{E}(X^2)}$ ergibt dann

$$\begin{aligned} 0 &\leq \mathbb{E}(Y^2)\mathbb{E}(X^2) + \mathbb{E}(X^2)\mathbb{E}(Y^2) + 2\sqrt{\mathbb{E}(Y^2)}\sqrt{\mathbb{E}(X^2)}\mathbb{E}(XY) \\ \Leftrightarrow -2\sqrt{\mathbb{E}(Y^2)}\sqrt{\mathbb{E}(X^2)}\mathbb{E}(XY) &\leq 2\mathbb{E}(X^2)\mathbb{E}(Y^2) \\ \Leftrightarrow -\sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)}\mathbb{E}(XY) &\leq \sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)} \\ \Leftrightarrow -\mathbb{E}(XY) &\leq \sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)}, \end{aligned} \quad (12)$$

und analog

$$\begin{aligned} 0 &\leq \mathbb{E}(Y^2)\mathbb{E}(X^2) + \mathbb{E}(X^2)\mathbb{E}(Y^2) - 2\sqrt{\mathbb{E}(Y^2)}\sqrt{\mathbb{E}(X^2)}\mathbb{E}(XY) \\ \Leftrightarrow 2\sqrt{\mathbb{E}(Y^2)}\sqrt{\mathbb{E}(X^2)}\mathbb{E}(XY) &\leq 2\mathbb{E}(X^2)\mathbb{E}(Y^2) \\ \Leftrightarrow \sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)}\mathbb{E}(XY) &\leq \sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)} \\ \Leftrightarrow \mathbb{E}(XY) &\leq \sqrt{\mathbb{E}(Y^2)\mathbb{E}(X^2)}. \end{aligned} \quad (13)$$

Zusammengenommen ergibt sich für den betrachteten Fall also die Cauchy-Schwarz-Ungleichung. Ein vollständiger Beweis findet sich bei DeGroot und Schervish (2012, Theorem 4.6.2).

□

Theorem (Korrelationsungleichung)

X und Y seien Zufallsvariablen mit $\mathbb{V}(X), \mathbb{V}(Y) > 0$. Dann gilt

$$\rho(X, Y)^2 = \frac{\mathbb{C}(X, Y)^2}{\mathbb{V}(X)\mathbb{V}(Y)} \leq 1. \quad (14)$$

Bemerkung

- Es gilt also $\rho(X, Y)^2 \leq 1 \Leftrightarrow |\rho(X, Y)| \leq 1 \Leftrightarrow \rho(X, Y) \in [-1, 1]$.

Beweis

Mit der Cauchy-Schwarz-Ungleichung für zwei Zufallsvariablen U und V gilt, dass

$$(\mathbb{E}(UV))^2 \leq \mathbb{E}(U^2)\mathbb{E}(V^2). \quad (15)$$

Wir definieren nun $U := X - \mathbb{E}(X)$ und $V := Y - \mathbb{E}(Y)$. Dann besagt die Cauchy-Schwarz Ungleichung, dass

$$\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))^2 \leq \mathbb{E}((X - \mathbb{E}(X))^2) \mathbb{E}((Y - \mathbb{E}(Y))^2). \quad (16)$$

Also gilt

$$\mathbb{C}(X, Y)^2 \leq \mathbb{V}(X)\mathbb{V}(Y) \Leftrightarrow \frac{\mathbb{C}(X, Y)^2}{\mathbb{V}(X)\mathbb{V}(Y)} \leq 1. \quad (17)$$

□

Ungleichungen und Grenzwerte

- Wahrscheinlichkeitsungleichungen
- Erwartungswertungleichungen
- **Gesetze der großen Zahl**
- Zentrale Grenzwertsätze
- Selbstkontrollfragen

Überblick

- Es gibt ein *Schwaches Gesetz der großen Zahl* und ein *Starkes Gesetz der großen Zahl*.
- Intuitiv besagen beide Gesetze, dass sich **das Stichprobenmittel** von unabhängigen und identisch verteilten Zufallsvariablen für eine große Anzahl an Zufallsvariablen dem Erwartungswert der zugrundeliegenden Verteilung nähert.
- Das Schwache und das Starke Gesetz der großen Zahl unterscheiden sich in Hinblick auf die zu ihrer Formulierung benutzen Formen der *Konvergenz von Zufallsvariablen*.
 - Das Schwache Gesetz basiert auf der *Konvergenz in Wahrscheinlichkeit*.
 - Das Starke Gesetz basiert auf der *fast sicheren Konvergenz*.
- Wir begnügen uns mit Konvergenz in Wahrscheinlichkeit und dem Schwachen Gesetz.

Definition (Konvergenz in Wahrscheinlichkeit)

Eine Folge von Zufallsvariable $X_1, X_2, \dots$ *konvergiert gegen eine Zufallsvariable X in Wahrscheinlichkeit*, wenn für jedes noch so kleine $\epsilon > 0$ gilt, dass

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| < \epsilon) = 1 \Leftrightarrow \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0 \quad (18)$$

Die Konvergenz von $X_1, X_2, \dots$ gegen X in Wahrscheinlichkeit wird geschrieben als

$$X_n \xrightarrow[n \rightarrow \infty]{P} X \quad (19)$$

Bemerkungen

- $X_n \xrightarrow[n \rightarrow \infty]{P} X$ heißt, dass sich die Wahrscheinlichkeit, dass X_n in dem zufälligen Intervall $]X - \epsilon, X + \epsilon[$ liegt, unabhängig davon, wie klein dieses Intervall sein mag, 1 nähert, wenn n gegen Unendlich strebt.
- Intuitiv heißt das, dass sich für eine konstante Zufallsvariable $X := a$ die Verteilung von X_n mehr und mehr um a konzentriert, wenn n gegen Unendlich strebt.

Theorem (Schwaches Gesetz der großen Zahl)

$X_1, \dots, X_n$ seien unabhängig und gleichverteilte Zufallsvariablen mit $\mathbb{E}(X_i) = \mu$ für alle $i = 1, \dots, n$. Weiterhin bezeichne

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i \quad (20)$$

das Stichprobenmittel der $X_i, i = 1, \dots, n$. Dann konvergiert $\bar{X}_n$ in Wahrscheinlichkeit gegen μ ,

$$\bar{X}_n \xrightarrow[n \rightarrow \infty]{P} \mu. \quad (21)$$

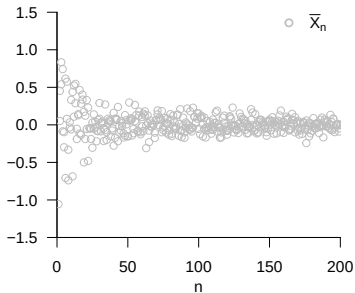
Bemerkung

- Für einen Beweis siehe z.B. Georgii (2009, Abschnitt 5.1).
- $\bar{X}_n \xrightarrow[n \rightarrow \infty]{P} \mu$ heißt, dass die Wahrscheinlichkeit, dass das Stichprobenmittel nahe dem Erwartungswert der zugrundeliegenden Verteilung liegt, sich 1 nähert, wenn $n \rightarrow \infty$.

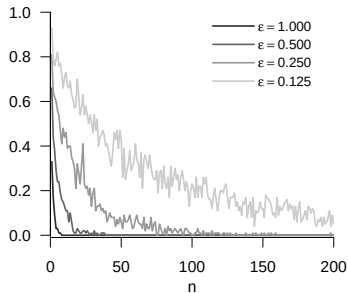
Beispiel ($X_1, \dots, X_n \sim N(0, 1)$)

- Die linke Abbildung zeigt Realisationen von $\bar{X}_n$ als Funktion von n .
- Die rechte Abbildung zeigt Schätzungen von $\mathbb{P}(|\bar{X}_n - \mu| \geq \epsilon)$ als Funktionen von n und ϵ .

Stichprobenmittelrealisationen



Wahrscheinlichkeitsschätzungen



Ungleichungen und Grenzwerte

- Wahrscheinlichkeitsungleichungen
- Erwartungswertungleichungen
- Gesetze der großen Zahl
- **Zentrale Grenzwertsätze**
- Selbstkontrollfragen

Überblick

- Die Zentralen Grenzwertsätze besagen grob, dass die Summe von unabhängigen Zufallsvariablen asymptotisch, d.h. für unendlich viele Zufallsvariablen, normalverteilt ist.
- Modelliert man eine Messgröße Y also als Summe eines deterministischen Einflusses μ und der Summe $\varepsilon := \sum_{i=1}^n X_i$ einer Vielzahl von unabhängigen Zufallsvariablen $X_i, i = 1, \dots, n$, welche unbekannte Störeinflüsse beschreiben, so ist für großes n die Annahme

$$Y = \mu + \varepsilon \text{ mit } \varepsilon \sim N(m, s^2) \quad (22)$$

also mathematisch gerechtfertigt. Wie wir später sehen werden, liegt die Annahme in Gleichung (22) vielen statischen Modellen zugrunde.

- In der “Lindenberg und Lévy” Form des Zentralen Grenzwertsatzes werden unabhängig und identische Zufallsvariablen vorausgesetzt. In der “Liapunov” Form werden nur unabhängige Zufallsvariablen vorausgesetzt. Der Beweis der “Lindenberg und Lévy” Form ist einfacher als der Beweis der “Liapunov” Form. Wir verzichten hier auf die Angabe von Beweisen.
- In beiden Formulierungen des Zentralen Grenzwertsatzes die betrachtete Konvergenz von Zufallsvariablen die *Konvergenz in Verteilung*, welche wir zunächst einführen.

Definition (Konvergenz in Verteilung)

Eine Folge $X_1, X_2, \dots$ von Zufallsvariablen *konvergiert in Verteilung gegen eine Zufallsvariable* X , wenn

$$\lim_{n \rightarrow \infty} P_{X_n}(x) = P_X(x). \quad (23)$$

für alle x an denen P_X stetig ist. Die Konvergenz in Verteilung von $X_1, X_2, \dots$ gegen X wird geschrieben als

$$X_n \xrightarrow[n \rightarrow \infty]{D} X, \quad (24)$$

Gilt $X_n \xrightarrow[n \rightarrow \infty]{D} X$, dann heißt die Verteilung von X die *asymptotische Verteilung der Folge* $X_1, X_2, \dots$.

Bemerkungen

- $X \xrightarrow[n \rightarrow \infty]{D} X$ ist eine Aussage über die Konvergenz von KVF's.
- Konvergenz in Wahrscheinlichkeit impliziert Konvergenz in Verteilung.

Theorem (Zentraler Grenzwertsatz nach Lindenberg und Lévy)

$X_1, \dots, X_n$ seien unabhängig und identisch verteilte Zufallsvariablen mit Erwartungswert $\mu := \mathbb{E}(X_i)$ und Varianz $\sigma^2 := \mathbb{V}(X_i)$, $0 < \sigma^2 < \infty$ für alle $i = 1, \dots, n$. Dann gilt für jedes feste x , dass

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq x \right) = \Phi(x), \quad (25)$$

wobei Φ KVF der Standardnormalverteilung bezeichnet.

Bemerkung

- Wir beweisen später noch folgende Implikationen des Zentralen Grenzwertsatzes:
 - Für $n \rightarrow \infty$ gilt asymptotisch $\sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2)$.
 - Für $n \rightarrow \infty$ gilt asymptotisch $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

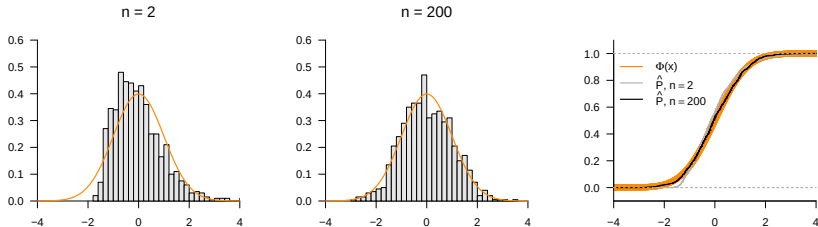
Beispiel ($X_1, \dots, X_n \sim \chi^2(k)$)

- Wir halten ohne Beweis fest, dass $\mathbb{E}(X_i) = k$ und $\mathbb{V}(X_i) = 2k$.
- Wir betrachten das Szenario $X_i \sim \chi(3)$ für $i = 1, \dots, n$.
- Die linken Abbildungen zeigen Histogrammschätzer der Wahrscheinlichkeitsdichte von

$$Y_n := \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \quad (26)$$

basierend auf 1000 Realisationen von Y_n für $n = 2$ und $n = 200$, sowie die WDF von $N(0, 1)$.

- Die rechte Abbildung zeigt die entsprechenden (empirischen) kumulativen Verteilungsfunktionen.



Theorem (Zentraler Grenzwertsatz nach Liapounov)

$X_1, \dots, X_n$ seien unabhängige aber nicht notwendigerweise identisch verteilten Zufallsvariablen mit den Eigenschaften

$$\mathbb{E}(|X_i - \mu_i|^3) < \infty \text{ und } \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \mathbb{E}(|X_i - \mu_i|^3)}{(\sum_{i=1}^n \sigma_i^2)^{3/2}} = 0. \quad (27)$$

Es sei weiterhin $\mu_i := \mathbb{E}(X_i)$ und $\sigma_i^2 = \mathbb{V}(X_i)$ für $i = 1, \dots, n$. Dann gilt für jedes feste x , dass

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \leq x \right) = \Phi(x), \quad (28)$$

wobei Φ KVF der Standardnormalverteilung bezeichnet.

Bemerkungen

- Für $n \rightarrow \infty$ gilt also asymptotisch $\sum_{i=1}^n X_i \sim N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$.
- Die Summe vieler unabhängiger Zufallsvariablen ist asymptotisch normalverteilt.

Ungleichungen und Grenzwerte

- Wahrscheinlichkeitsungleichungen
- Erwartungswertungleichungen
- Gesetze der großen Zahl
- Zentrale Grenzwertsätze
- **Selbstkontrollfragen**

1. Geben Sie die Markov Ungleichung wieder.
2. Geben Sie die Chebyshev Ungleichung wieder.
3. Geben Sie die Cauchy-Schwarz Ungleichung wieder.
4. Geben Sie die Korrelationsungleichung wieder.
5. Definieren Sie den Begriff der Konvergenz in Wahrscheinlichkeit.
6. Definieren Sie den Begriff der Konvergenz in Verteilung.
7. Geben Sie das Schwache Gesetz der großen Zahl wieder.
8. Erläutern Sie den Zentralen Grenzwertsatz nach Lindenbergl und Lévy.
9. Erläutern Sie den Zentralen Grenzwertsatz nach Liapunov.
10. Warum sind die Zentralen Grenzwertsätze für die statistische Modellbildung wichtig?

Literatur

Casella, G. and Berger, R. (2012). *Statistical Inference*. Duxbury.

DeGroot, M. H. and Schervish, M. J. (2012). *Probability and Statistics*. Addison-Wesley, Boston, 4th ed edition.

Georgii, H.-O. (2009). *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. DeGruyter-Lehrbuch. de Gruyter, Berlin, 4., überarb. und erw. aufl edition.

Wasserman, L. (2004). *All of Statistics*.



Inferenzstatistik

BSc Psychologie SoSe 2021

Prof. Dr. Dirk Ostwald

(6) Transformationen der Normalverteilung

Weiterführende Diskussionen der in dieser Vorlesungseinheit behandelten Inhalte finden sich z.B. in Georgii (2009, Kapitel 9), DeGroot and Schervish (2012, Sections 3.8 - 3.9) und Casella and Berger (2012, Section 5.3). Der Abschnitt Transformati-onstheoreme in dieser Vorlesungseinheit richtet sich primär Studierende, die an der Arbeit mit probabilistischen Modellen in der Statistik, dem maschinellen Lernen, oder der künstlichen Intelligenz interessiert sind. Bei den Standardtransformationen sind wie üblich die Theoreme, nicht jedoch die Beweise, prüfungsrelevant.

Transformationen der Normalverteilung

- Vorbemerkungen
- Transformationstheoreme
- Standardtransformationen
 - Summentransformation
 - Mittelwerttransformation
 - Z -Transformation
 - χ^2 -Transformation
 - T -Transformation
 - F -Transformation
- Selbstkontrollfragen

Transformationen der Normalverteilung

- **Vorbemerkungen**
- Transformationstheoreme
- Standardtransformationen
 - Summentransformation
 - Mittelwerttransformation
 - Z -Transformation
 - χ^2 -Transformation
 - T -Transformation
 - F -Transformation
- Selbstkontrollfragen

Realisierungen von Zufallsvariablen

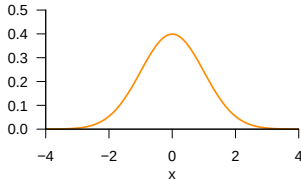
- Der einzelne Wert, den eine Zufallsvariable bei der Durchführung eines Zufallsexperiments annimmt, heißt eine *Realisierung der Zufallsvariable*. Mithilfe eines Computers lassen sich Zufallsexperimente simulieren und Realisierungen von Zufallsvariablen erhalten.
- Realisierungen von normalverteilten Zufallsvariablen erhält man in R mit **rnorm()**, wobei die Syntax für Realisierungen von n unabhängig und identisch verteilten Zufallsvariablen $X_i \sim N(\mu, \sigma^2)$, $i = 1, \dots, n$ durch **rnorm($n, \mu, \sqrt{\sigma^2}$)** gegeben ist.

```
x = rnorm(1,0,1)           # X1 \sim N(0,1)
[1] 0.13
x = rnorm(1,10,1)          # X1 \sim N(10,1)
[1] 8.87
x = rnorm(3,5,sqrt(2))      # Xi \sim N(5,2), i = 1,2,3 (u.i.v.)
[1] 5.25 3.17 6.62
x = rnorm(1e1,5,sqrt(2))    # Xi \sim N(5,2), i = 1,...,10 (u.i.v.)
[1] 2.63 8.48 5.60 7.88 6.04 4.08 5.34 3.22 6.42 5.30
x = rnorm(1e4,5,sqrt(2))    # Xi \sim N(5,2), i = 1,...,10.000 (u.i.v.)
[1] 6.78 5.43 5.10 2.89 4.28 4.22 4.31 1.94 5.65 6.26 4.49 4.82 4.93 5.66 ...
```

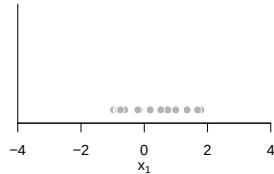
- Die empirische Verteilung unabhängig und identisch simulierter Zufallsvariablenrealisationen entspricht der Verteilung der Zufallsvariable. Die empirische Verteilung stellt man mit Histogrammen (Häufigkeitsverteilungen) oder histogramm-basierten Dichteschätzern dar.

Realisierungen von Zufallsvariablen

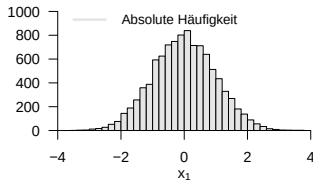
$N(x; 0, 1)$



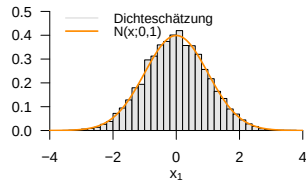
Realisierungen von $X_1, \dots, X_{20}$



$X_i \sim N(0,1), i = 1, \dots, 10^4$



$X_i \sim N(0,1), i = 1, \dots, 10^4$



Transformation von Zufallsvariablen

- Inhalt dieser Vorlesungseinheit sind einige Gesetzmäßigkeiten zur Transformation von normalverteilten Zufallsvariablen. Mit *Transformation* ist hier die Anwendung einer Funktion auf Zufallsvariablen sowie die arithmetische Verknüpfung mehrerer Zufallsvariablen gemeint. Die zentrale Fragestellung dabei ist folgende: "Wenn die Zufallsvariable X normalverteilt ist, wie ist dann eine Zufallsvariable Y , die sich durch Transformation von X ergibt, verteilt?"
- Für die in dieser Vorlesungseinheit behandelten Fälle gilt, dass man explizit Wahrscheinlichkeitsdichtefunktionen für die Verteilung der transformierten Zufallsvariable angeben kann. Diese gehören zu den klassischen Resultaten der frequentistischen Statistik und sind für das Verständnis von Konfidenzintervallen, Hypothesentests, und Varianzanalysen essentiell.
- Intuitiv kann man sich die Transformation einer Zufallsvariable anhand der Transformation ihrer u.i.v. Realisierungen klar machen. Betrachtet man z.B. $X \sim N(0, 1)$ und ihre Transformation $Y := X^2$ und sind $x_1 = 0.10, x_2 = -0.20, x_3 = 0.80$ drei u.i.v. Realisierungen von X , so entspricht dies den u.i.v. Realisierungen $y_1 = x_1^2 = 0.01, y_2 = x_2^2 = 0.04, y_3 = x_3^2 = 0.64$ von Y . In diesem Beispiel fällt auf, dass Y keine negativen Werte annimmt, die Verteilung von Y ordnet negativen Werten daher Wahrscheinlichkeitsdichten von 0 zu.

Transformation von Zufallsvariablen

- Simulation der Transformation normalverteilter Zufallsvariablen in R

```
# Simulationsspezifikation
n      = 1e4                      # Anzahl von u.i.v Realisierungen (ZVen)
mu      = 1                      # Erwartungswertparameter von X
sigsqr  = 2                      # Varianzparameter von X

# Quadration einer Zufallsvariable
X       = rnorm(n, mu, sigsqr)    # Realisierungen  $x_i$ ,  $i = 1, \dots, n$  von X
hist(X)                                # Histogramm der Realisierungen von X
Y       = X^2                    # Realisierungen  $y_i = x_i^2$  von Y
hist(Y)                                # Histogramm der Realisierungen von Y

# Addition zweier normalverteilter Zufallsvariablen
X_1     = rnorm(n, mu, sigsqr)    # Realisierungen  $x_{1i}$ ,  $i = 1, \dots, n$  von X1
X_2     = rnorm(n, mu, sigsqr)    # Realisierungen  $x_{2i}$ ,  $i = 1, \dots, n$  von X2
Y       = X_1 + X_2              # Realisierungen  $y_i = x_{1i} + x_{2i}$  von Y
hist(Y)                                # Histogramm der Realisierungen von Y
```

Überblick

- Im Abschnitt **Transformationstheoreme** stellen wir zunächst einige generelle Werkzeuge zum Berechnen der WDFen von transformierten Zufallsvariablen bereit. Diese Werkzeuge sind von der allgemeinen Form "Wenn X eine Zufallsvariable mit WDF p_X und $Y := f(X)$ die durch f transformierte Zufallsvariable ist, dann gilt für die WDF von Y die folgende Formel: $p_Y := \{\text{Formel}\}$ ".
- Im Abschnitt **Standardtransformationen** diskutieren wir sechs Standardtransformationen normalverteilter Zufallsvariablen, die in der klassischen frequentistischen Statistik und damit im weiteren Verlauf des Kurses zentrale Rollen spielen. Diese Aussagen sind von der allgemeinen Form "Wenn $X_i, i = 1, \dots, n$ unabhängig und identisch normalverteilte Zufallsvariablen sind und $Y := f(X_1, \dots, X_n)$ eine Transformation dieser Zufallsvariablen ist, dann ist die WDF von Y durch die Formel $p_Y := \{\text{Formel}\}$ gegeben und man nennt die Verteilung von Y {Verteilungsname}".
- Die Aussagen im Abschnitt **Standardtransformationen** sind für die klassische frequentistische Statistik zentral, weil
 - (1) der zentrale Grenzwertsatz die Annahme additiv unabhängig normalverteilter Störvariablen, und damit normalverteilter Daten, rechtfertigt,
 - (2) wie wir in der nächsten Vorlesungseinheit sehen werden, es sich bei Schätzern und Statistiken um Transformationen von Zufallsvariablen handelt, und
 - (3) Konfidenzintervalle und Hypothesentests durch die Verteilungen ihrer jeweiligen Statistiken charakterisiert und gerechtfertigt sind.

Transformationen der Normalverteilung

- Vorbemerkungen
- **Transformationstheoreme**
- Standardtransformationen
 - Summentransformation
 - Mittelwerttransformation
 - Z -Transformation
 - χ^2 -Transformation
 - T -Transformation
 - F -Transformation
- Selbstkontrollfragen

Überblick

- Das **univariate WDF Transformationstheorem bei bijektiven Abbildungen** liefert eine Formel zur Berechnung der WDF p_Y von $Y := f(X)$, wenn X eine Zufallsvariable mit WDF p_X ist und f eine bijektive Funktion ist.
- Das **univariate WDF Transformationstheorem bei linear affinen Abbildungen** gibt eine Formel zur Berechnung der WDF p_Y von $Y := f(X)$ an, wenn X eine Zufallsvariable mit WDF p_X ist und f eine linear-affine Funktion ist.
- Das **univariate WDF Transformationstheorem bei stückweisen bijektiven Abbildungen** gibt eine Formel zur Berechnung der WDF p_Y von $Y := f(X)$ an, wenn X eine Zufallsvariable mit WDF p_X ist und f zumindest in Teilen bijektiv ist.
- Das **multivariate WDF Transformationstheorem bei bijektiven Abbildungen** liefert eine Formel zur Berechnung der WDF p_Y von $Y := f(X)$, wenn X ein Zufallsvektor mit WDF p_X ist und f eine bijektive multivariate vektorwertige Funktion ist.
- Das **Faltungstheorem** liefert eine Formel zur Berechnung der WDF p_Y von $Y := X_1 + X_2$, wenn X_1 und X_2 zwei Zufallsvariablen mit WDFen p_{X_1} und p_{X_2} sind.

Theorem (Univariate WDF Transformation bei bijektiven Abbildungen)

X sei eine Zufallsvariable mit WDF p_X für die $\mathbb{P}(]a, b[) = 1$ gilt, wobei a und/oder b entweder endlich oder unendlich seien. Weiterhin sei

$$Y := f(X) \quad (1)$$

wobei die univariate reellwertige Funktion $f :]a, b[\rightarrow \mathbb{R}$ differenzierbar und bijektiv auf $]a, b[$ sei. $f(]a, b[)$ sei das Bild von $]a, b[$ unter f . Schließlich sei $f^{-1}(y)$ der Wert der Umkehrfunktion von $f(x)$ für $y \in f(]a, b[)$ und $f'(x)$ sei die Ableitung von f an der Stelle x . Dann ist die WDF von Y gegeben durch

$$p_Y : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, y \mapsto p_Y(y) := \begin{cases} \frac{1}{|f'(f^{-1}(y))|} p_X(f^{-1}(y)) & \text{für } y \in f(]a, b[) \\ 0 & \text{für } y \in \mathbb{R} \setminus f(]a, b[). \end{cases} \quad (2)$$

Bemerkungen

- Linear-affine Abbildungen sind ein wichtiger Anwendungsfall.
- Die Z -Transformation ist ein wichtiger Anwendungsfall.

Transformationstheoreme

Beweis

Wir halten zunächst fest, dass weil f eine differenzierbare bijektive Funktion auf $]a, b[$ ist, f entweder strikt wachsend oder strikt fallend ist. Nehmen wir zunächst an, dass f auf $]a, b[$ strikt wachsend ist. Dann ist auch f^{-1} für alle $y \in f(]a, b[)$ wachsend, und es gilt

$$P_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(f(X) \leq y) = \mathbb{P}(f^{-1}(f(X)) \leq f^{-1}(y)) = \mathbb{P}(X \leq f^{-1}(y)) = P_X(f^{-1}(y)). \quad (3)$$

P_Y ist also differenzierbar an allen Stellen y , an denen sowohl f^{-1} als auch P_X differenzierbar sind. Mit der Kettenregel und dem Satz von der Umkehrabbildung $(f^{-1}(x))' = 1/f'(f^{-1}(x))$, folgt dann, dass die WDF p_Y sich ergibt wie folgt:

$$p_Y(y) = \frac{d}{dy} P_Y(y) = \frac{d}{dy} P_X(f^{-1}(y)) = p_X(f^{-1}(y)) \frac{d}{dy} f^{-1}(y) = \frac{1}{f'(f^{-1}(y))} p_X(f^{-1}(y)), \quad (4)$$

Weil f^{-1} strikt wachsend ist, ist $d/dy(f^{-1}(y))$ positiv und das Theorem trifft zu. Analog gilt, dass wenn f auf $]a, b[$ strikt fallend ist, dann ist auch f^{-1} für alle $y \in f(]a, b[)$ fallend und es gilt

$$P_Y(y) = \mathbb{P}(f(X) \leq y) = \mathbb{P}(f^{-1}(f(X)) \geq f^{-1}(y)) = \mathbb{P}(X \geq f^{-1}(y)) = 1 - P_X(f^{-1}(y)), \quad (5)$$

Mit der Kettenregel und dem Satz von der Umkehrabbildung folgt dann

$$p_Y(y) = \frac{d}{dy} (1 - P_Y(y)) = -\frac{d}{dy} P_X(f^{-1}(y)) = -p_X(f^{-1}(y)) \frac{d}{dy} f^{-1}(y) = -\frac{1}{f'(f^{-1}(y))} p_X(f^{-1}(y)). \quad (6)$$

Weil f^{-1} strikt fallend ist, ist $d/dy(f^{-1}(y))$ negativ, so dass $-d/dy(f^{-1}(y))$ gleich $|d/dy(f^{-1}(y))|$ ist und das Theorem trifft zu.

□

Theorem (Univariate WDF Transformation bei linear-affinen Abbildungen)

X sei eine Zufallsvariable mit WDF p_X und es sei

$$Y = f(X) \text{ mit } f(X) := aX + b \text{ für } a \neq 0. \quad (7)$$

Dann ist die WDF von Y gegeben durch

$$p_Y : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, y \mapsto p_Y(y) := \frac{1}{|a|} p_X \left(\frac{y - b}{a} \right). \quad (8)$$

Bemerkung

- Das Theorem folgt direkt WDF Transformationstheorem bei bijektiven Abbildungen.
- Die Z -Transformation ist ein wichtiger Anwendungsfall.

Beweis

Wir halten zunächst fest, dass

$$f^{-1} : \mathbb{R} \rightarrow \mathbb{R}, y \mapsto f^{-1}(y) = \frac{y - b}{a} \quad (9)$$

ist, weil dann $f \circ f^{-1} = \text{id}_{\mathbb{R}}$ gilt, wie man anhand von

$$f(f^{-1}(x)) = a \left(\frac{x - b}{a} \right) + b = x - b + b = x \text{ für alle } x \in \mathbb{R} \quad (10)$$

einsieht. Wir halten weiterhin fest, dass

$$f' : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f'(x) = \frac{d}{dx}(ax + b) = a. \quad (11)$$

Also folgt mit dem Theorem zur WDF Transformation bei bijektiven Abbildungen, dass

$$\begin{aligned} p_Y : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, y \mapsto p_Y(y) &= \frac{1}{|f'(f^{-1}(y))|} p_X(f^{-1}(y)) \\ &= \frac{1}{|a|} p_X\left(\frac{y - b}{a}\right). \end{aligned} \quad (12)$$

□

Theorem (U.v. WDF Transformation bei stückweise bijektiven Abbildungen)

X sei eine Zufallsvariable mit Ergebnisraum $\mathcal{X}$ und WDF p_X . Weiterhin sei

$$Y = f(X), \quad (13)$$

wobei f so beschaffen sei, dass der Ergebnisraum von X in eine endliche Anzahl von Mengen $\mathcal{X}_1, \dots, \mathcal{X}_k$ mit einer entsprechenden Anzahl von Mengen $\mathcal{Y}_1 := f(\mathcal{X}_1), \dots, \mathcal{Y}_k := f(\mathcal{X}_k)$ im Ergebnisraum $\mathcal{Y}$ von Y partitioniert werden kann (wobei nicht notwendigerweise $\mathcal{Y}_i \cap \mathcal{Y}_j = \emptyset, 1 \leq i, j \leq k$ gelten muss), so dass die Abbildung f für alle $\mathcal{X}_1, \dots, \mathcal{X}_k$ bijektiv ist (d.h. f ist eine *stückweise* bijektive Abbildung). Für $i = 1, \dots, k$ bezeichne f_i^{-1} die Umkehrfunktion von f auf $\mathcal{Y}_i$. Schließlich nehmen wir an, dass die Ableitungen f'_i für alle $i = 1, \dots, k$ existieren und stetig sind. Dann ist eine WDF von Y durch

$$p_Y : \mathcal{Y} \rightarrow \mathbb{R}_{\geq 0}, y \mapsto p_Y(y) := \sum_{i=1}^k 1_{\mathcal{Y}_i}(y) \frac{1}{|f'_i(f_i^{-1}(y))|} p_X(f_i^{-1}(y)). \quad (14)$$

gegeben.

Bemerkungen

- Wir verzichten auf einen Beweis.
- Die χ^2 -Transformation ist ein wichtiger Anwendungsfall.

Theorem (Multivariate WDF Transformation bei bijektiven Abbildungen)

X sei ein n -dimensionaler Zufallsvektor mit Ergebnisraum $\mathbb{R}^n$ und WDF p_X . Weiterhin sei

$$Y := f(X), \quad (15)$$

wobei die multivariate vektorwertige Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ differenzierbar und bijektiv auf $]a, b[$ sei. Schließlich seien

$$J^f(x) = \left(\frac{\partial}{\partial x_j} f_i(x) \right)_{1 \leq i \leq n, 1 \leq j \leq n} \in \mathbb{R}^{n \times n} \quad (16)$$

die Jacobi-Matrix von f an der Stelle $x \in \mathbb{R}^n$, $|J^f(x)|$ die Determinante von $J^f(x)$, und es sei $|J^f(x)| \neq 0$ für alle $x \in \mathbb{R}^n$. Dann ist eine WDF von Y durch

$$p_Y : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}, y \mapsto p_Y(y) := \begin{cases} \frac{1}{|J^f(f^{-1}(y))|} p_X(f^{-1}(y)) & \text{for } y \in f(\mathbb{R}^n) \\ 0 & \text{for } y \in \mathbb{R}^n \setminus f(\mathbb{R}^n) \end{cases} \quad (17)$$

gegeben.

Bemerkungen

- Wir verzichten auf einen Beweis.
- Es handelt sich um eine direkte Generalisierung des univariaten Falls.
- Die T - und F -Transformationen sind wichtige Anwendungsfälle.

Theorem (Summe unabhängiger Zufallsvariable, Faltung)

X_1 und X_2 seien zwei kontinuierliche unabhängige Zufallsvariablen mit WDF p_{X_1} und p_{X_2} , respektive. $Y := X_1 + X_2$ sei die Summe von X_1 und X_2 . Dann ergibt sich eine WDF der Verteilung von Y als

$$p_Y(y) = \int_{-\infty}^{\infty} p_{X_1}(y - x_2)p_{X_2}(x_2) dx_2 = \int_{-\infty}^{\infty} p_{X_1}(x_1)p_{X_2}(y - x_1) dx_1 \quad (18)$$

Die Formel für die WDF p_Y heißt *Faltung* oder *Konvolution* von p_{X_1} und p_{X_2} .

Bemerkung

- Die Summen- und Mittelwerttransformation sind wichtige Anwendungsfälle.

Beweis

Wir nutzen das multivariate WDF Transformationstheorem für bijektive Abbildungen. Dazu definieren wir zunächst

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}^2, x \mapsto f(x) := \begin{pmatrix} x_1 + x_2 \\ x_2 \end{pmatrix} := \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \quad (19)$$

Die inverse Funktion von f ist dann gegeben durch

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}^2, z \mapsto f(z) := \begin{pmatrix} z_1 - z_2 \\ z_2 \end{pmatrix} \quad (20)$$

weil dann $f \circ f^{-1} = \text{id}_{\mathbb{R}^2}$ gilt, wie man anhand von

$$f^{-1}(f(x)) = f^{-1} \begin{pmatrix} x_1 + x_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 - x_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad (21)$$

einsieht. Die Jacobimatrix von f ergibt sich zu

$$J^f(x) = \begin{pmatrix} \frac{\partial}{\partial x_1} f_1(x) & \frac{\partial}{\partial x_2} f_1(x) \\ \frac{\partial}{\partial x_1} f_2(x) & \frac{\partial}{\partial x_2} f_2(x) \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x_1} (x_1 + x_2) & \frac{\partial}{\partial x_2} (x_1 + x_2) \\ \frac{\partial}{\partial x_1} x_2 & \frac{\partial}{\partial x_2} x_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad (22)$$

und die Jacobideterminante damit zu $|J^f(x)| = 1$.

Beweis (fortgeführt)

Wir halten weiterhin fest, dass die Unabhängigkeit von X_1 und X_2 impliziert, dass

$$p_{X_1, X_2}(x_1, x_2) = p_{X_1}(x_1)p_{X_2}(x_2) \quad (23)$$

impliziert. Einsetzen und Integration hinsichtlich x_2 ergibt dann ergibt dann für $z \in f(\mathbb{R}^2)$

$$\begin{aligned} p_Z(z) &= \frac{1}{|Jf(f^{-1}(z))|} p_X(f^{-1}(z)) \\ &= \frac{1}{1} p_{X_1, X_2}(z_1 - x_2, x_2) \\ &= p_{X_1}(z_1 - x_2)p_{X_2}(x_2) \end{aligned} \quad (24)$$

Integration über x_2 ergibt dann eine WDF für die marginale Verteilung von Z_1

$$p_{Z_1}(z_1) = \int_{-\infty}^{\infty} p_{X_1}(z_1 - x_2)p_{X_2}(x_2) dx_2 \quad (25)$$

Mit $Z_1 = X_1 + X_2 = Y$ ergibt sich dann die erste Form des Faltungstheorems zu

$$p_Y(y) = \int_{-\infty}^{\infty} p_{X_1}(y - x_2)p_{X_2}(x_2) dx_2. \quad (26)$$

□

Transformationen der Normalverteilung

- Vorbemerkungen
- Transformationstheoreme
- **Standardtransformationen**
 - Summentransformation
 - Mittelwerttransformation
 - Z -Transformation
 - χ^2 -Transformation
 - T -Transformation
 - F -Transformation
- Selbstkontrollfragen

Überblick

- Das **Summentransformationstheorem** besagt, dass die Summe unabhängig normalverteilter Zufallsvariablen wiederum normalverteilt ist und gibt die Parameter dieser Verteilung an.
- Das **Mittelwerttransformationstheorem** besagt, dass das Stichprobenmittel unabhängig normalverteilter Zufallsvariablen wiederum normalverteilt ist und gibt die Parameter dieser Verteilung an.
- Das **Z-Transformationstheorem** besagt, dass Subtraktion des Erwartungswertparameters und gleichzeitige Division mit der Wurzel des Varianzparameters die Verteilung einer normalverteilten Zufallsvariable in eine Standardnormalverteilung transformiere.
- Das **χ^2 -Transformationstheorem** besagt, dass die Summe quadrierter u.i.v. standardnormalverteilter Zufallsvariablen eine χ^2 -verteilte Zufallsvariable ist.
- Das **T-Transformationstheorem** besagt, dass die Zufallsvariable, die sich durch Division einer standardnormalverteilten Zufallsvariable durch die Quadratwurzel einer χ^2 -verteilten Zufallsvariable geteilt durch ein n , ergibt, eine t -verteilte Zufallsvariable ist.
- Das **F-Transformationstheorem** besagt, dass die Zufallsvariable, die sich durch Division zweier χ^2 verteilter Zufallsvariablen, jeweils geteilt durch ihre jeweiligen Freiheitsgradparameter, ergibt eine F -verteilte Zufallsvariable ist.

Transformationen der Normalverteilung

- Vorbemerkungen
- Transformationstheoreme
- **Standardtransformationen**
 - Summentransformation
 - Mittelwerttransformation
 - Z -Transformation
 - χ^2 -Transformation
 - T -Transformation
 - F -Transformation
- Selbstkontrollfragen

Theorem (Summe unabhängiger normalverteilter Zufallsvariablen)

Für $i = 1, \dots, n$ seien $X_i \sim N(\mu_i, \sigma_i^2)$ unabhängige normalverteilte Zufallsvariablen. Dann gilt für die Summe $Y := \sum_{i=1}^n X_i$, dass

$$Y \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right) \quad (27)$$

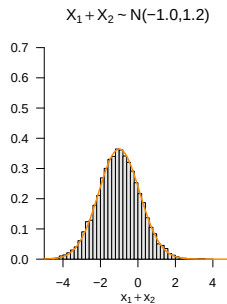
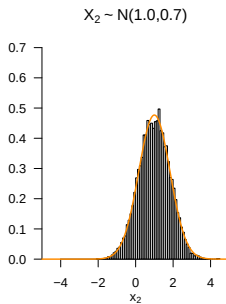
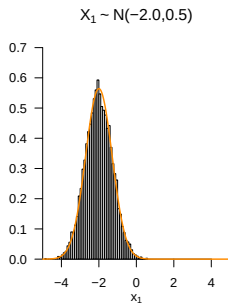
Für unabhängige und identisch normalverteilte Zufallsvariablen $X_i \sim N(\mu, \sigma^2)$ gilt folglich

$$Y \sim N(n\mu, n\sigma^2). \quad (28)$$

Bemerkung

- Die Mittelwerttransformation ist ein wichtiger Anwendungsfall.
- Die Generalisierung der zentralen Grenzwertsätze sind wichtige Anwendungsfälle.

Summe unabhängig normalverteilter Zufallsvariablen



Beweis

Wir skizzieren mithilfe der Faltungsformel, dass für $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, und $Y := X_1 + X_2$ gilt, dass $Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Für $n > 2$ folgt das Theorem dann durch Iteration. Mit der Definition der WDF der Normalverteilung erhalten wir zunächst

$$\begin{aligned} p_Y(y) &= \int_{-\infty}^{\infty} p_{X_1}(x_1) p_{X_2}(y - x_1) dx_1 \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2} \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2\right) \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2} \left(\frac{y - x_1 - \mu_2}{\sigma_2}\right)^2\right) dx_1 \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2} \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 - \frac{1}{2} \left(\frac{y - x_1 - \mu_2}{\sigma_2}\right)^2\right) dx_1. \end{aligned} \quad (29)$$

Mit einigem algebraischen Aufwand erhält man die Identität

$$\begin{aligned} -\frac{1}{2} \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 - \frac{1}{2} \left(\frac{y - x_1 - \mu_2}{\sigma_2}\right)^2 \\ = -\frac{(y - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)} - \frac{((\sigma_1^2 + \sigma_2^2)x_1 - \sigma_1^2 y + \mu_2 \sigma_1^2 - \mu_1 \sigma_2^2)^2}{2\sigma_1^2 \sigma_2^2 (\sigma_1^2 + \sigma_2^2)}, \end{aligned} \quad (30)$$

so dass weiterhin gilt, dass

Summentransformation

Beweis (fortgeführt)

$$\begin{aligned} p_Y(y) &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{(y - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)} - \frac{((\sigma_1^2 + \sigma_2^2)x_1 - \sigma_1^2 y + \mu_2\sigma_1^2 - \mu_1\sigma_2^2)^2}{2\sigma_1^2\sigma_2^2(\sigma_1^2 + \sigma_2^2)}\right) dx_1 \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{(y - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}\right) \exp\left(-\frac{((\sigma_1^2 + \sigma_2^2)x_1 - \sigma_1^2 y + \mu_2\sigma_1^2 - \mu_1\sigma_2^2)^2}{2\sigma_1^2\sigma_2^2(\sigma_1^2 + \sigma_2^2)}\right) dx_1 \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{(y - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{((\sigma_1^2 + \sigma_2^2)x_1 - \sigma_1^2 y + \mu_2\sigma_1^2 - \mu_1\sigma_2^2)^2}{2\sigma_1^2\sigma_2^2(\sigma_1^2 + \sigma_2^2)}\right) dx_1. \end{aligned} \quad (31)$$

Für das verbleibende Integral zeigt man mithilfe der Integration durch Substitution, dass

$$\int_{-\infty}^{\infty} \exp\left(-\frac{((\sigma_1^2 + \sigma_2^2)x_1 - \sigma_1^2 y + \mu_2\sigma_1^2 - \mu_1\sigma_2^2)^2}{2\sigma_1^2\sigma_2^2(\sigma_1^2 + \sigma_2^2)}\right) dx_1 = \frac{\sqrt{2\pi}\sigma_1\sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}. \quad (32)$$

Es ergibt sich also

$$\begin{aligned} p_Y(y) &= \frac{1}{2\pi\sigma_1\sigma_2} \frac{\sqrt{2\pi}\sigma_1\sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \exp\left(-\frac{(y - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}\right) \\ &= \frac{(2\pi)^{-1}(2\pi)^2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \exp\left(-\frac{(y - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}\right) \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2 + \sigma_2^2}} \exp\left(-\frac{(y - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}\right). \end{aligned} \quad (33)$$

Beweis (fortgeführt)

Schließlich folgt, dass

$$\begin{aligned} p_Y(y) &= \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp\left(-\frac{1}{2(\sigma_1^2 + \sigma_2^2)} (y - (\mu_1 + \mu_2))^2\right) \\ &= N(y; \mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) \end{aligned} \tag{34}$$

Ein einfacheres Vorgehen ergibt sich vermutlich nach Fouriertransformation der WDF im Sinne der sogenannten charakteristischen Funktion einer Zufallsvariable. In diesem Fall würde die Faltung der WDFen der Multiplikation der charakteristischen Funktionen entsprechen.

□

Transformationen der Normalverteilung

- Vorbemerkungen
- Transformationstheoreme
- **Standardtransformationen**
 - Summentransformation
 - **Mittelwerttransformation**
 - Z -Transformation
 - χ^2 -Transformation
 - T -Transformation
 - F -Transformation
- Selbstkontrollfragen

Theorem (Stichprobenmittel von u.i.v. normalverteilten Zufallsvariablen)

Für $i = 1, \dots, n$ seien $X_i \sim N(\mu, \sigma^2)$ unabhängig und identisch normalverteilte Zufallsvariablen. Dann gilt für **das Stichprobenmittel** $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$, dass

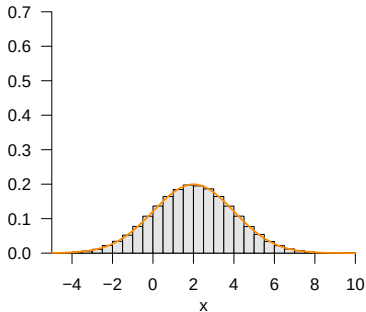
$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right). \quad (35)$$

Bemerkung

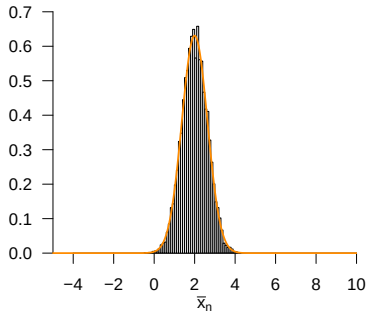
- Die Analyse von Erwartungswertschätzern ist ein wichtiger Anwendungsfall.
- Die Generalisierung der zentralen Grenzwertsätze sind wichtige Anwendungsfälle.

Stichprobenmittel u.i.v. normalverteilter Zufallsvariablen

$$X_i \sim N(2,4), i = 1, \dots, 10$$



$$\bar{X}_n \sim N(2, 4/10)$$



Mittelwerttransformation

Beweis

Wir halten zunächst fest, dass mit dem Theorem zur Summe von unabhängig normalverteilten Zufallsvariablen gilt, dass $\bar{X}_n = \frac{1}{n} Y$ mit

$Y := \sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2)$. Einsetzen in das univariate WDF Transformationstheorem für lineare Funktionen ergibt dann

$$\begin{aligned} p_{\bar{X}_n}(\bar{x}_n) &= \frac{1}{|1/n|} N\left(n\bar{x}_n; n\mu, n\sigma^2\right) \\ &= \frac{n}{\sqrt{2\pi n\sigma^2}} \exp\left(-\frac{1}{2n\sigma^2} (n\bar{x}_n - n\mu)^2\right) \\ &= \frac{n}{\sqrt{2\pi n\sigma^2}} \exp\left(-\frac{1}{2n\sigma^2} (n\bar{x}_n - n\mu)^2\right) \\ &= n n^{-\frac{1}{2}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(n\bar{x}_n)^2}{2n\sigma^2} + \frac{2(n\bar{x}_n)(n\mu)}{2n\sigma^2} - \frac{(n\mu)^2}{2n\sigma^2}\right) \\ &= \sqrt{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{n\bar{x}_n^2}{2\sigma^2} + \frac{2n\bar{x}_n\mu}{2\sigma^2} - \frac{n\mu^2}{2\sigma^2}\right) \\ &= \frac{1}{1/\sqrt{n}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\bar{x}_n^2}{2(\sigma^2/n)} + \frac{2\bar{x}_n\mu}{2(\sigma^2/n)} - \frac{\mu^2}{2(\sigma^2/n)}\right) \\ &= \frac{1}{\sqrt{2\pi(\sigma^2/n)}} \exp\left(-\frac{1}{2(\sigma^2/n)} (\bar{x}_n - \mu)^2\right) \\ &= N\left(\bar{x}_n; \mu, \sigma^2/n\right) \end{aligned} \tag{36}$$

also, dass $\bar{X}_n \sim N\left(\mu, \sigma^2/n\right)$.

□

Transformationen der Normalverteilung

- Vorbemerkungen
- Transformationstheoreme
- **Standardtransformationen**
 - Summentransformation
 - Mittelwerttransformation
 - ***Z*-Transformation**
 - χ^2 -Transformation
 - *T*-Transformation
 - *F*-Transformation
- Selbstkontrollfragen

Definition (Z -Zufallsvariable)

Z sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}$ und WDF

$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, z \mapsto p(z) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right). \quad (37)$$

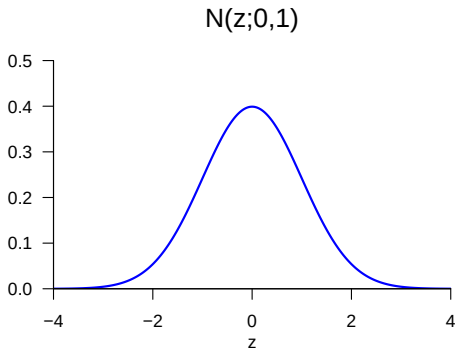
Dann sagen wir, dass Z einer Z -Verteilung (oder *Standardnormalverteilung*) unterliegt und nennen Z eine Z -Zufallsvariable. Wir kürzen dies mit $Z \sim N(0, 1)$ ab. Die WDF einer Z -Zufallsvariable bezeichnen wir mit

$$N(z; 0, 1) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right). \quad (38)$$

Bemerkung

- Eine Z -Zufallsvariable ist eine normalverteilte Zufallsvariable mit $\mu := 0$ und $\sigma^2 := 1$.

Wahrscheinlichkeitsdichtefunktion der Z-Verteilung



- Die Z -Verteilung hat keine Parameter.

Theorem (Z-Transformation)

Es sei $X \sim N(\mu, \sigma^2)$ eine normalverteilte Zufallsvariable. Dann ist die Zufallsvariable

$$Z := \frac{X - \mu}{\sigma} \quad (39)$$

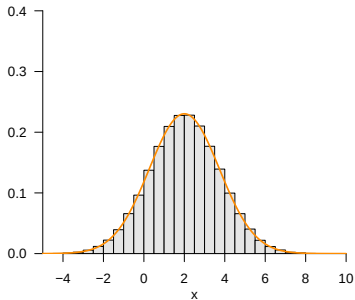
eine Z-verteilte Zufallsvariable, es gilt also $Z \sim N(0, 1)$.

Bemerkungen

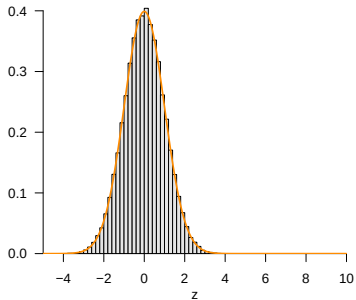
- Z wird hier als $(X - \mu)/\sigma$ definiert. Dass ein solches Z aber eine Z-Zufallsvariable ist, muss bewiesen werden und ergibt sich nicht einfach durch die Wahl des Bezeichners für $(X - \mu)/\sigma$, welcher hier zufällig auch Z lautet. In analoger Form gilt diese Bemerkung auch für alle weiteren betrachteten Transformationen.
- Die T-Statistik und T-Tests sind wichtige Anwendungsfälle.

Z-Transformation

$$X \sim N(2,3)$$



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



Beweis

Wir nutzen das univariate WDF Transformationstheorem für **linear-affine** Funktionen. Dazu halten wir zunächst fest, dass die Z-Transformation einer Funktion der Form

$$\zeta : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \zeta(x) := \frac{x - \mu}{\sigma} =: z \quad (40)$$

entspricht. Wir stellen weiterhin fest, dass die Umkehrfunktion von ζ durch

$$\zeta^{-1} : \mathbb{R} \rightarrow \mathbb{R}, z \mapsto \zeta^{-1}(z) := \sigma z + \mu \quad (41)$$

gegeben ist, da für alle $z \in \mathbb{R}$ mit $z = \frac{x - \mu}{\sigma}$ gilt, dass

$$\zeta^{-1}(z) = \zeta^{-1}\left(\frac{x - \mu}{\sigma}\right) = \frac{\sigma(x - \mu)}{\sigma} + \mu = x - \mu + \mu = x. \quad (42)$$

Schließlich stellen wir fest, dass für die Ableitung ζ' von ζ gilt, dass

$$\zeta'(x) = \frac{d}{dx} \left(\frac{x - \mu}{\sigma} \right) = \frac{d}{dx} \left(\frac{x}{\sigma} - \frac{\mu}{\sigma} \right) = \frac{1}{\sigma}. \quad (43)$$

Beweis (fortgeführt)

Einsetzen in das univariate WDF Transformationstheorem für lineare Funktionen ergibt dann

$$\begin{aligned} p_Z(z) &= \frac{1}{|1/\sigma|} N(\sigma z + \mu; \mu, \sigma^2) \\ &= \frac{1}{1/\sqrt{\sigma^2}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (\sigma z + \mu - \mu)^2\right) \\ &= \frac{\sqrt{\sigma^2}}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \sigma^2 z^2\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2\right) \\ &= N(z; 0, 1) \end{aligned} \tag{44}$$

also, dass $Z \sim N(0, 1)$. Z ist also eine Z -Zufallsvariable.

□

Transformationen der Normalverteilung

- Vorbemerkungen
- Transformationstheoreme
- **Standardtransformationen**
 - Summentransformation
 - Mittelwerttransformation
 - Z -Transformation
 - χ^2 -Transformation
 - T -Transformation
 - F -Transformation
- Selbstkontrollfragen

Definition (χ^2 -Zufallsvariable)

U sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}_{>0}$ und WDF

$$p : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}, u \mapsto p(u) := \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} u^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}u\right), \quad (45)$$

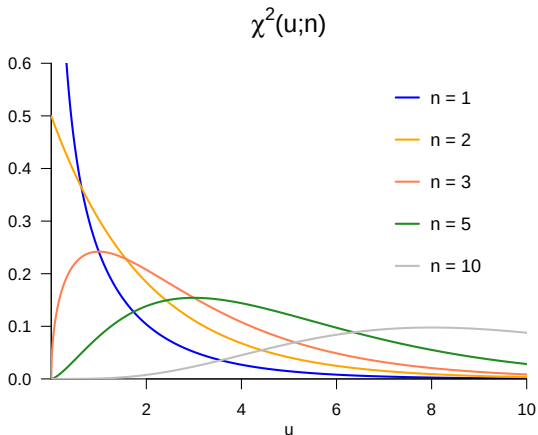
wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass U einer χ^2 -Verteilung mit n Freiheitsgraden unterliegt und nennen U eine χ^2 -Zufallsvariable mit n Freiheitsgraden. Wir kürzen dies mit $U \sim \chi^2(n)$ ab. Die WDF einer χ^2 -Zufallsvariable bezeichnen wir mit

$$\chi^2(u; n) := \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} u^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}u\right). \quad (46)$$

Bemerkung

- Die WDF der χ^2 -Verteilung entspricht der WDF $G\left(u; \frac{n}{2}, 2\right)$ einer Gammaverteilung.

Wahrscheinlichkeitsdichtefunktionen der χ^2 -Verteilung



- Steigendes n verbreitert $\chi^2(u; n)$ und verschiebt Masse zur größeren Werten.

Theorem (χ^2 -Transformation)

$Z_1, \dots, Z_n \sim N(0, 1)$ seien unabhängig und identisch verteilte Z -Zufallsvariablen. Dann ist die Zufallsvariable

$$U := \sum_{i=1}^n Z_i^2 \quad (47)$$

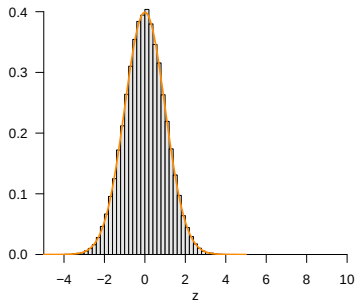
eine χ^2 -verteilte Zufallsvariable mit n Freiheitsgraden, es gilt also $U \sim \chi^2(n)$. Insbesondere gilt für $Z \sim N(0, 1)$ und $U := Z^2$, dass $U \sim \chi^2(1)$.

Bemerkungen

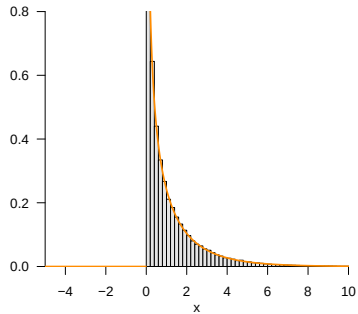
- Die T-Statistik und T-Tests sind wichtige Anwendungsfälle.
- Die F-Statistik und Varianzanalysen sind wichtige Anwendungsfälle.

χ^2 -Transformation

$$Z \sim N(0,1)$$



$$U = Z^2 \sim \chi^2(1)$$



Beweis

Wir zeigen das Theorem nur für den Fall $n := 1$ mithilfe des WDF Transformationstheorems für stückweise bijektive Abbildungen. Danach ist die WDF einer Zufallsvariable $U := f(Z)$, welche aus der Transformation einer Zufallsvariable Z mit WDF p_Z durch eine stückweise bijektive Abbildung hervorgeht, gegeben durch

$$p_U(u) = \sum_{i=1}^k 1_{\mathcal{U}_i} \frac{1}{|f'_i(f_i^{-1}(u))|} p_Z\left(f_i^{-1}(u)\right). \quad (48)$$

Wir definieren

$$\mathcal{U}_1 :=]-\infty, 0[, \mathcal{U}_2 :=]0, \infty[, \text{ und } \mathcal{U}_i := \mathbb{R}_{>0} \text{ für } i = 1, 2, \quad (49)$$

sowie

$$f_i : \mathcal{Z}_i \rightarrow \mathcal{U}_i, x \mapsto f_i(z) := z^2 =: u \text{ für } i = 1, 2. \quad (50)$$

Die Ableitung und die Umkehrfunktion der f_i ergeben sich zu

$$f'_i : \mathcal{Z}_i \rightarrow \mathcal{Z}_i, x \mapsto f'_i(z) = 2z \text{ für } i = 1, 2, \quad (51)$$

und

$$f_1^{-1} : \mathcal{U}_1 \rightarrow \mathcal{U}_1, u \mapsto f_1^{-1}(u) = -\sqrt{u} \text{ und } f_2^{-1} : \mathcal{U}_2 \rightarrow \mathcal{U}_2, u \mapsto f_2^{-1}(u) = \sqrt{u}, \quad (52)$$

respektive.

Beweis (fortgeführt)

Einsetzen in Gleichung (48) ergibt dann

$$\begin{aligned} p_U(u) &= 1_{\mathcal{U}_1}(u) \frac{1}{|f'_1(f_1^{-1}(u))|} p_Z\left(f_1^{-1}(u)\right) + 1_{\mathcal{U}_2}(u) \frac{1}{|f'_2(f_2^{-1}(u))|} p_Z\left(f_2^{-1}(u)\right) \\ &= \frac{1}{|2(-\sqrt{u})|} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(-\sqrt{u})^2\right) + \frac{1}{|2(\sqrt{u})|} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\sqrt{u})^2\right) \\ &= \frac{1}{2\sqrt{u}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u\right) + \frac{1}{2\sqrt{u}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u\right) \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{u}} \exp\left(-\frac{1}{2}u\right). \end{aligned} \tag{53}$$

Andererseits gilt, dass mit $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, die PDF einer χ^2 -Zufallsvariable U mit $n = 1$ durch

$$\frac{1}{\Gamma\left(\frac{1}{2}\right) 2^{\frac{1}{2}}} u^{\frac{1}{2}-1} \exp\left(-\frac{1}{2}u\right) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{u}} \exp\left(-\frac{1}{2}u\right) \tag{54}$$

gegeben ist. Also gilt, dass wenn $Z \sim N(0, 1)$ ist, dann ist $U := Z^2 \sim \chi^2(1)$.

□

Transformationen der Normalverteilung

- Vorbemerkungen
- Transformationstheoreme
- **Standardtransformationen**
 - Summentransformation
 - Mittelwerttransformation
 - Z -Transformation
 - χ^2 -Transformation
 - **T -Transformation**
 - F -Transformation
- Selbstkontrollfragen

Definition (t -Zufallsvariable)

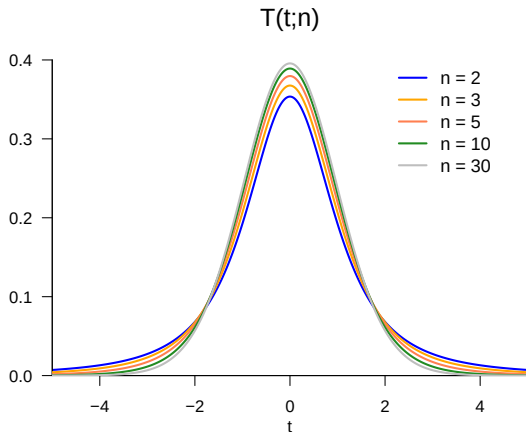
T sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}$ und WDF

$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, t \mapsto p(t) := \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad (55)$$

wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass T einer t -Verteilung mit n Freiheitsgraden unterliegt und nennen T eine t -Zufallsvariable mit n Freiheitsgraden. Wir kürzen dies mit $T \sim t(n)$ ab. Die WDF einer t -Zufallsvariable bezeichnen wir mit

$$T(t; n) := \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}. \quad (56)$$

Wahrscheinlichkeitsdichtefunktionen der t -Verteilung



- Die Verteilung ist um 0 symmetrisch
- Steigendes n verschiebt Wahrscheinlichkeitsmasse aus den Ausläufen zum Zentrum.
- Ab $n = 30$ gilt $T(t; n) \approx N(0, 1)$.

Theorem (T -Transformation)

$Z \sim N(0, 1)$ sei eine Z -Zufallsvariable, $U \sim \chi^2(n)$ sei eine χ^2 -Zufallsvariable mit n Freiheitsgraden, und Z und U seien unabhängige Zufallsvariablen. Dann ist die Zufallsvariable

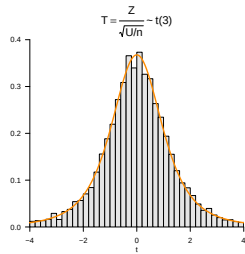
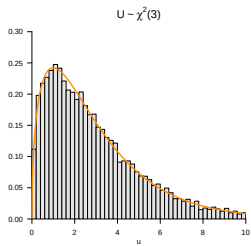
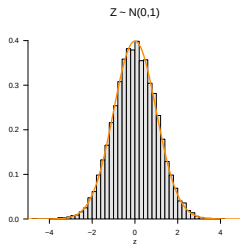
$$T := \frac{Z}{\sqrt{U/n}} \quad (57)$$

eine t -verteilte Zufallsvariable mit n Freiheitsgraden, es gilt also $T \sim t(n)$.

Bemerkungen

- Das Theorem geht auf (Student, 1908) zurück.
- Das Theorem ist ein zentrales Resultat der frequentistischen Statistik.
- ZABELL (2008) gibt hierzu einen historischen Überblick.
- Die T -Statistik und T -Tests sind wichtige Anwendungsfälle.

T-Transformation



Beweis

Wir halten zunächst fest, dass die zweidimensionale WDF der gemeinsamen (unabhängigen) Verteilung von Z und U durch

$$p_{Z,U}(z, u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} u^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}u\right). \quad (58)$$

gegeben ist. Wir betrachten dann die multivariate vektorwertige Abbildung

$$f: \mathbb{R}^2 \rightarrow \mathbb{R}^2, (z, u) \mapsto f(z, u) := \left(\frac{z}{\sqrt{u/n}}, u \right) =: (t, w) \quad (59)$$

und benutzen das multivariate WDF Transformationstheorem für bijektive Abbildungen um die WDF von (t, w) herzuleiten. Dazu erinnern wir uns, dass wenn X ein n -dimensionaler Zufallsvektor mit WDF p_X und $Y := f(X)$ für eine differenzierbare und bijektive Abbildung $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ ist, die WDF des Zufallsvektors Y durch

$$p_Y: \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}, y \mapsto p_Y(y) := \frac{1}{|Jf(f^{-1}(y))|} p_X(f^{-1}(y)) \quad (60)$$

gegeben ist. Für die im vorliegenden Fall betrachtete Abbildung halten wir zunächst fest, dass

$$f^{-1}: \mathbb{R}^2 \rightarrow \mathbb{R}^2, (t, w) \mapsto f^{-1}(t, w) := \left(\sqrt{w/nt}, w \right). \quad (61)$$

Beweis (fortgeführt)

Dies ergibt sich direkt aus

$$f^{-1}(f(z, u)) = f^{-1}\left(\frac{z}{\sqrt{u/n}}, u\right) = \left(\frac{\sqrt{u/n}z}{\sqrt{u/n}}, u\right) = (z, u) \text{ für alle } (z, u) \in \mathbb{R}^2. \quad (62)$$

Wir halten dann fest, dass die Determinante der Jacobi-Matrix von f an der Stelle (z, u) durch

$$|J^f(z, u)| = \left| \begin{array}{cc} \frac{\partial}{\partial z} \left(\frac{z}{\sqrt{u/n}} \right) & \frac{\partial}{\partial u} \left(\frac{z}{\sqrt{u/n}} \right) \\ \frac{\partial}{\partial z} u & \frac{\partial}{\partial u} u \end{array} \right| = \left(\frac{v}{n} \right)^{-1/2}, \quad (63)$$

gegeben ist, sodass folgt, dass

$$\frac{1}{|J^f(f^{-1}(z, u))|} = \left(\frac{w}{n} \right)^{1/2}. \quad (64)$$

Einsetzen in Gleichung (60) ergibt dann

$$p_{T,W}(t, w) = \left(\frac{w}{n} \right)^{1/2} p_{Z,V} \left(\sqrt{w/nt}, w \right), \quad (65)$$

Beweis (fortgeführt)

Es folgt also

$$\begin{aligned} p_T(t) &= \int_0^\infty p_{T,W}(t, w) dw \\ &= \int_0^\infty \left(\frac{w}{n}\right)^{1/2} p_{Z,V}\left(\sqrt{w/nt}, w\right) dw \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\sqrt{w/nt})^2\right) \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} w^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}w\right) \left(\frac{w}{n}\right)^{1/2} dw \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}} n^{\frac{1}{2}}} \int_0^\infty \exp\left(-\frac{1}{2}\frac{w}{n}t^2\right) w^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}w\right) w^{1/2} dw \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}} n^{\frac{1}{2}}} \int_0^\infty \exp\left(-\frac{1}{2}\frac{w}{n}t^2 - \frac{1}{2}w\right) w^{\frac{n}{2}-1} w^{\frac{1}{2}} dw \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}} n^{\frac{1}{2}}} \int_0^\infty \exp\left(-\frac{1}{2}\left(\frac{w}{n}t^2 + w\right)\right) w^{\frac{n+1}{2}-1} dw \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}} n^{\frac{1}{2}}} \int_0^\infty \exp\left(-\frac{1}{2}\left(1 + \frac{t^2}{n}\right)w\right) w^{\frac{n+1}{2}-1} dw \end{aligned} \tag{66}$$

Beweis (fortgeführt)

Wir stellen dann fest, dass der Integrand auf der linken Seite der obigen Gleichung dem Kern einer Gamma WDF mit Parametern $\alpha = \frac{n+1}{2}$ und $\beta = \frac{2}{1+\frac{t^2}{n}}$ entspricht, wie man leicht einsieht:

$$\begin{aligned}\Gamma(w; \alpha, \beta) &= \frac{1}{\Gamma(\alpha)\beta^\alpha} w^{\alpha-1} \exp\left(-\frac{w}{\beta}\right) \\ \Rightarrow \Gamma\left(w; \frac{n+1}{2}, \frac{2}{1+\frac{t^2}{n}}\right) &= \frac{1}{\Gamma\left(\frac{n+1}{2}\right) \left(\frac{2}{1+\frac{t^2}{n}}\right)^{\frac{n+1}{2}}} w^{\frac{n+1}{2}-1} \exp\left(-\frac{w}{\frac{2}{1+\frac{t^2}{n}}}\right) \\ &= \frac{1}{\Gamma\left(\frac{n+1}{2}\right) \left(\frac{2}{1+\frac{t^2}{n}}\right)^{\frac{n+1}{2}}} \exp\left(-\frac{1}{2} \left(1 + \frac{t^2}{n}\right)\right) w^{\frac{n+1}{2}-1}.\end{aligned}$$

Beweis (fortgeführt)

Es ergibt sich also

$$p_T(t) = \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}} n^{\frac{1}{2}}} \int_0^\infty \Gamma\left(w; \frac{n+1}{2}, \frac{2}{1 + \frac{t^2}{n}}\right) dw. \quad (67)$$

Schließlich stellen wir fest, dass der Integralterm in obiger Gleichung dem Normalisierungsterm einer Gamma WDF entspricht. Abschließend ergibt sich also

$$p_T(t) = \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}} n^{\frac{1}{2}}} \Gamma\left(\frac{n+1}{2}\right) \left(\frac{2}{1 + \frac{t^2}{n}}\right)^{\frac{n+1}{2}}. \quad (68)$$

Die Verteilung von $Z/\sqrt{U/n}$ hat also die WDF einer T -Zufallsvariable.



Transformationen der Normalverteilung

- Vorbemerkungen
- Transformationstheoreme
- **Standardtransformationen**
 - Summentransformation
 - Mittelwerttransformation
 - Z -Transformation
 - χ^2 -Transformation
 - T -Transformation
 - **F -Transformation**
- Selbstkontrollfragen

Definition (f -Zufallsvariable)

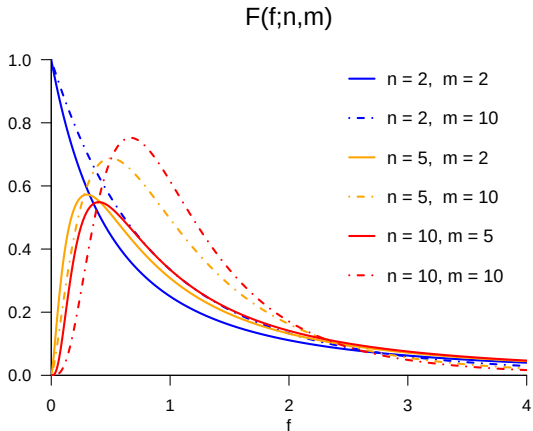
F sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}_{>0}$ und WDF

$$p_F : \mathbb{R} \rightarrow \mathbb{R}_{>0}, f \mapsto p_F(f) := m^{\frac{m}{2}} n^{\frac{n}{2}} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} \frac{f^{\frac{m}{2}-1}}{\left(1 + \frac{m}{n} f\right)^{\frac{m+n}{2}}}, \quad (69)$$

wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass F einer f -Verteilung mit n, m Freiheitsgraden unterliegt und nennen F eine f -Zufallsvariable mit n, m Freiheitsgraden. Wir kürzen dies mit $F \sim f(n, m)$ ab. Die WDF einer f -Zufallsvariable bezeichnen wir mit

$$F(f; n, m) := m^{\frac{m}{2}} n^{\frac{n}{2}} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} \frac{f^{\frac{m}{2}-1}}{\left(1 + \frac{m}{n} f\right)^{\frac{m+n}{2}}}. \quad (70)$$

Wahrscheinlichkeitsdichtefunktionen der f -Verteilung



Theorem (F -Transformation)

$V \sim \chi^2(n)$ und $W \sim \chi^2(m)$ seien zwei unabhängige χ^2 -Zufallsvariablen mit n und m Freiheitsgraden, respektive. Dann ist die Zufallsvariable

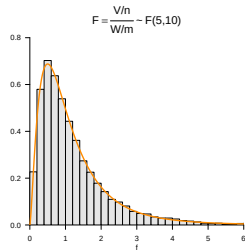
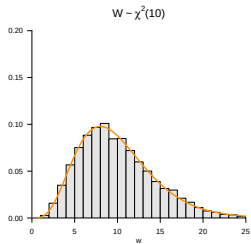
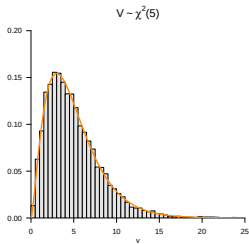
$$F := \frac{V/n}{W/m} \quad (71)$$

eine f -verteilte Zufallsvariable mit n, m Freiheitsgraden, es gilt also $F \sim f(n, m)$.

Bemerkungen

- Wir verzichten auf einen expliziten Beweis.
- Das Theorem kann bewiesen werden, in dem man zunächst ein Transformation für Quotienten von Zufallsvariablen mithilfe des multivariaten Transformationstheorems und Marginalisierung herleitet und dieses Theorem dann auf die WDF von χ^2 -verteilten ZVen anwendet. Dabei ist die Regel zur Integration durch Substitution von zentraler Bedeutung.

F-Transformation



Transformationen der Normalverteilung

- Vorbemerkungen
- Transformationstheoreme
- Standardtransformationen
 - Summentransformation
 - Mittelwerttransformation
 - Z -Transformation
 - χ^2 -Transformation
 - T -Transformation
 - F -Transformation
- **Selbstkontrollfragen**

1. Erläutern Sie den Begriff der Transformation einer Zufallsvariable.
2. Erläutern Sie die zentrale Idee der Transformationstheoreme.
3. Erläutern Sie die Bedeutung der Standardtransformationen für die Statistik.
4. Geben sie das Summentransformationstheorem wieder.
5. Geben sie das Mittelwerttransformationstheorem wieder.
6. Geben sie das Z -Transformationstheorem wieder.
7. Geben sie das χ^2 -Transformationstheorem wieder.
8. Beschreiben Sie die WDF der t -Verteilung in Abhängigkeit ihrer Freiheitsgrade.
9. Geben sie das T -Transformationstheorem wieder.
10. Geben sie das F -Transformationstheorem wieder.

Literatur

Casella, G. and Berger, R. (2012). *Statistical Inference*. Duxbury.

DeGroot, M. H. and Schervish, M. J. (2012). *Probability and Statistics*. Addison-Wesley, Boston, 4th ed edition.

Georgii, H.-O. (2009). *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. De-Gruyter-Lehrbuch. de Gruyter, Berlin, 4., überarb. und erw. aufl edition.

Student (1908). The Probable Error of a Mean. *Biometrika*, 6(1):1–25.

Zabell, S. L. (2008). On Student's 1908 Article "The Probable Error of a Mean". *Journal of the American Statistical Association*, 103(481):1–7.



Inferenzstatistik

BSc Psychologie SoSe 2021

Prof. Dr. Dirk Ostwald

(7) Statistische Modelle, Statistiken und Schätzer

Die in dieser Einheit behandelten Konzepte sind grundlegend und werden unter anderem in Georgii (2009, Kapitel 7) diskutiert.

Statistische Modelle, Statistiken und Schätzer

- Statistische Modelle
- Statistiken und Schätzer
- Standardprobleme frequentistischer Inferenz
- Maximum-Likelihood Schätzer
- Selbstkontrollfragen

Statistische Modelle, Statistiken und Schätzer

- **Statistische Modelle**
- Statistiken und Schätzer
- Standardprobleme frequentistischer Inferenz
- Maximum-Likelihood Schätzer
- Selbstkontrollfragen

Definition (Statistisches Modell)

Ein *statistisches Modell* ist ein Tripel

$$\mathcal{M} := (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\}) \quad (1)$$

bestehend aus einem *Stichprobenraum* $\mathcal{X}$, einer σ -Algebra $\mathcal{A}$ auf $\mathcal{X}$ und einer mindestens zweielementigen Menge $\{\mathbb{P}_\theta | \theta \in \Theta\}$ von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{A})$, die mit einer gewissen Indexmenge Θ indiziert sind.

Bemerkungen

- Im Gegensatz zum W-Raum betrachten wir zwei oder mehr W-Maße.
- Das jeweils zugrundeliegende Wahrscheinlichkeitsmaß ist mit $\theta \in \Theta$ indiziert.
- Für Erwartungswerte und (Ko)Varianzen bezüglich $\mathbb{P}_\theta$ schreiben wir $\mathbb{E}_\theta$, $\mathbb{V}_\theta$, $\mathbb{C}_\theta$.

Definition (Parametrische, Standard- und Produktmodelle)

- Ein statistisches Modell $\mathcal{M}$ heißt ein *parametrisches Modell*, wenn $\Theta \subset \mathbb{R}^d$ für ein $d \in \mathbb{N}$ ist. Für $d = 1$ heißt $\mathcal{M}$ ein *einparametrisches Modell*.
- Ein statistisches Modell $\mathcal{M}$ heißt ein *diskretes Modell*, wenn $\mathcal{X}$ diskret ist. Jedes $\mathbb{P}_\theta$ besitzt dann eine WMF p_θ definiert durch $p_\theta := \mathbb{P}_\theta(x)$. Ein statistisches Modell $\mathcal{M}$ heißt ein *stetiges Modell*, wenn $\mathcal{X} \subset \mathbb{R}^n$ ist und jedes $\mathbb{P}_\theta$ eine WDF p_θ besitzt.
- Für ein statistisches Modell $\mathcal{M}_0 := (\mathcal{X}_0, \mathcal{A}_0, \{\mathbb{P}_\theta^0 | \theta \in \Theta\})$ und $n > 1$ heißt das statistische Modell $\mathcal{M} := (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\})$, für welches $\mathcal{X} := \times_{i=1}^n \mathcal{X}_0$ das n -fache kartesische Produkt von $\mathcal{X}_0$ mit sich selbst ist, $\mathcal{A}$ die entsprechende Produkt- σ -Algebra ist, und $\{\mathbb{P}_\theta | \theta \in \Theta\}$ die entsprechende Menge an Produktmaßen $\prod_{i=1}^n \mathbb{P}_\theta^0$ ist, das zu $\mathcal{M}_0$ gehörige *Produktmodell*.

Bemerkungen

- Der Vorgang des Beobachtens, der ein zufälliges Ergebnis liefert, wird mit Zufallsvektoren X , die Werte in $\mathcal{X}$ annehmen, beschrieben.
- Im Kontext statistischer Modelle nennt man diese Zufallsvektoren *Beobachtung*, *Messung*, oder *Stichprobe*. Ihre Realisierungen $x \in \mathcal{X}$, also konkret vorliegende Datenwerte, werden *Beobachtungswert*, *Messwert*, *Stichprobenwert* oder *Datensatz* genannt.
- Produktmodelle modellieren die unabhängige Wiederholung eines Einzelexperiments. Der entsprechende Zufallsvektor $X := (X_1, \dots, X_n)$ entspricht dann einer Menge von unabhängigen Zufallsvariablen/vektoren $X_1, \dots, X_n \sim \mathbb{P}_\theta^0$.
- Wenn für ein Produktmodell die Menge $\mathcal{X}_0$ eindimensional ist, also z.B. $\mathcal{X}_0 = \mathbb{R}$ gilt, spricht man von einem *univariaten statistischen Modell*.

Bemerkungen

- Wenn für ein Produktmodell die Menge $\mathcal{X}_0$ mehrdimensional ist, also z.B. $\mathcal{X}_0 = \mathbb{R}^m, m > 1$ ist, spricht man auch von *multivariaten statistischen Modellen*. Wir betrachten in diesem Kurs nur univariate statistische Modelle.
- Die am häufigsten angewendeten statistischen Modelle sind parametrische stetige Produktmodelle.
- In einem konkreten Datenanalyseproblem nimmt man an, dass die beobachteten Stichprobenwerte $x = (x_1, \dots, x_n)$ der Stichprobe $X = (X_1, \dots, X_n)$ durch $\mathbb{P}_{\tilde{\theta}}$ generiert wurde, wobei $\tilde{\theta}$ hier den *wahren, aber unbekannten, Parameterwert* bezeichnet. Der wahre, aber unbekannten, Parameterwert $\tilde{\theta}$ bleibt auch nach der statistischen Analyse unbekannt.
- In der mathematischen Analyse von Inferenzmethoden betrachtet man alle möglichen wahren, aber unbekannten, Parameterwerte, schreibt also einfach $\{\mathbb{P}_{\theta} | \theta \in \Theta\}$.

Beispiele

Bernoullimodell

$\mathcal{M} := (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\})$ mit $\mathcal{X} := \{0, 1\}^n$, $\mathcal{A} := \mathcal{P}(\{0, 1\}^n)$, $\theta := \mu$, $\Theta :=]0, 1[$, also

$$\{\mathbb{P}_\theta | \theta \in \Theta\} := \left\{ \prod_{i=1}^n \text{Bern}(\mu) | \mu \in]0, 1[\right\}, \quad (2)$$

und damit $X_1, \dots, X_n \sim \text{Bern}(\mu)$.

Normalverteilungsmodell

$\mathcal{M} := (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\})$ mit $\mathcal{X} := \mathbb{R}^n$, $\mathcal{A} := \mathcal{B}(\mathbb{R}^n)$, $\theta := (\mu, \sigma^2)$, $\Theta := \mathbb{R} \times \mathbb{R}_{>0}$, also

$$\{\mathbb{P}_\theta | \theta \in \Theta\} := \left\{ \prod_{i=1}^n N(\mu, \sigma^2) | (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0} \right\}, \quad (3)$$

und damit $X_1, \dots, X_n \sim N(\mu, \sigma^2)$.

Statistische Modelle, Statistiken und Schätzer

- Statistische Modelle
- **Statistiken und Schätzer**
- Standardprobleme frequentistischer Inferenz
- Maximum-Likelihood Schätzer
- Selbstkontrollfragen

Definition (Statistik)

$\mathcal{M} := (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\})$ sei ein statistisches Modell und $(\Sigma, \mathcal{S})$ sei ein Messraum. Dann heißt eine Zufallsvariable $S : \mathcal{X} \rightarrow \Sigma$ eine *Statistik*.

Bemerkungen

- Beobachtungen und Statistiken werden durch Zufallsvariablen modelliert.
- Beobachtungen modellieren die stochastische Generation von Daten; Statistiken modellieren von **Datenanalyst:innen** konstruierte Funktionen, die aufgrund von Beobachtungen essentielle Größen extrahieren, aus denen sich sinnvolle Schlüsse ziehen lassen.

Beispiele (Statistiken)

$\mathcal{M}$ sei das Normalverteilungsmodell. Dann sind zum Beispiel

- das *Stichprobenmittel*

$$\bar{x}_n : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto \bar{x}_n(x) := \frac{1}{n} \sum_{i=1}^n x_i, \quad (4)$$

- die *Stichprobenvarianz*

$$s_n^2 : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}, x \mapsto s_n^2(x) := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2, \quad (5)$$

- und die *T-Statistik* für $\mu_0 \in \mathbb{R}$ und Stichprobenstandardabweichung s_n

$$t : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto t(x) := \sqrt{n} \left(\frac{\bar{x}_n - \mu_0}{s_n} \right) \quad (6)$$

Statistiken.

Definition (Schätzer)

$\mathcal{M} := (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\})$, $(\Sigma, \mathcal{S})$ sei ein Messraum, und $\tau : \Theta \rightarrow \Sigma$ sei eine Abbildung, die jedem $\theta \in \Theta$ eine Kenngröße $\tau(\theta) \in \Sigma$ zuordnet. Eine Statistik $\hat{\tau} : \mathcal{X} \rightarrow \Sigma$ heißt dann ein *Schätzer* für τ .

Bemerkungen

- Typische Beispiele für τ sind
 - $\tau(\theta) := \theta$ für die Schätzung von θ ,
 - $\tau(\theta) := \theta_i$ mit $\theta \in \mathbb{R}^d$, $d > 1$ für die Schätzung einer Komponente von θ ,
 - $\tau(\theta) := \mathbb{E}_\theta(X_0)$ für die Schätzung des Erwartungswerts einer Beobachtung,
 - $\tau(\theta) := \mathbb{V}_\theta(X_0)$ für die Schätzung der Varianz einer Beobachtung.
- Schätzer nehmen Zahlwerte in Σ an und heißen deshalb auch *Punktschätzer*.
- Nicht jeder Schätzer ist ein guter Schätzer, man definiert deshalb *Schätzgütekriterien*.
- Für $\hat{\tau}$ bei $\tau(\theta) := \theta$ schreibt man auch $\hat{\theta}$

Beispiele (Schätzer)

$\mathcal{M}$ sei das Normalverteilungsmodell.

- Dann ist zum Beispiel das Stichprobenmittel $\bar{x}_n : \mathbb{R}^n \rightarrow \mathbb{R}$ ein Schätzer für

$$\tau : \mathbb{R} \times \mathbb{R}_{>0} \rightarrow \mathbb{R}, (\mu, \sigma^2) \mapsto \tau(\mu, \sigma^2) := \mu. \quad (7)$$

Ebenso ist $\bar{x}_n$ ein Schätzer für

$$\tau : \mathbb{R} \times \mathbb{R}_{>0} \rightarrow \mathbb{R}, (\mu, \sigma^2) \mapsto \tau(\mu, \sigma^2) := \mathbb{E}_{\mu, \sigma^2}(X_0) = \mu. \quad (8)$$

- Weiterhin ist die konstante Funktion

$$\hat{\tau} : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto \hat{\tau}(x) := 42 \quad (9)$$

ein Schätzer für

$$\tau : \mathbb{R} \times \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}, (\mu, \sigma^2) \mapsto \tau(\mu, \sigma^2) := \sigma^2. \quad (10)$$

Dass eine Funktion $\hat{\tau} : \mathcal{X} \rightarrow \Sigma$ ein Schätzer ist, heißt nicht, dass sie ein guter Schätzer ist!

Gütekriterien für Schätzer sind der Inhalt von Vorlesungseinheit (8) Schätzereigenschaften.

Statistische Modelle, Statistiken und Schätzer

- Statistische Modelle
- Statistiken und Schätzer
- **Standardprobleme frequentistischer Inferenz**
- Maximum-Likelihood Schätzer
- Selbstkontrollfragen

(1) Parameterschätzung

Ziel der Parameterschätzung ist es, einen möglichst guten Tipp für den wahren, aber unbekannten, Parameterwert (oder eine Funktion dessen) abzugeben, typischerweise basierend auf der Beobachtung einer Realisierung von $X_1, \dots, X_n \sim p_\theta$.

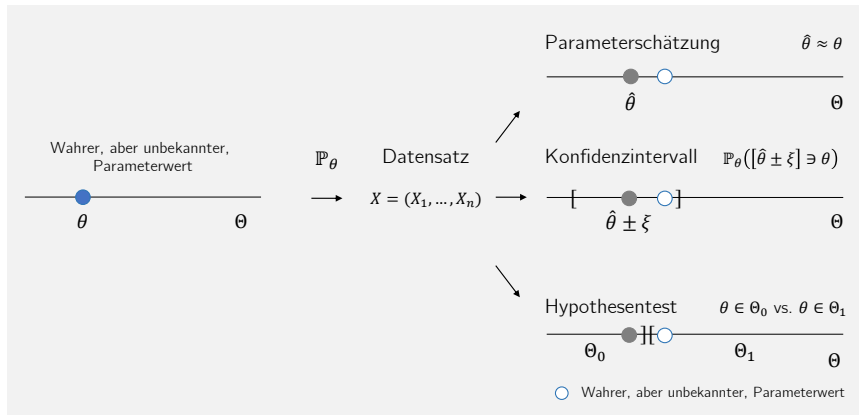
(2) Konfidenzintervalle

Das Ziel der Bestimmung von Konfidenzintervallen ist es, basierend auf der Verteilung möglicher Parameterschätzwerte eine quantitative Aussage über die mit dem Schätzwert assoziierte Unsicherheit zu treffen.

(3) Hypothesentest

Das Ziel der Auswertung von Hypothesentests ist es, basierend auf der angenommenen Verteilung der Beobachtungen $X_1, \dots, X_n$ in einer möglichst sinnvollen Form zu entscheiden, ob der wahre, aber unbekannte Parameterwert, in einer von zwei sich gegenseitig ausschließenden Untermengen des Parameterraumes, welche man als Hypothesen bezeichnet, liegt.

Standardprobleme frequentistischer Inferenz



Die Standardannahmen frequentistischer Inferenz

- $\mathcal{M}$ sei ein statistisches Modell mit $X_1, \dots, X_n \sim p_\theta$. Es wird angenommen, dass ein vorliegender Datensatz eine der möglichen Realisierungen von $X_1, \dots, X_n \sim p_\theta$ ist.
- Aus frequentistischer Sicht kann man das Experiment unendlich oft wiederholen und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. das Stichprobenmittel:

$$\text{Datensatz (1)} : x^{(1)} = (x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}) \text{ mit } \bar{x}_n^{(1)} = \frac{1}{n} \sum_{i=1}^n x_i^{(1)}$$

$$\text{Datensatz (2)} : x^{(2)} = (x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)}) \text{ mit } \bar{x}_n^{(2)} = \frac{1}{n} \sum_{i=1}^n x_i^{(2)}$$

$$\text{Datensatz (3)} : x^{(3)} = (x_1^{(3)}, x_2^{(3)}, \dots, x_n^{(3)}) \text{ mit } \bar{x}_n^{(3)} = \frac{1}{n} \sum_{i=1}^n x_i^{(3)}$$

$$\text{Datensatz (4)} : x^{(4)} = (x_1^{(4)}, x_2^{(4)}, \dots, x_n^{(4)}) \text{ mit } \bar{x}_n^{(4)} = \frac{1}{n} \sum_{i=1}^n x_i^{(4)}$$

$$\text{Datensatz (5)} : x^{(5)} = \dots$$

- Um die Qualität statischer Methoden zu beurteilen betrachtet die frequentistische Statistik deshalb die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken unter Annahme von $X_1, \dots, X_n \sim p_\theta$. Was zum Beispiel ist die Verteilung von $\bar{x}_n^{(1)}$, $\bar{x}_n^{(2)}$, $\bar{x}_n^{(3)}$, $\bar{x}_n^{(4)}$, ... also die Verteilung der Zufallsvariable $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$?
- Wenn eine statistische Methode im Sinne der frequentistischen Standardannahme gut ist, dann heißt das also, dass sie bei häufiger Anwendung "im Mittel" gut ist. Im Einzelfall, also bei nur einem Datensatz, kann sie auch recht schlecht sein.

Anwendungsbeispiel

Beck Depressions-Inventar II (BDI II)

- Weltweit verbreitetster Fragebogen zur Selbstbeurteilung des Depressionsschweregrades
- Konzipiert von Beck (1961), seitdem kleinere Revisionen
- 21 Fragen ("Items") mit inhaltlichen Aussagen aufsteigendem Schweregrads (0-3) zu diagnoserelevanten Themengebieten wie Traurigkeit, Versagensgefühle, Schlafverhalten, Selbstmordgedanken, ...
- Beispielitem (14) Wertlosigkeit
 - 0 Ich fühle mich nicht wertlos.
 - 1 Ich halte mich für weniger wertvoll und nützlich als sonst.
 - 2 Verglichen mit anderen Menschen fühle ich mich viel weniger wert.
 - 3 Ich fühle mich völlig wertlos.
- Auswertung durch Bestimmung des Summenwerts (BDI II Σ) über Items.
- Ergebnisklassifikation: 0-8 keine Depression, 9-13 minimale Depression, 14-19 leichte Depression, 20-28 mittelschwere Depression, 29-63 schwere Depression

Standardprobleme frequentistischer Inferenz

Anwendungsbeispiel

Beck Depressions-Inventar II (BDI II)

Fragebogen			
Name	Aber	Gesichte m / w	Datum
Anleitung: Dieser Fragebogen enthält 21 Gruppen von Aussagen. Bitte lesen Sie jede dieser Gruppen von Aussagen sorgfältig durch und suchen Sie sich dann in jeder Gruppe eine Aussage heraus, die am besten beschreibt, wie Sie sich in den letzten zwei Wochen, einschließlich heute, gefühlt haben. Kreuzen Sie die Zahl neben der Aussage an, die Sie sich am besten zugehörig fühlen (0, 1, 2 oder 3). Falls in einer Gruppe mehrere Aussagen gleichermaßen auf Sie zutreffen, kreuzen Sie die Aussage mit der höheren Zahl an. Achten Sie bitte darauf, dass Sie in jeder Gruppe nicht mehr als eine Aussage ankreuzen, das gilt auch für Gruppe 18 (Veränderungen der Schlafgewohnheiten) oder Gruppe 19 (Veränderungen des Appetits).			
1.) Traurigkeit 0 Ich bin nicht traurig. 1 Ich bin oft traurig. 2 Ich bin ständig traurig. 3 Ich bin so traurig oder unglücklich, dass ich es nicht aushalte.	6.) Bestrafungsgefühle 0 Ich habe nicht das Gefühl, für etwas bestraft zu sein. 1 Ich habe das Gefühl, vielleicht bestraft zu werden. 2 Ich erwarte, bestraft zu werden. 3 Ich habe das Gefühl, bestraft zu sein.		
2.) Pessimismus 0 Ich sehe nicht mutlos in die Zukunft. 1 Ich sehe mutloser in die Zukunft als sonst. 2 Ich bin mutlos und erwarte nicht, dass meine Situation besser wird. 3 Ich glaube, dass meine Zukunft hoffnungslos ist und nur noch schlechter wird.	7.) Selbstablehnung 0 Ich halte von mir genauso viel wie immer. 1 Ich habe Vertrauen in mich verloren. 2 Ich bin von mir enttäuscht. 3 Ich lehne mich völlig ab.		
3.) Versagensgefühle 0 Ich fühle mich nicht als Versager. 1 Ich habe häufiger Versagensgefühle. 2 Wenn ich zurückblicke, sehe ich eine Menge Fehlchancen. 3 Ich habe das Gefühl, als Mensch ein völliger Versager zu sein.	8.) Selbstvorwürfe 0 Ich kritisiere oder tadle mich nicht mehr als sonst. 1 Ich bin mir gegenüber kritischer als sonst. 2 Ich kritisiere mich für all meine Mängel. 3 Ich gebe mir die Schuld für alles Schlimme, was passiert.		
4.) Verlust von Freude 0 Ich kann die Dinge genauso gut genießen wie früher. 1 Ich kann die Dinge nicht mehr so genießen wie früher. 2 Dinge, die mir früher Freude gemacht haben, kann ich kaum mehr genießen. 3 Dinge, die mir früher Freude gemacht haben, kann ich überhaupt nicht mehr genießen.	9.) Selbstmordgedanken 0 Ich denke nicht daran, mir etwas anzutun. 1 Ich denke manchmal an Selbstmord, aber ich würde es nicht tun. 2 Ich möchte mich am liebsten umbringen. 3 Ich würde mich umbringen, wenn ich die Gelegenheit dazu hätte.		
5.) Schuldgefühle 0 Ich habe keine besonderen Schuldgefühle. 1 Ich habe oft Schuldgefühle wegen Dingen, die ich getan habe oder hätte tun sollen. 2 Ich habe die meiste Zeit Schuldgefühle. 3 Ich habe ständig Schuldgefühle.	10.) Weinen 0 Ich weine nicht öfter als früher. 1 Ich weine jetzt mehr als früher. 2 Ich weine beim geringsten Anlass. 3 Ich möchte gern weinen, aber ich kann nicht.		

- 11.) Unruhe**
0 Ich bin nicht unruhiger als sonst.
1 Ich bin unruhiger als sonst.
2 Ich bin so unruhig, dass es mir schwerfällt, still zu sitzen.
3 Ich bin so unruhig, dass ich mich ständig bewegen oder etwas tun muss.

- 12.) Interessenverlust**
0 Ich habe das Interesse an anderen Menschen oder an Tätigkeiten nicht verloren.
1 Ich habe weniger Interesse an anderen Menschen oder an Dingen als sonst.
2 Ich habe das Interesse an anderen Menschen oder Dingen zum größten Teil verloren.
3 Es fällt mir schwer, mich überhaupt für irgend etwas zu interessieren.

- 13.) Entschlussunfähigkeit**
0 Ich bin so entscheidungsfreudig wie immer.
1 Es fällt mir schwerer als sonst, Entscheidungen zu treffen.
2 Es fällt mir sehr viel schwerer als sonst, Entscheidungen zu treffen.
3 Ich habe Mühe, überhaupt Entscheidungen zu treffen.

- 14.) Wertlosigkeit**
0 Ich fühle mich nicht wertlos.
1 Ich habe mich für weniger wertvoll und nützlich als sonst.
2 Verglichen mit anderen Menschen fühle ich mich viel weniger wert.
3 Ich fühle mich völlig wertlos.

- 15.) Energieverlust**
0 Ich habe so viel Energie wie immer.
1 Ich habe weniger Energie als sonst.
2 Ich habe so wenig Energie, dass ich kaum noch etwas schaffe.
3 Ich habe keine Energie mehr, um überhaupt noch etwas zu tun.

- 16.) Veränderungen der Schlafgewohnheiten**
0 Meine Schlafgewohnheiten haben sich nicht verändert.
1a Ich schlafe etwas mehr als sonst.
1b Ich schlafe etwas weniger als sonst.
2a Ich schlafe viel mehr als sonst.
2b Ich schlafe viel weniger als sonst.
3a Ich schlafe fast den ganzen Tag.
3b Ich wache 1-2 Stunden früher auf als gewöhnlich und kann dann nicht mehr einschlafen.

- 17.) Reizbarkeit**
0 Ich bin nicht reizbarer als sonst.
1 Ich bin reizbarer als sonst.
2 Ich bin viel reizbarer als sonst.
3 Ich fühle mich dauernd gereizt.

- 18.) Veränderungen des Appetits**
0 Mein Appetit hat sich nicht verändert.
1a Mein Appetit ist etwas schlechter als sonst.
1b Mein Appetit ist etwas größer als sonst.
2a Mein Appetit ist viel schlechter als sonst.
2b Mein Appetit ist viel größer als sonst.
3a Ich habe überhaupt keinen Appetit.
3b Ich habe ständig Heißhunger.

- 19.) Konzentrationsschwierigkeiten**
0 Ich kann mich so gut konzentrieren wie immer.
1 Ich kann mich nicht mehr so gut konzentrieren wie sonst.
2 Es fällt mir schwer, mich längere Zeit auf irgend etwas zu konzentrieren.
3 Ich kann mich überhaupt nicht mehr konzentrieren.

- 20.) Ermüdung oder Erschöpfung**
0 Ich fühle mich nicht müde oder erschöpfter als sonst.
1 Ich werde schneller müde oder erschöpft als sonst.
2 Für viele Dinge, die ich üblicherweise tue, bin ich zu müde oder erschöpft.
3 Ich bin so müde oder erschöpft, dass ich fast nichts mehr tun kann.

- 21.) Verlust an sexuellem Interesse**
0 Mein Interesse an Sexualität hat sich in letzter Zeit nicht verändert.
1 Ich interessiere mich weniger für Sexualität als früher.
2 Ich interessiere mich jetzt viel weniger für Sexualität.
3 Ich habe das Interesse an Sexualität völlig verloren.

Anwendungsbeispiel

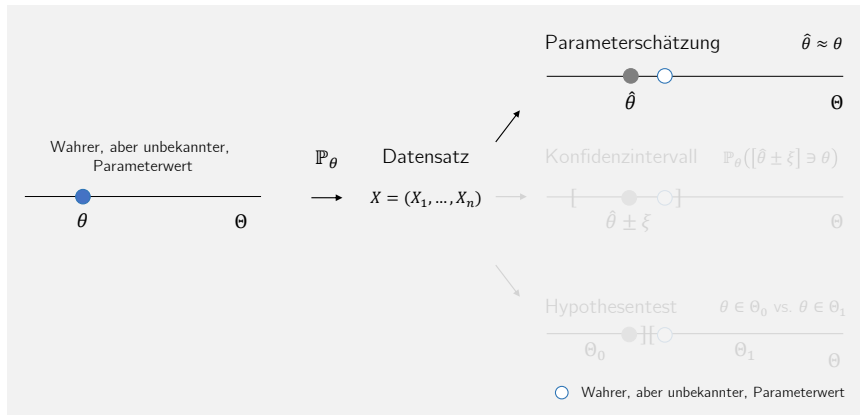
Beck Depressions-Inventar II (BDI II)

VP	BDI II Σ	Modellannahmen
1	11	Wir legen das Normalverteilungsmodell zugrunde, d.h. wir nehmen an, dass die BDI II Σ Werte Realisierungen von u.i.v. Zufallsvariablen $X_1, \dots, X_{12} \sim N(\mu, \sigma^2)$ sind. Dies entspricht der Annahme, dass sich der BDI II Σ Wert einer VP durch Addition einer normalverteilten Fehlervariable mit Erwartungswertparameter 0 und Varianzparameter σ^2 zu dem über VP identischen Wert μ ergibt.
2	14	
3	10	
4	9	
5	8	
6	17	
7	3	Fragestellungen (1) Was sind sinnvolle Tipps für die wahren, aber unbekannten, Parameterwerte μ und σ^2 ? (2) Wie hoch ist im frequentistischen Sinn die mit diesen Tipps assoziierte Unsicherheit? (3) Entscheiden wir uns sinnvollerweise für die Hypothese, dass $\mu \leq 8$ ist oder dass $\mu > 8$ ist?
8	20	
9	13	
10	3	
11	6	
12	2	

Statistische Modelle, Statistiken und Schätzer

- Statistische Modelle
- Statistiken und Schätzer
- Standardprobleme frequentistischer Inferenz
- **Maximum-Likelihood Schätzer**
- Selbstkontrollfragen

Maximum-Likelihood Schätzer



Definition (Parameterpunktschätzer)

$\mathcal{M} := (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\})$ sei ein statistisches Modell, $(\Theta, \mathcal{S})$ sei ein Messraum, und $\hat{\theta} : \mathcal{X} \rightarrow \Theta$ sei eine Abbildung. Dann nennen wir $\hat{\theta}$ einen *Parameterpunktschätzer* für θ .

- Parameterpunktschätzer nennt man auch einfach *Parameterschätzer*.
- Parameterpunktschätzer sind Schätzer mit $\tau := \text{id}_\Theta$
- Parameterschätzer nehmen Zahlwerte in Θ an.
- Notationstechnisch wird oft nicht zwischen $\hat{\theta}$ und $\hat{\theta}(x)$ unterschieden.

Alternative Visualisierung der Parameterpunktschätzung (Georgii, 2009, Kapitel 7)

- Man beachte, dass hier die Notation $\vartheta = \theta$, $P_{\vartheta} = \mathbb{P}_{\theta}$, $T(x) = \hat{\theta}(x)$ benutzt wird.

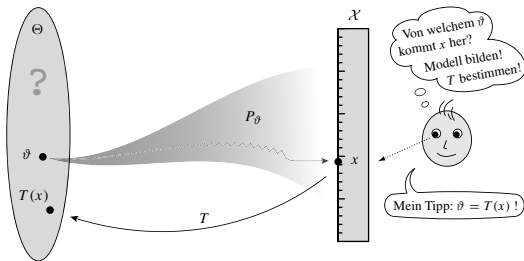


Abbildung 7.1: Das Prinzip des Schätzens: Eine unbekannte Größe ϑ soll ermittelt werden; die möglichen ϑ 's bilden eine Menge Θ . Das tatsächlich vorliegende ϑ bestimmt ein Zufallsgesetz P_{ϑ} , welches das Zufallsexperiment steuert und im konkreten Fall zum Messwert x führt. Um auf ϑ zurückschließen zu können, muss der Statistiker zunächst für jedes ϑ das Zufallsgesetz P_{ϑ} analysieren – das geschieht bei der Modellbildung. Die ermittelte Struktur aller P_{ϑ} 's muss er dann ausnutzen, um eine Abbildung $T : \mathcal{X} \rightarrow \Theta$ so geschickt zu konstruieren, dass der Wert $T(x)$ dem tatsächlich vorliegenden ϑ typischerweise möglichst nahe kommt – ganz egal, welches ϑ das nun wirklich ist.

Prinzipien zur Gewinnung von Parameterschätzern

- Die Definition eines Parameterschätzers macht keine Aussage darüber, wie man Parameterschätzer findet. Zur Gewinnung von Parameterschätzern haben sich deshalb verschiedene Prinzipien etabliert.
- Populäre Methoden zur Gewinnung von Parameterschätzern sind
 - Momentenmethode ($\approx$ est. 1890)
 - Maximum-Likelihood Methode ($\approx$ est. 1920)
 - M-, Z-, W-Schätzung ($\approx$ est. 1960)
- *Per se* garantiert keine der obengenannten Methoden, dass die mit ihrer Hilfe generierten Parameterschätzer in einem wohldefinierten Sinn gute Schätzer sind.
- Die Eigenschaften von durch die Maximum-Likelihood Methode generierten Schätzern sind generell wünschenswert. Wir betrachten also in der Folge nur die Maximum-Likelihood Methode genauer. Mithilfe der Maximum-Likelihood Methode generierte Parameterpunktschätzer nennen wir *Maximum-Likelihood (ML) Schätzer*.

Definition (Likelihood-Funktion und Log-Likelihood-Funktion)

$\mathcal{M}$ sei ein parametrisches statistisches Produktmodell mit WMF oder WDF p_θ , so dass $X_1, \dots, X_n \sim p_\theta$. Dann ist die *Likelihood Funktion* definiert als

$$L_n : \Theta \rightarrow [0, \infty[, \theta \mapsto L_n(\theta) := \prod_{i=1}^n p_\theta(x_i). \quad (11)$$

Die *Log-Likelihood-Funktion* ist definiert als

$$\ell_n : \Theta \rightarrow \mathbb{R}, \theta \mapsto \ell_n(\theta) := \ln L_n(\theta). \quad (12)$$

Bemerkungen

- L_n ist eine Funktion des Parameters eines statistischen Modells.
- Werte von L_n sind die gemeinsamen W-Massen bzw. W-Dichten von Daten $x_1, \dots, x_n$.
- Generell gibt es keinen Grund anzunehmen, dass L_n über Θ zu 1 integriert.
- Die Likelihood-Funktion ist also keine WMF oder WDF.
- Die Log-Likelihood-Funktion ist die logarithmierte Likelihood-Funktion.

Definition (Maximum-Likelihood Schätzer)

$\mathcal{M}$ sei ein parametrisches statistisches Produktmodell mit Parameter $\theta \in \Theta$. Ein *Maximum-Likelihood (ML) Schätzer* von θ ist definiert als

$$\hat{\theta}_n^{\text{ML}} : \mathcal{X} \rightarrow \Theta, x \mapsto \hat{\theta}_n^{\text{ML}}(x) := \arg \max_{\theta \in \Theta} L_n(\theta) = \arg \max_{\theta \in \Theta} \ell_n(\theta). \quad (13)$$

Bemerkungen

- $L_n(\theta) := \prod_{i=1}^n p_\theta(x_i)$ hängt von $x_1, \dots, x_n$ ab, also hängt auch $\hat{\theta}_n^{\text{ML}}(x)$ von $x_1, \dots, x_n$ ab.
- Weil $\ln$ monoton steigend ist, entspricht eine Maximumstelle von ℓ_n einer Maximumstelle von L_n .
- Das Arbeiten mit der Log-Likelihood-Funktion ist oft einfacher als mit der Likelihood Funktion.
- Multiplikation von L_n mit einer positive Konstante, die nicht von θ abhängt, verändert einen ML Schätzer nicht, konstante additive Terme in der Log-Likelihood können also vernachlässigt werden.
- Maximum-Likelihood Schätzung ist ein Optimierungsproblem

Maximum-Likelihood Schätzung für parametrische statistische Produktmodelle

- (1) Formulierung der Log-Likelihood-Funktion.
- (2) Auswertung der Ableitung der Log-Likelihood-Funktion und Nullsetzen.
- (3) Auflösen nach potentiellen Maximumstellen.

Dabei nutzt man typischerweise

- Methoden der analytische Optimierung in klassischen Beispielen und
- Methoden der numerische Optimierung im Anwendungskontext.

Beispiel (Bernoullimodell)

$\mathcal{M}$ sei das Bernoullimodell, also $X_1, \dots, X_n \sim \text{Bern}(\mu)$.

(1) Formulierung der Log-Likelihood-Funktion

Es gilt

$$L_n :]0, 1[\rightarrow]0, 1[, \mu \mapsto L_n(\mu) := \prod_{i=1}^n \mu^{x_i} (1 - \mu)^{1-x_i} = \mu^{\sum_{i=1}^n x_i} (1 - \mu)^{n - \sum_{i=1}^n x_i}. \quad (14)$$

Logarithmieren ergibt

$$\ell_n :]0, 1[\rightarrow \mathbb{R}, \mu \mapsto \ell_n(\mu) = \ln \mu \sum_{i=1}^n x_i + \ln(1 - \mu) \left(n - \sum_{i=1}^n x_i \right). \quad (15)$$

Beispiel (Bernoullimodell)

(2) Auswertung der Ableitung der Log-Likelihood-Funktion und Nullsetzen Es gilt

$$\begin{aligned}\frac{d}{d\mu}\ell_n(\mu) &= \frac{d}{d\mu} \left(\ln \mu \sum_{i=1}^n x_i + \ln(1 - \mu) \left(n - \sum_{i=1}^n x_i \right) \right) \\ &= \frac{d}{d\mu} \ln \mu \sum_{i=1}^n x_i + \frac{d}{d\mu} \ln(1 - \mu) \left(n - \sum_{i=1}^n x_i \right) \\ &= \frac{1}{\mu} \sum_{i=1}^n x_i - \frac{1}{1 - \mu} \left(n - \sum_{i=1}^n x_i \right).\end{aligned}\tag{16}$$

Die *Maximum-Likelihood Gleichung* ergibt sich also zu

$$\frac{1}{\hat{\mu}_n^{\text{ML}}} \sum_{i=1}^n x_i - \frac{1}{1 - \hat{\mu}_n^{\text{ML}}} \left(n - \sum_{i=1}^n x_i \right) = 0.\tag{17}$$

Beispiel (Bernoullimodell)

(3) Auflösen nach potentiellen Maximumstellen

Es gilt

$$\begin{aligned} & \frac{1}{\hat{\mu}_n^{\text{ML}}} \sum_{i=1}^n x_i - \frac{1}{1 - \hat{\mu}_n^{\text{ML}}} \left(n - \sum_{i=1}^n x_i \right) = 0 \\ \Leftrightarrow & \hat{\mu}_n^{\text{ML}} (1 - \hat{\mu}_n^{\text{ML}}) \left(\frac{1}{\hat{\mu}_n^{\text{ML}}} \sum_{i=1}^n x_i - \frac{1}{1 - \hat{\mu}_n^{\text{ML}}} \left(n - \sum_{i=1}^n x_i \right) \right) = 0 \\ \Leftrightarrow & \sum_{i=1}^n x_i - \hat{\mu}_n^{\text{ML}} \sum_{i=1}^n x_i - n \hat{\mu}_n^{\text{ML}} + \hat{\mu}_n^{\text{ML}} \sum_{i=1}^n x_i = 0 \quad (18) \\ \Leftrightarrow & n \hat{\mu}_n^{\text{ML}} = \sum_{i=1}^n x_i \\ \Leftrightarrow & \hat{\mu}_n^{\text{ML}} = \frac{1}{n} \sum_{i=1}^n x_i. \end{aligned}$$

Beispiel (Bernoullimodell)

$\hat{\mu}_n^{\text{ML}} = \frac{1}{n} \sum_{i=1}^n x_i$ ist also ein potentieller Maximum-Likelihood Schätzer von μ . Dies kann durch Betrachten der zweiten Ableitung von ℓ_n verifiziert werden, worauf hier verzichtet werden soll.

$$\hat{\mu}_n^{\text{ML}} : \{0, 1\}^n \rightarrow [0, 1], x \mapsto \hat{\mu}_n^{\text{ML}}(x) := \frac{1}{n} \sum_{i=1}^n x_i \quad (19)$$

ist also ein Maximum-Likelihood Schätzer von μ im Bernoullimodell.

Beispiel (Normalverteilungsmodell)

$\mathcal{M}$ sei das Normalverteilungsmodell, also $X_1, \dots, X_n \sim N(\mu, \sigma^2)$

(1) Formulierung der Log-Likelihood-Funktion

Es gilt

$$\begin{aligned} L_n : \mathbb{R} \times \mathbb{R}_{>0} &\rightarrow \mathbb{R}_{>0}, (\mu, \sigma^2) \mapsto L_n(\mu, \sigma^2) := \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right). \end{aligned} \quad (20)$$

Logarithmieren ergibt

$$\ell_n : \mathbb{R} \times \mathbb{R}_{>0} \rightarrow \mathbb{R}, (\mu, \sigma^2) \mapsto \ell_n(\mu, \sigma^2) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \quad (21)$$

Beispiel (Normalverteilungsmodell)

(2) Auswertung der Ableitung der Log-Likelihood-Funktion und Nullsetzen Es ergibt sich

$$\frac{d}{d\mu} \ell_n(\mu) = -\frac{d}{d\mu} \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 = -\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{d}{d\mu} (x_i - \mu)^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu). \quad (22)$$

Weiterhin ergibt sich

$$\frac{d}{d\sigma^2} \ell_n(\sigma^2) = -\frac{n}{2} \frac{d}{d\sigma^2} \ln \sigma^2 - \frac{d}{d\sigma^2} \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2. \quad (23)$$

Die Maximum-Likelihood Gleichungen haben also die Form

$$\begin{aligned} \sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}}) &= 0 \\ -\frac{n}{2\hat{\sigma}_n^{\text{ML}2}} + \frac{1}{2\hat{\sigma}_n^{\text{ML}4}} \sum_{i=1}^n (x_i - \mu)^2 &= 0 \end{aligned} \quad (24)$$

Beispiel (Normalverteilungsmodell)

(3) Auflösen nach potentiellen Maximumstellen Es ergibt sich

$$\sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}}) = 0 \Leftrightarrow \sum_{i=1}^n x_i = n\hat{\mu}_n^{\text{ML}} \Leftrightarrow \hat{\mu}_n^{\text{ML}} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (25)$$

Also ist $\hat{\mu}_n^{\text{ML}} = n^{-1} \sum_{i=1}^n x_i$ ein potentieller Maximum-Likelihood Schätzer von μ .

Einsetzen ergibt dann weiterhin

$$\begin{aligned} -\frac{n}{2\hat{\sigma}_n^{2\text{ML}}} + \frac{1}{2\hat{\sigma}_n^{4\text{ML}}} \sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}})^2 &= 0 \\ \Leftrightarrow -n\hat{\sigma}_n^{2\text{ML}} + \sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}})^2 &= 0 \\ \Leftrightarrow \hat{\sigma}_n^{2\text{ML}} &= \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}})^2. \end{aligned} \quad (26)$$

$\hat{\sigma}_n^{2\text{ML}} = n^{-1} \sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}})^2$ ist also potentieller Maximum-Likelihood Schätzer von σ^2 .

Beispiel (Normalverteilungsmodell)

Beide potentiellen Maximum-Likelihood Schätzer können durch Betrachten der zweiten Ableitung von ℓ_n verifiziert werden, worauf hier verzichtet werden soll. Also sind

$$\hat{\mu}_n^{\text{ML}} : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto \hat{\mu}_n^{\text{ML}}(x) := \frac{1}{n} \sum_{i=1}^n x_i \quad (27)$$

und

$$\hat{\sigma}_n^{2\text{ML}} : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}, x \mapsto \hat{\sigma}_n^{2\text{ML}}(x) := \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}})^2. \quad (28)$$

die Maximum-Likelihood Schätzer von μ und σ^2 im Normalverteilungsmodell. $\hat{\mu}_n^{\text{ML}}$ ist identisch mit dem Stichprobenmittel $\bar{X}_n$, $\hat{\sigma}_n^{2\text{ML}}$ ist dagegen nicht identisch mit der Stichprobenvarianz S_n^2 .

Anwendungsbeispiel

Beck Depressions-Inventar II (BDI II)

VP	BDI II Σ	Modellannahmen
1	11	Wir legen das Normalverteilungsmodell zugrunde, d.h. wir nehmen an, dass die BDI II Σ Werte Realisierungen von u.i.v. Zufallsvariablen $X_1, \dots, X_{12} \sim N(\mu, \sigma^2)$ sind. Dies entspricht der Annahme, dass sich der BDI II Σ Wert einer VP durch Addition einer normalverteilten Fehlervariable mit Erwartungswertparameter 0 und Varianzparameter σ^2 zu dem über VP identischen Wert μ ergibt.
2	14	
3	10	
4	9	
5	8	
6	17	
7	3	Fragestellungen (1) Was sind sinnvolle Tipps für die wahren, aber unbekannten, Parameterwerte μ und σ^2 ? (2) Wie hoch ist im frequentistischen Sinne die mit diesen Tipps assoziierte Unsicherheit? (3) Entscheiden wir uns sinnvollerweise für die Hypothese, dass $\mu \leq 8$ ist oder dass $\mu > 8$ ist?
8	20	
9	13	
10	3	
11	6	
12	2	

Anwendungsbeispiel

Beck Depressions-Inventar II (BDI II)

VP	BDI II Σ	Modellannahmen
1	11	Wir legen das Normalverteilungsmodell zugrunde, d.h. wir nehmen an, dass die BDI II Σ Werte Realisierungen von u.i.v. Zufallsvariablen $X_1, \dots, X_{12} \sim N(\mu, \sigma^2)$ sind. Dies entspricht der Annahme, dass sich der BDI II Σ Wert einer VP durch Addition einer normalverteilten Fehlervariable mit Erwartungswertparameter 0 und Varianzparameter σ^2 zu dem über VP identischen Wert μ ergibt.
2	14	
3	10	
4	9	
5	8	
6	17	
7	3	Fragestellungen (1) Was sind sinnvolle Tipps für die wahren, aber unbekannten, Parameterwerte μ und σ^2 ? (2) Wie hoch ist im frequentistischen Sinne die mit diesen Tipps assoziierte Unsicherheit? (3) Entscheiden wir uns sinnvollerweise für die Hypothese, dass $\mu \leq 8$ ist oder dass $\mu > 8$ ist?
8	20	
9	13	
10	3	
11	6	
12	2	

Anwendungsbeispiel

Beck Depressions-Inventar II (BDI II)

VP	BDI II Σ	Modellannahmen
1	11	Wir legen das Normalverteilungsmodell zugrunde, d.h. wir nehmen an, dass die BDI II Σ Werte Realisierungen von u.i.v. Zufallsvariablen $X_1, \dots, X_{12} \sim N(\mu, \sigma^2)$ sind. Dies entspricht der Annahme, dass sich der BDI II Σ Wert einer VP durch Addition einer normalverteilten Fehlervariable mit Erwartungswertparameter 0 und Varianzparameter σ^2 zu dem über VP identischen Wert μ ergibt.
2	14	
3	10	
4	9	
5	8	
6	17	
7	3	Fragestellungen (1) Was sind sinnvolle Tipps für die wahren, aber unbekannten, Parameterwerte μ und σ^2 ? Basierend auf dem Prinzip der Maximum-Likelihood Schätzung sind
8	20	
9	13	
10	3	
11	6	
12	2	

$$\hat{\mu}_n^{\text{ML}} = 9.6 \text{ und } \hat{\sigma}_n^{2\text{ML}} = 29.7 \quad (29)$$

sinnvolle Tipps für μ und σ^2 .

Statistische Modelle, Statistiken und Schätzer

- Statistische Modelle
- Statistiken und Schätzer
- Standardprobleme frequentistischer Inferenz
- Maximum-Likelihood Schätzer
- **Selbstkontrollfragen**

Selbstkontrollfragen

1. Definieren und erläutern Sie den Begriff des statistischen Modells.
2. Definieren und erläutern Sie den Begriff eines parametrischen statistischen Produktmodells.
3. Erläutern Sie den Unterschied zwischen univariaten und multivariaten statistischen Modellen.
4. Formulieren Sie das Bernoulli-Modell.
5. Formulieren Sie das Normalverteilungsmodell.
6. Definieren und erläutern Sie den Begriff der Statistik.
7. Definieren und erläutern Sie den Begriff des Schätzers.
8. Nennen und erläutern Sie die Standardprobleme der frequentistischen Inferenz.
9. Erläutern Sie die Standardannahmen der frequentistischen Inferenz.
10. Definieren und erläutern Sie den Begriff des Parameterpunktschätzers.
11. Definieren Sie die Begriffe der Likelihood-Funktion und der Log-Likelihood Funktion.
12. Definieren Sie den Begriff des Maximum-Likelihood Schätzers.
13. Erläutern Sie das Vorgehen zur ML Schätzung für ein parametrisches statistisches Produktmodell.
14. Geben Sie den ML Schätzer für den Parameter μ des Bernoulli-Modells an.
15. Geben Sie den ML Schätzer für den Parameter μ des Normalverteilungsmodells an.
16. Geben Sie den ML Schätzer für den Parameter σ^2 des Normalverteilungsmodells an.

Literatur

- Beck, A. T. (1961). An Inventory for Measuring Depression. *Archives of General Psychiatry*, 4(6):561.
- Georgii, H.-O. (2009). *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. DeGruyter-Lehrbuch. de Gruyter, Berlin, 4., überarb. und erw. aufl edition.



Inferenzstatistik

BSc Psychologie SoSe 2021

Prof. Dr. Dirk Ostwald

(8) Schätzeigenschaften

Die Lektüre von Georgii (2009, Abschnitte 7.4, 7.5, und 7.7) wird empfohlen. Die hier gewählte Darstellung von Schätzereigenschaften bei endlichen Stichproben folgt Wasserman (2004, Chapter 9), Held and Sabanés Bové (2014, Sections 3.1,3.2,4.1,4.2), und DeGroot and Schervish (2012, Section 8.8). Die hier gewählte Darstellung von asymptotischen Schätzereigenschaften folgt Casella and Berger (2012, Section 10.1).

Schätzereigenschaften

- Vorbemerkungen
- Schätzereigenschaften bei endlichen Stichproben
- Asymptotische Schätzereigenschaften
- Eigenschaften von Maximum-Likelihood Schätzern
- Selbstkontrollfragen

Schätzereigenschaften

- **Vorbemerkungen**
- Schätzereigenschaften bei endlichen Stichproben
- Asymptotische Schätzereigenschaften
- Eigenschaften von Maximum-Likelihood Schätzern
- Selbstkontrollfragen

Vorbemerkungen

Wir gehen von einem statistischem parametrischem Produktmodell $\mathcal{M} := \{\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\}\}$ mit n -dimensionalen Stichprobenraum (z.B. $\mathcal{X} := \mathbb{R}^n$), d -dimensionalen Parameterraum $\Theta \subset \mathbb{R}^d$ und gegebener WMF oder WDF p_θ für alle $\theta \in \Theta$ aus. $X_1, \dots, X_n \sim p_\theta$ bezeichnet die zu $\mathcal{M}$ gehörende Stichprobe unabhängig und identisch verteilter Zufallsvariablen, es gilt also $X_0 \sim p_\theta$ und $X_i \sim p_\theta$ für alle $i = 1, \dots, n$.

Für einen Messraum $(\Sigma, \mathcal{S})$ sei $\hat{\tau} : \mathcal{X} \rightarrow \Sigma$ ein Schätzer von $\tau : \Theta \rightarrow \Sigma$. Wir betrachten Erwartungswerts-Varianz-, und Standardabweichungsschätzer, also Schätzer für

$$\tau : \Theta \rightarrow \Sigma, \theta \mapsto \tau(\theta) \text{ mit } \tau(\theta) := \mathbb{E}_\theta(X_0), \tau(\theta) := \mathbb{V}_\theta(X_0), \text{ und } \tau(\theta) := \mathbb{S}_\theta(X_0) \quad (1)$$

respektive, sowie Parameterschätzer, also Schätzer für

$$\tau : \Theta \rightarrow \Sigma, \tau(\theta) := \theta. \quad (2)$$

In dieser Vorlesungseinheit führen wir frequentistische Schätzereigenschaften ein. Frequentistische Schätzereigenschaften betrachten die Verteilung der Schätzwerte $\hat{\tau}(x_1, \dots, x_n)$ in Abhängigkeit von der Verteilung der Stichprobenwerte $x_1, \dots, x_n$. Weil die Stichprobenwerte zufällig sind, sind auch die Schätzwerte zufällig; ein Schätzer $\hat{\tau}$ ist also eine Zufallsvariable.

Wir unterscheiden zwischen *Schätzereigenschaften bei endlichen Stichproben*, d.h. Eigenschaften von $\hat{\tau}_n$ für ein fixes $n \in \mathbb{N}$ (z.B. $n = 12$) und *Asymptotischen Schätzereigenschaften*, d.h. Eigenschaften von $\hat{\tau}_n$ für unendlich groß werdende Stichproben mit $n \rightarrow \infty$.

Schätzereigenschaften

- Vorbemerkungen
- **Schätzereigenschaften bei endlichen Stichproben**
- Asymptotische Schätzereigenschaften
- Eigenschaften von Maximum-Likelihood Schätzern
- Selbstkontrollfragen

Schätzereigenschaften bei endlichen Stichproben

Ein Schätzer $\hat{\tau}_n$ heißt **erwartungstreu**, wenn sein Erwartungswert dem wahren, aber unbekannten, Wert $\tau(\theta)$ für alle $\theta \in \Theta$ gleicht.

Die **Varianz** eines Schätzers $\hat{\tau}_n$ ist die Varianz der Zufallsvariable $\hat{\tau}_n(X)$; der **Standardfehler** eines Schätzers $\hat{\tau}_n$ ist die Standardabweichung der Zufallsvariable $\hat{\tau}_n(X)$.

Die **Cramér-Rao-Ungleichung** gibt eine untere Schranke für die Varianz erwartungstreuer Schätzer an. Ein erwartungstreuer Schätzer mit Varianz gleich der in der Cramér-Rao-Ungleichung gegebenen unteren Schranke hat die kleinstmögliche Varianz aller erwartungstreuen Schätzer und ist in diesem Sinne ein “optimaler” Schätzer.

Der **mittlere quadratische Fehler** von $\hat{\tau}_n$ ist der Erwartungswert der quadrierten Abweichung von $\hat{\tau}_n(X)$ von $\tau(\theta)$ über Stichproben vom Umfang n .

Schätzereigenschaften bei endlichen Stichproben

Ein Schätzer $\hat{\tau}_n$ heißt **erwartungstreu**, wenn sein Erwartungswert dem wahren, aber unbekannten, Wert $\tau(\theta)$ für alle $\theta \in \Theta$ gleicht.

Die **Varianz** eines Schätzers $\hat{\tau}_n$ ist die Varianz der Zufallsvariable $\hat{\tau}_n(X)$; der **Standardfehler** eines Schätzers $\hat{\tau}_n$ ist die Standardabweichung der Zufallsvariable $\hat{\tau}_n(X)$.

Die **Cramér-Rao-Ungleichung** gibt eine untere Schranke für die Varianz erwartungstreuer Schätzer an. Ein erwartungstreuer Schätzer mit Varianz gleich der in der Cramér-Rao-Ungleichung gegebenen unteren Schranke hat die kleinstmögliche Varianz aller erwartungstreuen Schätzer und ist in diesem Sinne ein “optimaler” Schätzer.

Der **mittlere quadratische Fehler** von $\hat{\tau}_n$ ist der Erwartungswert der quadrierten Abweichung von $\hat{\tau}_n(X)$ von $\tau(\theta)$ über Stichproben vom Umfang n .

Definition (Fehler, Systematischer Fehler, und Erwartungstreue)

$X_1, \dots, X_n \sim p_\theta$ sei eine Stichprobe und $\hat{\tau}_n$ sei ein Schätzer für τ .

- Der *Fehler* von $\hat{\tau}_n$ ist definiert als

$$\hat{\tau}_n(X) - \tau(\theta). \quad (3)$$

- Der *systematische Fehler (Bias)* von $\hat{\tau}_n$ ist definiert als

$$B(\hat{\tau}_n) := \mathbb{E}_\theta(\hat{\tau}_n(X)) - \tau(\theta). \quad (4)$$

- $\hat{\tau}_n$ heißt *erwartungstreu (unbiased)*, wenn

$$B(\hat{\tau}_n) = 0 \Leftrightarrow \mathbb{E}_\theta(\hat{\tau}_n(X)) = \tau(\theta) \text{ für alle } \theta \in \Theta, n \in \mathbb{N}. \quad (5)$$

Andernfalls heißt $\hat{\tau}_n$ *verzerrt (biased)*.

Bemerkungen

- Der Fehler hängt von einer Realisation der Stichprobe ab.
- Der systematische Fehler ist der erwartete Fehler über viele Stichprobenrealisationen.
- Ein Parameterschätzer ist erwartungstreu, wenn $\mathbb{E}_\theta(\hat{\theta}_n(X)) = \theta$.

Theorem (Erwartungstreue von Stichprobenmittel und Stichprobenvarianz)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines statistischen parametrischen Produktmodells $\mathcal{M}$. Dann gilt:

- Das Stichprobenmittel

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i \quad (6)$$

ist ein erwartungstreuer Schätzer des Erwartungswerts $\mathbb{E}_\theta(X_0)$.

- Die Stichprobenvarianz

$$S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (7)$$

ist ein erwartungstreuer Schätzer der Varianz $\mathbb{V}_\theta(X_0)$.

Beweis

Um die Notation zu vereinfachen, definieren wir $\mathbb{E} := \mathbb{E}_\theta$ und $\mathbb{V} := \mathbb{V}_\theta$. Mit der Linearität von Erwartungswerten ergibt sich dann

$$\mathbb{E}(\bar{X}_n) = \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_0) = \frac{1}{n} n \mathbb{E}(X_0) = \mathbb{E}(X_0).$$

Dies zeigt die Erwartungstreue des Stichprobenmittels als Schätzer des Erwartungswertes.

Um die Erwartungstreue der Stichprobenvarianz zu zeigen, halten wir zunächst fest, dass

$$\mathbb{V}(\bar{X}_n) = \mathbb{V}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(X_i) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(X_0) = \frac{1}{n^2} n \mathbb{V}(X_0) = \frac{\mathbb{V}(X_0)}{n}.$$

gilt. Weiterhin halten wir ohne Beweis fest, dass

$$\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sum_{i=1}^n (X_i - \mu - \bar{X}_n + \mu)^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X}_n - \mu)^2.$$

Beweis

Es ergibt sich somit

$$\begin{aligned}\mathbb{E}\left((n-1)S_n^2\right) &= \mathbb{E}\left(\sum_{i=1}^n (X_i - \bar{X}_n)^2\right) \\&= \mathbb{E}\left(\sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X}_n - \mu)^2\right) \\&= \sum_{i=1}^n \mathbb{E}\left((X_i - \mu)^2\right) - n\mathbb{E}\left((\bar{X}_n - \mu)^2\right) \\&= n\mathbb{V}(X_0) - n\mathbb{V}(\bar{X}_n) \\&= n\mathbb{V}(X_0) - n\frac{\mathbb{V}(X_0)}{n} \\&= n\mathbb{V}(X_0) - \mathbb{V}(X_0) \\&= (n-1)\mathbb{V}(X_0)\end{aligned}$$

Schließlich ergibt sich

$$\mathbb{E}(S_n^2) = \mathbb{E}\left(\frac{1}{n-1}(n-1)S_n^2\right) = \frac{1}{n-1}\mathbb{E}\left((n-1)S_n^2\right) = \frac{1}{n-1}(n-1)\mathbb{V}(X_0) = \mathbb{V}(X_0)$$

und damit die Erwartungstreue der Stichprobenvarianz als Schätzer der Varianz.

□

Theorem (Verzerrtheit der Stichprobenstandardabweichung)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines statistischen parametrischen Produktmodells $\mathcal{M}$. Dann ist die Stichprobenstandardabweichung

$$S_n := \sqrt{S_n^2} \quad (8)$$

ein verzerrter Schätzer der Standardabweichung $\mathbb{S}_\theta(X_0)$.

Beweis

Wir halten zunächst fest, dass $\sqrt{\cdot}$ eine strikt konkave Funktion und $\sigma^2 > 0$ ist. Dann aber gilt mit der Jensenschen Ungleichung $\mathbb{E}(f(X)) < f(\mathbb{E}(X))$ für strikt konkave Funktionen, dass

$$\mathbb{E}(S) = \mathbb{E}\left(\sqrt{S_n^2}\right) < \sqrt{\mathbb{E}(S_n^2)} = \sqrt{\mathbb{V}_\theta(X_0)} = \mathbb{S}_\theta(X_0). \quad (9)$$

□

Bemerkung

- Nichtlineare Transformationen von erwartungstreuen Schätzern liefern oft verzerrte Schätzer.

Simulation ($X_1, \dots, X_{12} \sim N(\mu, \sigma^2)$ mit $n = 12$, $\mu = 1.7$, $\sigma^2 = 2$, $\sigma \approx 1.41$)

```
# Modellformulierung
mu      = 1.7                # wahrer, aber unbekannter, Erwartungswertparameter
sigsqr  = 2                  # wahrer, aber unbekannter, Varianzparameter
n       = 12                 # Stichprobengroesse n
ns      = 1e4                # Anzahl der Simulationen
X_bar   = rep(NaN, ns)       # Stichprobenmittellarray
S_sqr   = rep(NaN, ns)       # Stichprobenvarianzarray
S       = rep(NaN, ns)       # Stichprobenstandardabweichungarray

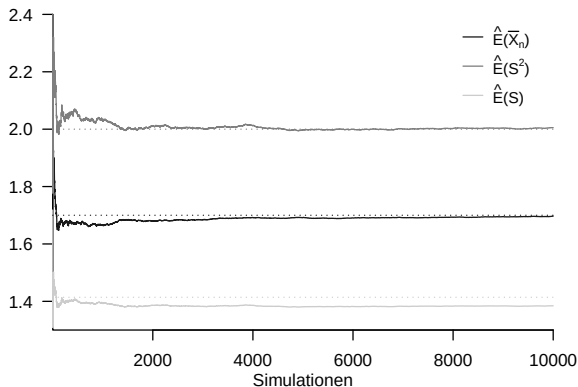
# Simulationsiterationen
for(s in 1:ns){

  # Stichprobenrealisation von  $X_1, \dots, X_{12}$ 
  x      = rnorm(n, mu, sqrt(sigsqr))

  # Erwartungswert-, Varianz-, Standardabweichungsschaetzer
  X_bar[s] = mean(x)          # Stichprobenmittel
  S_sqr[s] = var(x)           # Stichprobenvarianz
  S[s]     = sd(x)            # Stichprobenstandardabweichung
}

# Erwartungswertschaetzung
E_hat_X_bar = cumsum(X_bar)/(1:ns) #  $\mathbb{E}(\bar{X}_n)$  Schaetzungen
E_hat_S_sqr = cumsum(S_sqr)/(1:ns) #  $\mathbb{E}(S^2)$  Schaetzungen
E_hat_S     = cumsum(S)/(1:ns)    #  $\mathbb{E}(S)$  Schaetzungen
```

Simulation ($X_1, \dots, X_{12} \sim N(\mu, \sigma^2)$ mit $n = 12$, $\mu = 1.7$, $\sigma^2 = 2$, $\sigma \approx 1.41$)



Schätzereigenschaften bei endlichen Stichproben

Ein Schätzer $\hat{\tau}_n$ heißt **erwartungstreu**, wenn sein Erwartungswert dem wahren, aber unbekannten, Wert $\tau(\theta)$ für alle $\theta \in \Theta$ gleicht.

Die **Varianz** eines Schätzers $\hat{\tau}_n$ ist die Varianz der Zufallsvariable $\hat{\tau}_n(X)$; der **Standardfehler** eines Schätzers $\hat{\tau}_n$ ist die Standardabweichung der Zufallsvariable $\hat{\tau}_n(X)$.

Die **Cramér-Rao-Ungleichung** gibt eine untere Schranke für die Varianz erwartungstreuer Schätzer an. Ein erwartungstreuer Schätzer mit Varianz gleich der in der Cramér-Rao-Ungleichung gegebenen unteren Schranke hat die kleinstmögliche Varianz aller erwartungstreuen Schätzer und ist in diesem Sinne ein “optimaler” Schätzer.

Der **mittlere quadratische Fehler** von $\hat{\tau}_n$ ist der Erwartungswert der quadrierten Abweichung von $\hat{\tau}_n(X)$ von $\tau(\theta)$ über Stichproben vom Umfang n .

Definition (Varianz und Standardfehler)

$X_1, \dots, X_n \sim p_\theta$ sei eine Stichprobe und $\hat{\tau}_n$ sei ein Schätzer von τ .

- Die *Varianz* von $\hat{\tau}_n$ ist definiert als

$$\mathbb{V}_\theta(\hat{\tau}_n) := \mathbb{E}_\theta \left((\hat{\tau}_n(X) - \mathbb{E}_\theta(\hat{\tau}_n(X)))^2 \right). \quad (10)$$

- Der *Standardfehler* von $\hat{\tau}_n$ ist definiert als

$$\text{SE}(\hat{\tau}_n) := \sqrt{\mathbb{V}_\theta(\hat{\tau}_n)} \quad (11)$$

Bemerkungen

- Die Varianz eines Schätzers $\hat{\tau}_n$ ist die Varianz der Zufallsvariable $\hat{\tau}_n(X)$.
- Der Standardfehler eines Schätzers $\hat{\tau}_n$ ist die Standardabweichung von $\hat{\tau}_n(X)$.

Theorem (Standardfehler des Stichprobenmittels)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{N}$. Dann ist der *Standardfehler des Stichprobenmittels* gegeben durch

$$\text{SE}(\bar{X}_n) = \frac{\mathbb{S}_\theta(X_0)}{\sqrt{n}}. \quad (12)$$

Der Standardfehler des Stichprobenmittels heißt auch *Standardfehler des Mittelwertes*.

Beweis

Per definitionem und mit $\mathbb{V}_\theta(\bar{X}_n) = \mathbb{V}_\theta(X_0)/n$, ergibt sich

$$\text{SE}(\bar{X}_n) = \sqrt{\mathbb{V}_\theta(\bar{X}_n)} = \sqrt{\frac{\mathbb{V}_\theta(X_0)}{n}} = \frac{\mathbb{S}_\theta(X_0)}{\sqrt{n}}. \quad (13)$$

Bemerkung

- Der Standardfehler des Mittelwerts beschreibt die Variabilität des Stichprobenmittels.
- Da $\mathbb{S}_\theta(X_0)$ unbekannt ist, ist auch $\text{SE}(\bar{X}_n)$ unbekannt.
- Ein verzerrter Schätzer für den Standardfehler des Stichprobenmittels ist gegeben durch

$$\hat{\text{SE}}(\bar{X}_n) = \frac{S_n}{\sqrt{n}}. \quad (14)$$

Beispiel (Standardfehler des Bernoulli Parameter ML Schätzers)

Es sei $X_1, \dots, X_n \sim \text{Bern}(\mu)$ und $\hat{\mu}_n^{\text{ML}}$ der ML Schätzer für μ . Dann ist

$$\text{SE}(\hat{\mu}_n^{\text{ML}}) = \sqrt{\frac{\mu(1-\mu)}{n}}. \quad (15)$$

Beweis

Es gilt

$$\text{SE}(\hat{\mu}_n^{\text{ML}}) = \sqrt{\mathbb{V}_{\mu}(\hat{\mu}_n^{\text{ML}})} = \sqrt{\mathbb{V}_{\mu}\left(\frac{1}{n} \sum_{i=1}^n X_i\right)} = \sqrt{\frac{1}{n^2} \sum_{i=1}^n \mathbb{V}_{\mu}(X_i)} = \sqrt{\frac{n\mu(1-\mu)}{n^2}} = \sqrt{\frac{\mu(1-\mu)}{n}}, \quad (16)$$

wobei die dritte Gleichung mit der Unabhängigkeit der $X_i, i = 1, \dots, n$ und die vierte Gleichung mit der Varianz $\mathbb{V}_{\mu}(X_0) = \mathbb{V}_{\mu}(X_i) = \mu(1-\mu), i = 1, \dots, n$ der Bernoulli Stichprobenvariablen folgt.

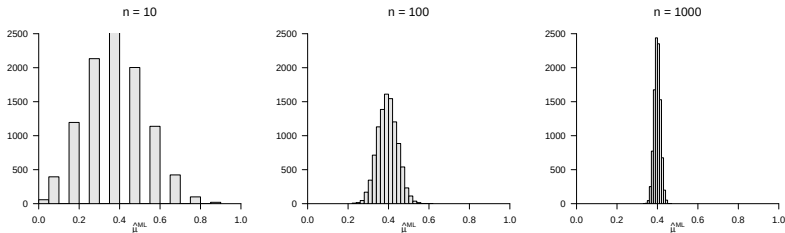
□

Bemerkung

- Ein Schätzer für den Standardfehler $\text{SE}(\hat{\mu}_n^{\text{ML}})$ ist

$$\hat{\text{SE}}(\hat{\mu}_n^{\text{ML}}) = \sqrt{\frac{\hat{\mu}_n^{\text{ML}}(1 - \hat{\mu}_n^{\text{ML}})}{n}}$$

Simulation ($X_1, \dots, X_n \sim \text{Bern}(\mu)$ mit $\mu = 0.4$)



- Die Varianz bzw. der Standardfehler von $\hat{\mu}^{ML}$ hängen von n ab.

Schätzereigenschaften bei endlichen Stichproben

Ein Schätzer $\hat{\tau}_n$ heißt **erwartungstreu**, wenn sein Erwartungswert dem wahren, aber unbekannten, Wert $\tau(\theta)$ für alle $\theta \in \Theta$ gleicht.

Die **Varianz** eines Schätzers $\hat{\tau}_n$ ist die Varianz der Zufallsvariable $\hat{\tau}_n(X)$; der **Standardfehler** eines Schätzers $\hat{\tau}_n$ ist die Standardabweichung der Zufallsvariable $\hat{\tau}_n(X)$.

Die **Cramér-Rao-Ungleichung** gibt eine untere Schranke für die Varianz erwartungstreuer Schätzer an. Ein erwartungstreuer Schätzer mit Varianz gleich der in der Cramér-Rao-Ungleichung gegebenen unteren Schranke hat die kleinstmögliche Varianz aller erwartungstreuen Schätzer und ist in diesem Sinne ein “optimaler” Schätzer.

Der **mittlere quadratische Fehler** von $\hat{\tau}_n$ ist der Erwartungswert der quadrierten Abweichung von $\hat{\tau}_n(X)$ von $\tau(\theta)$ über Stichproben vom Umfang n .

Vorbemerkungen zur Cramér-Rao-Ungleichung

- Je kleiner die Varianz eines Schätzers, desto besser. Weil aber Stichproben streuen, kann die Varianz von erwartungstreuen Schätzern nicht beliebig klein sein.
- Die **Cramér-Rao-Ungleichung** gibt eine untere Schranke für die Varianz erwartungstreuer Schätzer an. Ein erwartungstreuer Schätzer mit Varianz gleich der in der Cramér-Rao-Ungleichung gegebenen unteren Schranke hat die kleinstmögliche Varianz aller erwartungstreuer Schätzer und ist in diesem Sinne “optimal”.
- Die Cramér-Rao-Ungleichung basiert auf dem Begriff der **Fisher-Information**. Wir diskutieren deshalb zunächst die Begriffe der **Scorefunktion** und der darauf basierenden **Fisher-Information**
- Die vorgestellten Resultate gelten im Allgemeinen nur unter eine Reihe von Annahmen, den sogenannten **Fisher-Regularitätsbedingungen**

Fisher-Regularitätsbedingungen

1. Θ ist ein offenes Intervall, d.h. θ liegt nicht an einer Parameterraumgrenze.
2. Der Träger von p_θ hängt nicht von θ ab.
3. WMFs oder WDF mit unterschiedlichem $\theta \in \Theta$ sind unterschiedlich.
4. Die Likelihood-Funktion ist zweimal stetig differenzierbar.
5. Integration und Differentiation dürfen vertauscht werden.

Definition (Scorefunktion und Fisher-Information)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ mit eindimensionalem Parameter θ und ℓ_n sei die zugehörige Log-Likelihood-Funktion.

- Die erste Ableitung der Log-Likelihood-Funktion ℓ_n wird *Scorefunktion der Stichprobe* $X_1, \dots, X_n$ genannt und mit

$$S_n(\theta) := \frac{d}{d\theta} \ell_n(\theta). \quad (17)$$

bezeichnet. Für $n = 1$ schreiben wir $S(\theta) := S_1(\theta)$ und nennen $S(\theta)$ *Scorefunktion einer Zufallsvariable*.

- Die negative zweite Ableitung der Log-Likelihood-Funktion ℓ_n wird *Fisher-Information der Stichprobe* $X_1, \dots, X_n$ genannt und mit

$$I_n(\theta) := -\frac{d^2}{d\theta^2} \ell_n(\theta). \quad (18)$$

bezeichnet. Für $n = 1$ schreiben wir $I(\theta) := I_1(\theta)$ und nennen $I(\theta)$ die *Fisher-Information einer Zufallsvariable*.

Definition (Erwartete und beobachtete Fisher-Information)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ mit eindimensionalem Parameter θ , ℓ_n sei die zugehörige Log-Likelihood-Funktion und $\hat{\theta}_n^{\text{ML}}(X_1, \dots, X_n)$ sei ein ML-Schätzer von θ .

- Die *beobachtete Fisher-Information der Stichprobe* $X_1, \dots, X_n$ ist definiert als

$$I_n \left(\hat{\theta}_n^{\text{ML}} \right) := - \frac{d^2}{d\theta^2} \ell_n \left(\hat{\theta}_n^{\text{ML}} \right), \quad (19)$$

d.h. die beobachtete Fisher-Information der Stichprobe $X_1, \dots, X_n$ ist die Fisher-Information an der Stelle des ML-Schätzers $\hat{\theta}_n^{\text{ML}}(X_1, \dots, X_n)$.

- Die *erwartete Fisher-Information der Stichprobe* $X_1, \dots, X_n$ ist definiert als

$$J_n(\theta) := \mathbb{E}_\theta(I_n(\theta)). \quad (20)$$

Für $n = 1$ schreiben wir $J(\theta) := J_1(\theta)$ und nennen $J(\theta)$ die *erwartete Fisher-Information einer Zufallsvariable*.

Theorem (Additivität der Fisher-Information)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ mit eindimensionalem Parameter θ , ℓ_n sei die zugehörige Log-Likelihood-Funktion, und $I_n(\theta)$ und $J_n(\theta)$ seien die Fisher-Information und die erwartete Fisher-Information der Stichprobe $X_1, \dots, X_n$, respektive. Dann gilt

$$I_n(\theta) = nI_1(\theta) \text{ und } J_n(\theta) = nJ_1(\theta). \quad (21)$$

Bemerkungen

- Um $I_n(\theta)$ oder $J_n(\theta)$ zu berechnen, genügt es also $I(\theta)$ oder $J(\theta)$ zu berechnen.
- Die Additivität der beobachteten Fisher-Information ist in der Additivität von $I_n(\theta)$ implizit.

Beweis

Wir zeigen das Resultat für die erwartete Fisher-Information, das Resultat für die Fisher-Information ist dann implizit. Per definitionem und mit der Linearität von Ableitungen und Erwartungswerte gilt

$$\begin{aligned} J_n(\theta) &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \ell_n(\theta) \right) \\ &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \ln \left(\prod_{i=1}^n p_\theta(x_i) \right) \right) \\ &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \sum_{i=1}^n \ln p_\theta(x_i) \right) \\ &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \sum_{i=1}^n \ln p_\theta(x_1) \right) \tag{22} \\ &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \ell_1(\theta) n \right) \\ &= n \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \ell_1(\theta) \right) \\ &= n J_n(\theta). \end{aligned}$$

□

Theorem (Erwartungswert und Varianz der Scorefunktion)

Der Erwartungswert der Scorefunktion einer Zufallsvariable ist

$$\mathbb{E}_{\theta}(S(\theta)) = 0 \quad (23)$$

und die Varianz der Scorefunktion einer Zufallsvariable ist

$$\mathbb{V}_{\theta}(S(\theta)) = J(\theta). \quad (24)$$

Bemerkungen

- Der Erwartungswert der Ableitung der Log-Likelihood-Funktion ist Null.
- Die erwartete Fisher-Information ist gleich der Varianz der Scorefunktion.

Beweis

Wir betrachten nur den Fall, dass p_θ eine WDF ist und zeigen zunächst, dass $\mathbb{E}_\theta(S(\theta)) = 0$ ist:

$$\begin{aligned}\mathbb{E}_\theta(S(\theta)) &= \int S(\theta) p_\theta(x) dx \\&= \int \frac{d}{d\theta} \ell(\theta) p_\theta(x) dx \\&= \int \frac{d}{d\theta} \ln L(\theta) p_\theta(x) dx \\&= \int \frac{1}{L(\theta)} \frac{d}{d\theta} L(\theta) p_\theta(x) dx \\&= \int \frac{1}{p_\theta(x)} \frac{d}{d\theta} L(\theta) p_\theta(x) dx & (25) \\&= \int \frac{d}{d\theta} L(\theta) dx \\&= \frac{d}{d\theta} \int p_\theta(x) dx \\&= \frac{d}{d\theta} 1 \\&= 0.\end{aligned}$$

Mit der Definition der Varianz folgt dann sofort, dass $\mathbb{V}_\theta(S(\theta)) = \mathbb{E}_\theta(S(\theta)^2)$ ist.

Beweis

Als nächstes zeigen wir, dass $J(\theta) = \mathbb{E}_\theta (S(\theta)^2)$ und deshalb $\mathbb{V}_\theta(S(\theta)) = J(\theta)$ ist:

$$\begin{aligned} J(\theta) &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \ln L(\theta) \right) \\ &= \mathbb{E}_\theta \left(-\frac{d}{d\theta} \frac{\frac{d}{d\theta} L(\theta)}{L(\theta)} \right) \\ &= \mathbb{E}_\theta \left(-\frac{\frac{d^2}{d\theta^2} L(\theta) L(\theta) - \frac{d}{d\theta} L(\theta) \frac{d}{d\theta} L(\theta)}{L(\theta) L(\theta)} \right) \\ &= -\mathbb{E}_\theta \left(\frac{\frac{d^2}{d\theta^2} L(\theta)}{L(\theta)} \right) + \mathbb{E}_\theta \left(\frac{\left(\frac{d}{d\theta} L(\theta) \right)^2}{(L(\theta))^2} \right) \\ &= -\int \frac{\frac{d^2}{d\theta^2} L(\theta)}{L(\theta)} p_\theta(x) dx + \int \frac{\left(\frac{d}{d\theta} L(\theta) \right)^2}{(L(\theta))^2} p_\theta(x) dx \\ &= -\frac{d^2}{d\theta^2} \int p_\theta(x) dx + \int \left(\frac{1}{L(\theta)} \frac{d}{d\theta} L(\theta) \right)^2 p_\theta(x) dx \\ &= -\frac{d^2}{d\theta^2} 1 + \int \left(\frac{d}{d\theta} \ln L(\theta) \right)^2 p_\theta(x) dx \\ &= \mathbb{E}_\theta (S(\theta)^2) . \end{aligned} \tag{26}$$

□

Beispiel (Scorefunktion und Fisher-Information für den Bernoulli Parameter)

Es sei $X_1, \dots, X_n \sim \text{Bern}(\mu)$ mit $\mu \in]0, 1[$. Dann gilt:

- Die Scorefunktion der Stichprobe ist

$$S_n :]0, 1[\rightarrow \mathbb{R}, \mu \mapsto S_n(\mu) := \frac{1}{\mu} \sum_{i=1}^n x_i - \frac{1}{1-\mu} \left(n - \sum_{i=1}^n x_i \right). \quad (27)$$

- Die Fisher-Information der Stichprobe ist

$$I_n :]0, 1[\rightarrow \mathbb{R}, \mu \mapsto I_n(\mu) := \frac{x}{\mu^2} + \frac{(1-x)^2}{1-\mu}. \quad (28)$$

- Die beobachtete Fisher-Information der Stichprobe ist

$$I_n :]0, 1[\rightarrow \mathbb{R}, \hat{\mu}_n^{\text{ML}} \mapsto I_n(\hat{\mu}_n^{\text{ML}}) := \frac{x}{\hat{\mu}_n^{\text{ML}^2}} + \frac{(1-x)}{1-\hat{\mu}_n^{\text{ML}}}. \quad (29)$$

- Die erwartete Fisher-Information der Stichprobe ist

$$J_n :]0, 1[\rightarrow \mathbb{R}, \mu \mapsto J_n(\mu) := \frac{n}{\mu(1-\mu)}. \quad (30)$$

Beispiel (Scorefunktion und Fisher-Information für den Bernoulli Parameter)

Beweis

Die Scorefunktion wurde bereits in Vorlesung (7) im Kontext der Maximum-Likelihood-Schätzung von μ hergeleitet. Wir betrachten die Fisher-Information einer einzelnen Bernoulli Zufallsvariable X :

$$\begin{aligned} I(\mu) &:= -\frac{d^2}{d\mu^2} \ell_1(\mu) \\ &= -\frac{d^2}{d\mu^2} \ln p_\mu(x) \\ &= -\frac{d^2}{d\mu^2} (x \ln \mu + (1-x) \ln(1-\mu)) \\ &= -\frac{d}{d\mu} \left(\frac{d}{d\mu} (x \ln \mu + (1-x) \ln(1-\mu)) \right) \\ &= -\frac{d}{d\mu} \left(\frac{x}{\mu} + \frac{(1-x)}{1-\mu} \right) \\ &= -\left(-\frac{x}{\mu^2} - \frac{(1-x)^2}{(1-\mu)^2} \right) \\ &= \frac{x}{\mu^2} + \frac{(1-x)^2}{(1-\mu)^2}. \end{aligned} \tag{31}$$

Beispiel (Scorefunktion und Fisher-Information für den Bernoulli Parameter)

Beweis

Damit ergibt sich die erwartete Fisher-Information der Zufallsvariable X als

$$\begin{aligned} J(\mu) &= \mathbb{E}_{\mu}(I(\mu)) \\ &= \mathbb{E}_{\mu} \left(\frac{X}{\mu^2} + \frac{(1-X)^2}{1-\mu} \right) \\ &= \frac{\mathbb{E}_{\mu}(X)}{\mu^2} + \frac{(1 - \mathbb{E}_{\mu}(X))^2}{1-\mu} \\ &= \frac{\mu}{\mu^2} + \frac{(1-\mu)^2}{1-\mu} \\ &= \frac{1}{\mu(1-\mu)}. \end{aligned} \tag{32}$$

Mit der Additivitätseigenschaft der Fisher-Information und der Definition der beobachteten Fisher-Information ergibt sich dann sofort

$$I_n(\mu) = \frac{nx}{\mu^2} + \frac{n(1-x)^2}{1-\mu}, \text{ und } J_n(\mu) = \frac{n}{\mu(1-\mu)}. \tag{33}$$

□

Beispiel (Erwartungswert-Parameter der Normalverteilung)

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ und σ^2 als bekannt vorausgesetzt. Dann gilt:

- Die Scorefunktion der Stichprobe ist

$$S_n : \mathbb{R} \rightarrow \mathbb{R}, \mu \mapsto S_n(\mu) := \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu). \quad (34)$$

- Die Fisher-Information der Stichprobe ist

$$I_n : \mathbb{R} \rightarrow \mathbb{R}, \mu \mapsto I_n(\mu) := \frac{n}{\sigma^2}. \quad (35)$$

- Die beobachtete Fisher-Information der Stichprobe ist

$$I_n(\hat{\mu}_n^{\text{ML}}) = \frac{n}{\sigma^2}. \quad (36)$$

- Die erwartete Fisher-Information der Stichprobe ist

$$J_n : \mathbb{R} \rightarrow \mathbb{R}, \mu \mapsto J_n(\mu) := \frac{n}{\sigma^2}. \quad (37)$$

Beispiel (Erwartungswert-Parameter der Normalverteilung)

Beweis

Wir erinnern uns, dass die Log-Likelihood-Funktion der Stichprobe $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ bei bekanntem Varianzparameter σ^2 durch

$$\ell_n : \mathbb{R} \rightarrow \mathbb{R}, \mu \mapsto \ell_n(\mu) := -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \quad (38)$$

gegeben ist. Damit ergibt sich die Scorefunktion als

$$S_n(\mu) = \frac{d}{d\mu} \ell_n(\mu) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \quad (39)$$

Die Fisher-Information der Stichprobe $X_1, \dots, X_n$ ergibt sich als

$$I_n(\mu) = -\frac{d^2}{d\mu^2} \ell_n(\mu) = -\frac{d}{d\mu} S_n(\mu) = -\frac{1}{\sigma^2} \frac{d}{d\mu} \left(\sum_{i=1}^n x_i - n\mu \right) = \frac{n}{\sigma^2}. \quad (40)$$

Die beobachtete Fisher-Information ist die Fisher-Information an der Stelle des Maximum-Likelihood-Schätzers $\hat{\mu}_n^{\text{ML}}$. Die erwartete Fisher-Information schließlich ergibt sich als

$$J_n(\mu) = \mathbb{E}_\mu(I_n(\mu)) = \mathbb{E}_\mu \left(\frac{n}{\sigma^2} \right) = \frac{n}{\sigma^2}. \quad (41)$$

□

Beispiel (Varianz-Parameter der Normalverteilung)

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ und μ als bekannt vorausgesetzt. Dann gilt:

- die Scorefunktion ist

$$S_n : \mathbb{R}_{>0} \rightarrow \mathbb{R}, \sigma^2 \mapsto S_n(\sigma^2) := -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \quad (42)$$

- die Fisher-Information der Stichprobe ist

$$I_n : \mathbb{R}_{>0} \rightarrow \mathbb{R}, \sigma^2 \mapsto I_n(\sigma^2) := \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2 - \frac{n}{2\sigma^4} \quad (43)$$

- die beobachtete Fisher-Information der Stichprobe ist

$$I_n(\hat{\sigma}_n^{2\text{ML}}) = \frac{n}{2\hat{\sigma}_{\text{ML}}^4} \quad (44)$$

- die erwartete Fisher-Information der Stichprobe ist

$$J_n : \mathbb{R}_{>0} \rightarrow \mathbb{R}, \sigma^2 \mapsto J_n(\sigma^2) := \frac{n}{2\sigma^4}. \quad (45)$$

Beispiel (Varianz-Parameter der Normalverteilung)

Beweis

Wir erinnern uns, dass die Log-Likelihood-Funktion der Stichprobe $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ bei bekanntem Erwartungswert-Parameter μ durch

$$\ell_n : \mathbb{R}_{>0} \rightarrow \mathbb{R}, \sigma^2 \mapsto \ell_n(\sigma^2) := -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \quad (46)$$

gegeben ist. Die Scorefunktion ergibt sich also als

$$S_n(\sigma^2) = \frac{d}{d\sigma^2} \ell_n(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2. \quad (47)$$

Die Fisher-Information der Stichprobe $X_1, \dots, X_n$ ergibt sich als

$$\begin{aligned} I_n(\sigma^2) &= -\frac{d}{d\sigma^2} S_n(\sigma^2) = -\left(\frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2 \right) \\ &= \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2 - \frac{n}{2\sigma^4}. \end{aligned} \quad (48)$$

Beispiel (Varianz-Parameter der Normalverteilung)

Beweis

Die beobachtete Fisher-Information ist die Fisher-Information an der Stelle des Maximum-Likelihood-Schätzers $\hat{\sigma}_n^{2\text{ML}}$.

$$\begin{aligned} I_n(\hat{\sigma}_n^{2\text{ML}}) &= \frac{\sum_{i=1}^n (x_i - \mu)^2}{(\hat{\sigma}_n^{2\text{ML}})^3} - \frac{n}{2 (\hat{\sigma}_n^{2\text{ML}})^2} \\ &= \frac{\sum_{i=1}^n (x_i - \mu)^2}{\frac{1}{n^3} (\sum_{i=1}^n (x_i - \mu)^2)^3} - \frac{n}{2 (\hat{\sigma}_n^{2\text{ML}})^2} \\ &= \frac{1}{\frac{1}{n^3} (\sum_{i=1}^n (x_i - \mu)^2)^2} - \frac{n}{2 (\hat{\sigma}_n^{2\text{ML}})^2} \\ &= \frac{n}{(\hat{\sigma}_n^{2\text{ML}})^2} - \frac{n}{2 (\hat{\sigma}_n^{2\text{ML}})^2} \\ &= \frac{n}{2 (\hat{\sigma}_n^{2\text{ML}})^2} \\ &= \frac{n}{2 \hat{\sigma}_n^4 \text{ML}}. \end{aligned} \tag{49}$$

Beispiel (Varianz-Parameter der Normalverteilung)

Beweis

Die erwartete Fisher-Information ergibt sich als

$$\begin{aligned} J_n(\sigma^2) &= \mathbb{E}_{\sigma^2}(I_n(\sigma^2)) \\ &= \mathbb{E}_{\sigma^2} \left(\frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2 - \frac{n}{2\sigma^4} \right) \\ &= \frac{1}{\sigma^6} \sum_{i=1}^n \mathbb{E}_{\sigma^2} \left((x_i - \mu)^2 \right) - \frac{n}{2\sigma^4} \\ &= \frac{1}{\sigma^6} \sum_{i=1}^n \sigma^2 - \frac{n}{2\sigma^4} \\ &= \frac{n\sigma^2}{\sigma^6} - \frac{n}{2\sigma^4} \\ &= \frac{n}{\sigma^4} - \frac{n}{2\sigma^4} \\ &= \frac{n}{2\sigma^4}. \end{aligned} \tag{50}$$

Theorem (Cramér-Rao-Ungleichung)

$\mathcal{M}$ sei ein parametrisches statistisches Modell mit WMF oder WDF p_θ und $\hat{\tau}_n$ sei ein erwartungstreuer Schätzer von $\tau(\theta)$. Dann gilt

$$\mathbb{V}_\theta(\hat{\tau}_n) \geq \frac{\left(\frac{d}{d\theta}\tau(\theta)\right)^2}{J(\theta)}. \quad (51)$$

Im Speziellen gilt für $\tau(\theta) := \theta$ und somit $\hat{\tau}_n = \hat{\theta}_n$ und $\left(\frac{d}{d\theta}\tau(\theta)\right)^2 = 1$, dass

$$\mathbb{V}_\theta(\hat{\theta}_n) \geq \frac{1}{J(\theta)}. \quad (52)$$

Die rechte Seite obiger Ungleichungen heißt *Cramér-Rao-Schranke*.

Bemerkungen

- Die Varianz eines erwartungstreuen Schätzers $\hat{\theta}$ von θ ist größer oder gleich der reziproken erwarteten Fisher-Information $J(\theta)$.
- Wenn $\mathbb{V}_\theta(\hat{\theta}_n) = \frac{1}{J(\theta)}$ ist, ist die Varianz des Schätzers minimal.

Beweis

Wir halten zunächst fest, dass für die Zufallsvariablen $S(\theta)$ und $\hat{\tau}_n$ mit der Korrelationsungleichung und $\mathbb{V}_\theta(S(\theta)) = J(\theta)$ gilt, dass

$$\begin{aligned}\frac{\mathbb{C}_\theta(S_n(\theta), \hat{\tau}_n)^2}{\mathbb{V}_\theta(S(\theta))\mathbb{V}_\theta(\hat{\tau}_n)} &\leq 1 \\ \Leftrightarrow \mathbb{V}_\theta(\hat{\tau}_n) &\geq \frac{\mathbb{C}_\theta(S(\theta), \hat{\tau}_n)^2}{J(\theta)}.\end{aligned}\tag{53}$$

Mit dem Translationstheorem für Kovarianzen, $\mathbb{E}_\theta(S(\theta)) = 0$, und der Erwartungstreue von $\hat{\tau}_n$ ergibt sich dann

$$\mathbb{C}_\theta(S(\theta), \hat{\tau}_n) = \frac{d}{d\theta} \tau(\theta)\tag{54}$$

wie unten gezeigt wird. Also gilt

$$\mathbb{V}_\theta(\hat{\tau}_n) \geq \frac{\left(\frac{d}{d\theta} \tau(\theta)\right)^2}{J(\theta)}.\tag{55}$$

Beweis

Es bleibt zu zeigen, dass $\mathbb{C}_\theta(S(\theta), \hat{\tau}_n) = \frac{d}{d\theta} \tau(\theta)$. Dies ergibt sich mit

$$\begin{aligned}\mathbb{C}_\theta(S(\theta), \hat{\tau}_n) &= \mathbb{E}_\theta(S(\theta) \hat{\tau}_n) - \mathbb{E}_\theta(S(\theta)) \mathbb{E}_\theta(\hat{\tau}_n) \\&= \mathbb{E}_\theta(S(\theta) \hat{\tau}_n) \\&= \int S(\theta) \hat{\tau}_n p_\theta(x) dx \\&= \int \frac{d}{d\theta} \ln L(\theta) \hat{\tau}_n p_\theta(x) dx \\&= \int \frac{\frac{d}{d\theta} L(\theta)}{L(\theta)} \hat{\tau}_n p_\theta(x) dx \\&= \int \frac{\frac{d}{d\theta} L(\theta)}{p_\theta(x)} \hat{\tau}_n p_\theta(x) dx \\&= \int \frac{d}{d\theta} L(\theta) \hat{\tau}_n dx \\&= \frac{d}{d\theta} \int L(\theta) \hat{\tau}_n dx \\&= \frac{d}{d\theta} \int \hat{\tau}_n p_\theta(x) dx \\&= \frac{d}{d\theta} \mathbb{E}_\theta(\hat{\tau}_n) \\&= \frac{d}{d\theta} \tau(\theta)\end{aligned}\tag{56}$$



Schätzereigenschaften bei endlichen Stichproben

Ein Schätzer $\hat{\tau}_n$ heißt **erwartungstreu**, wenn sein Erwartungswert dem wahren, aber unbekannten, Wert $\tau(\theta)$ für alle $\theta \in \Theta$ gleicht.

Die **Varianz** eines Schätzers $\hat{\tau}_n$ ist die Varianz der Zufallsvariable $\hat{\tau}_n(X)$; der **Standardfehler** eines Schätzers $\hat{\tau}_n$ ist die Standardabweichung der Zufallsvariable $\hat{\tau}_n(X)$.

Die **Cramér-Rao-Ungleichung** gibt eine untere Schranke für die Varianz erwartungstreuer Schätzer an. Ein erwartungstreuer Schätzer mit Varianz gleich der in der Cramér-Rao-Ungleichung gegebenen unteren Schranke hat die kleinstmögliche Varianz aller erwartungstreuen Schätzer und ist in diesem Sinne ein “optimaler” Schätzer.

Der **mittlere quadratische Fehler** von $\hat{\tau}_n$ ist der Erwartungswert der quadrierten Abweichung von $\hat{\tau}_n(X)$ von $\tau(\theta)$ über Stichproben vom Umfang n .

Definition (Mittlerer quadratischer Fehler)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\tau}_n$ ein Schätzer für τ . Dann ist der *mittlere quadratische Fehler* (engl. *mean squared error*) von $\hat{\tau}_n$ definiert als

$$\text{MQF}(\hat{\tau}_n) := \mathbb{E}_\theta \left((\hat{\tau}_n(X) - \tau(\theta))^2 \right). \quad (57)$$

Bemerkungen

- Der MQF von $\hat{\tau}_n$ ist die erwartete quadrierte Abweichung von $\hat{\tau}_n(X)$ von $\tau(\theta)$.
- Die Varianz von $\hat{\tau}_n$ ist die erwartete quadrierte Abweichung von $\hat{\tau}_n$ von $\mathbb{E}_\theta(\hat{\tau}_n(X))$.
- $\mathbb{E}_\theta(\hat{\tau}_n(X))$ kann mit $\tau(\theta)$ übereinstimmen, muss es aber nicht.

Theorem (Zerlegung des mittleren quadratischen Fehlers)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$, $\hat{\tau}_n$ sei ein Schätzer für τ , und $\text{MQF}(\hat{\tau}_n)$ sei der mittlere quadratische Fehler von $\hat{\tau}_n$. Dann gilt

$$\text{MQF}(\hat{\tau}_n) = \text{B}(\hat{\tau}_n)^2 + \mathbb{V}_\theta(\hat{\tau}_n) \quad (58)$$

Bemerkungen

- $\text{MQF} = \text{Bias}^2 + \text{Varianz}$.
- Der MQF kann als Bias-Variance Abwägungskriterium benutzt werden.
- Kleine Verzerrungen können gegenüber großen Varianz präferiert werden.

Mittlerer quadratischer Fehler

Beweis

Zur Vereinfachung der Notation seien $\tau := \tau(\theta)$, $\hat{\tau}_n := \hat{\tau}_n(X)$ und $\bar{\tau}_n := \mathbb{E}_\theta(\hat{\tau}_n(X))$ zur Vereinfachung der Notation. Dann gilt:

$$\begin{aligned}\mathbb{E}_\theta \left((\hat{\tau}_n - \tau)^2 \right) &= \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n + \bar{\tau}_n - \tau)^2 \right) \\&= \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)^2 + 2(\hat{\tau}_n - \bar{\tau}_n)(\bar{\tau}_n - \tau) + (\bar{\tau}_n - \tau)^2 \right) \\&= \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)^2 \right) + 2\mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)(\bar{\tau}_n - \tau) \right) + \mathbb{E}_\theta \left((\bar{\tau}_n - \tau)^2 \right) \\&= \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)^2 \right) + 2\mathbb{E}_\theta \left(\hat{\tau}_n \bar{\tau}_n - \hat{\tau}_n \tau - \bar{\tau}_n \bar{\tau}_n + \bar{\tau}_n \tau \right) + \mathbb{E}_\theta \left((\bar{\tau}_n - \tau)^2 \right) \\&= \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)^2 \right) + 2(\bar{\tau}_n \bar{\tau}_n - \bar{\tau}_n \tau) + \mathbb{E}_\theta \left((\bar{\tau}_n - \tau)^2 \right) \\&= \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)^2 \right) + 0 + \mathbb{E}_\theta \left((\bar{\tau}_n - \tau)^2 \right) \\&= \mathbb{E}_\theta \left((\bar{\tau}_n - \tau)^2 \right) + \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)^2 \right) \\&= \mathbb{E}_\theta \left((\mathbb{E}_\theta(\hat{\tau}_n) - \tau)^2 \right) + \mathbb{E}_\theta \left((\hat{\tau}_n - \mathbb{E}_\theta(\hat{\tau}_n))^2 \right) \\&= (\mathbb{E}_\theta(\hat{\tau}_n) - \tau)^2 + \mathbb{V}_\theta(\hat{\tau}_n) \\&= \mathbf{B}(\hat{\tau}_n)^2 + \mathbb{V}_\theta(\hat{\tau}_n).\end{aligned}$$

□

Schätzereigenschaften

- Vorbemerkungen
- Schätzereigenschaften bei endlichen Stichproben
- **Asymptotische Schätzereigenschaften**
- Eigenschaften von Maximum-Likelihood Schätzern
- Selbstkontrollfragen

Motivation

- Dieser Abschnitt ist eine Kurzeinführung in die *Asymptotische Statistik (AS)*.
- Die AS befasst sich mit dem Verhalten von Statistiken bei großen Stichproben.
- Methoden der AS werden benutzt, um
 - qualitative Schätzereigenschaften zu studieren und
 - Schätzereigenschaften für große Stichprobengrößen zu approximieren.
- Moderne Stichproben sind üblicherweise groß.
- Die Methoden der AS sind also praktisch einsetzbar und gerechtfertigt.
- van der Vaart (1998) gibt eine ausführliche Einführung in die AS.

Asymptotische Schätzereigenschaften

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch erwartungstreu**, wenn der Erwartungswert von $\hat{\tau}_n$ für große Stichprobengrößen $n \rightarrow \infty$ gleich dem wahren, aber unbekannten, Wert $\tau(\theta)$ ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **konsistent**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Wahrscheinlichkeit dafür, dass $\hat{\tau}_n(X)$ vom wahren, aber unbekannten, Wert $\tau(\theta)$ **abweicht** beliebig klein wird.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch normalverteilt**, wenn für große Stichprobengrößen $n \rightarrow \infty$, die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung gegeben ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch effizient**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung mit Erwartungswertparameter $\tau(\theta)$ und Varianzparameter gleich der Cramér-Rao-Schranke gegeben ist.

Asymptotische Schätzereigenschaften

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch erwartungstreu**, wenn der Erwartungswert von $\hat{\tau}_n$ für große Stichprobengrößen $n \rightarrow \infty$ gleich dem wahren, aber unbekannten, Wert $\tau(\theta)$ ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **konsistent**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Wahrscheinlichkeit dafür, dass $\hat{\tau}_n(X)$ vom wahren, aber unbekannten, Wert $\tau(\theta)$ **abweicht** beliebig klein wird.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch normalverteilt**, wenn für große Stichprobengrößen $n \rightarrow \infty$, die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung gegeben ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch effizient**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung mit Erwartungswertparameter $\tau(\theta)$ und Varianzparameter gleich der Cramér-Rao-Schranke gegeben ist.

Definition (Asymptotische Erwartungstreue)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\tau}_n$ sei ein Schätzer für τ . $\hat{\tau}_n$ heißt *asymptotisch erwartungstreu*, wenn

$$\lim_{n \rightarrow \infty} \mathbb{E}_\theta(\hat{\tau}_n(X)) = \tau(\theta) \text{ für alle } \theta \in \Theta. \quad (59)$$

Bemerkungen

- Asymptotisch erwartungstreue Schätzer sind für “unendlich große” Stichproben erwartungstreu.
- Erwartungstreue Schätzer sind immer auch asymptotisch erwartungstreu.

Beispiel (Ein asymptotisch erwartungstreuer Schätzer)

Es sei $\mathcal{M}$ das Normalverteilungsmodell, also $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ und

$$\hat{\sigma}_n^{2\text{ML}} := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (60)$$

sei der ML Schätzer von σ^2 . Mit der Erwartungstreue der Stichprobenvarianz ergibt sich dann

$$\mathbb{E}_{\mu, \sigma^2} \left(\hat{\sigma}_n^{2\text{ML}} \right) = \mathbb{E}_{\mu, \sigma^2} \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right) = \frac{1}{n} \mathbb{E}_{\mu, \sigma^2} \left(\sum_{i=1}^n (X_i - \bar{X}_n)^2 \right) = \frac{n-1}{n} \sigma^2 \neq \sigma^2.$$

$\hat{\sigma}_n^{2\text{ML}}$ ist also ein verzerrter Schätzer von σ^2 . Allerdings gilt $(n-1)/n \rightarrow 1$ für $n \rightarrow \infty$, so dass

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mu, \sigma^2} \left(\hat{\sigma}_n^{2\text{ML}} \right) = \lim_{n \rightarrow \infty} \frac{n-1}{n} \sigma^2 = \sigma^2 \lim_{n \rightarrow \infty} \frac{n-1}{n} = \sigma^2. \quad (61)$$

Beispiel (Ein asymptotisch erwartungstreuer Schätzer)

Es sei $\mathcal{M}$ das Normalverteilungsmodell, also $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ und

$$\hat{\sigma}_n^{2\text{ML}} := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (60)$$

sei der ML Schätzer von σ^2 . Mit der Erwartungstreue der Stichprobenvarianz ergibt sich dann

$$\mathbb{E}_{\mu, \sigma^2} \left(\hat{\sigma}_n^{2\text{ML}} \right) = \mathbb{E}_{\mu, \sigma^2} \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right) = \frac{1}{n} \mathbb{E}_{\mu, \sigma^2} \left(\sum_{i=1}^n (X_i - \bar{X}_n)^2 \right) = \frac{n-1}{n} \sigma^2 \neq \sigma^2.$$

$\hat{\sigma}_n^{2\text{ML}}$ ist also ein verzerrter Schätzer von σ^2 . Allerdings gilt $(n-1)/n \rightarrow 1$ für $n \rightarrow \infty$, so dass

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mu, \sigma^2} \left(\hat{\sigma}_n^{2\text{ML}} \right) = \lim_{n \rightarrow \infty} \frac{n-1}{n} \sigma^2 = \sigma^2 \lim_{n \rightarrow \infty} \frac{n-1}{n} = \sigma^2. \quad (61)$$

Im Limit großer Stichprobenumfänge ist der ML Schätzer des Varianzparameters einer Normalverteilung also erwartungstreu.

Simulation $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit $\mu = 1, \sigma^2 = 2$

```
# Modellformulierung
mu      = 1                # wahrer, aber unbekannter, Erwartungswertparameter
sigsqr  = 2                # wahrer, aber unbekannter, Varianzparameter
n       = seq(1,100, by = 2) # Stichprobengroessen
ns      = 1e3              # Anzahl Simulation pro Stichprobengroesse
sigsqr_ml = matrix(        # \hat{\sigma^2}^{ML} Array
  rep(NaN, length(n)*ns),
  ncol = length(n))

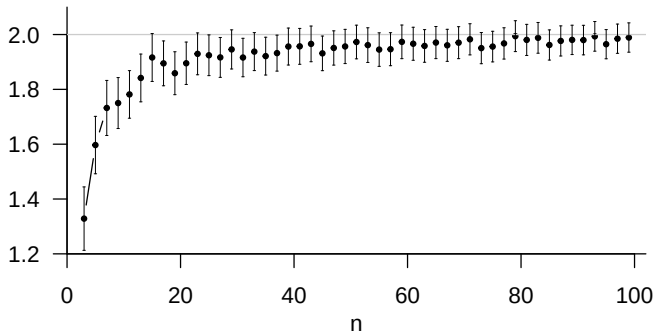
# Stichprobengroesseniterationen
for(i in seq_along(n)){

  # Simulationsiterationen
  for(s in 1:ns){
    # Stichprobenrealisation
    x = rnorm(n[i], mu, sqrt(sigsqr))

    # \hat{\sigma^2}^{ML}
    sigsqr_ml[s,i] = ((n[i]-1)/n[i])*var(x)
  }
}
E_sigsqr_ml = colMeans(sigsqr_ml) # Erwartungswertschaetzung
```

Asymptotische Erwartungstreue

Simulation $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit $\mu = 1, \sigma^2 = 2$



Asymptotische Schätzereigenschaften

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch erwartungstreu**, wenn der Erwartungswert von $\hat{\tau}_n$ für große Stichprobengrößen $n \rightarrow \infty$ gleich dem wahren, aber unbekannten, Wert $\tau(\theta)$ ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **konsistent**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Wahrscheinlichkeit dafür, dass $\hat{\tau}_n(X)$ vom wahren, aber unbekannten, Wert $\tau(\theta)$ **abweicht** beliebig klein wird.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch normalverteilt**, wenn für große Stichprobengrößen $n \rightarrow \infty$, die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung gegeben ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch effizient**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung mit Erwartungswertparameter $\tau(\theta)$ und Varianzparameter gleich der Cramér-Rao-Schranke gegeben ist.

Definition (Konsistenz)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\tau}_n$ sei ein Schätzer von τ . Eine Folge von Schätzern $\hat{\tau}_1, \hat{\tau}_2, \dots$ wird dann eine *konsistente Folge von Schätzern* genannt, wenn für jedes $\epsilon > 0$ und jedes $\theta \in \Theta$ gilt, dass

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta (|\hat{\tau}_n(X) - \tau(\theta)| \geq \epsilon) = 0.$$

Wenn $\hat{\tau}_1, \hat{\tau}_2, \dots$ eine konsistente Folge von Schätzern ist, dann heißt $\hat{\tau}_n$ *konsistenter Schätzer*.

Bemerkungen

- Für $n \rightarrow \infty$ wird die Wahrscheinlichkeit, dass $\hat{\tau}_n(X)$ beliebig nah bei $\tau(\theta)$ liegt, groß.
- Für $n \rightarrow \infty$ wird die Wahrscheinlichkeit, dass $\hat{\tau}_n(X)$ von $\tau(\theta)$ abweicht, klein.
- Diese Eigenschaften gelten für alle möglichen wahren, aber unbekannten, Parameterwerte.
- Die Konvergenz ist *Konvergenz in Wahrscheinlichkeit*.
- Konsistenz von Schätzern kann direkt oder mit Kriterien nachgewiesen werden.

Theorem (Mittlerer-Quadratischer-Fehler-Kriterium für Konsistenz)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\tau}_n$ sei ein Schätzer von τ . Wenn $\lim_{n \rightarrow \infty} \text{MQF}(\hat{\tau}_n) = 0$ gilt, dann ist $\hat{\tau}_n$ ein konsistenter Schätzer

Beweis

Mit der Chebychev-Ungleichung folgt, dass

$$\mathbb{P}_\theta (|\hat{\tau}_n(X) - \tau(\theta)| \geq \epsilon) \leq \frac{\mathbb{E}_\theta ((\hat{\tau}_n(X) - \tau(\theta))^2)}{\epsilon^2} \quad (62)$$

Grenzwertbildung ergibt dann

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta (|\hat{\tau}_n(X) - \tau(\theta)| \geq \epsilon) \leq \frac{1}{\epsilon^2} \lim_{n \rightarrow \infty} \mathbb{E}_\theta ((\hat{\tau}_n(X) - \tau(\theta))^2). \quad (63)$$

Wenn also $\lim_{n \rightarrow \infty} \mathbb{E}_\theta ((\hat{\tau}_n(X) - \tau(\theta))^2) = 0$ gilt, dann gilt mit $\mathbb{P}_\theta (|\hat{\tau}_n(X) - \tau(\theta)| \geq \epsilon) \geq 0$, dass

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta (|\hat{\tau}_n(X) - \tau(\theta)| \geq \epsilon) = 0. \quad (64)$$

Also ist $\hat{\tau}_n$ ein konsistenter Schätzer.



Theorem (Bias-Varianz-Kriterium für Konsistenz)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\tau}_n$ sei ein Schätzer von τ . Wenn

$$\lim_{n \rightarrow \infty} B(\hat{\tau}_n) = 0 \text{ und } \lim_{n \rightarrow \infty} \mathbb{V}_\theta(\hat{\tau}_n) = 0 \quad (65)$$

gilt, dann ist $\hat{\tau}_n$ ein konsistenter Schätzer

Beweis

Wenn $n \rightarrow \infty$, dann gilt $B(\hat{\tau}_n) \rightarrow 0$, also auch $B(\hat{\tau}_n)^2 \rightarrow 0$. Wenn für $n \rightarrow \infty$ sowohl $B(\hat{\tau}_n)^2 \rightarrow 0$ als auch $\mathbb{V}_\theta(\hat{\tau}_n) \rightarrow 0$, dann gilt auch $\lim_{n \rightarrow \infty} \text{MQF}(\hat{\tau}_n) = 0$. Also gilt mit dem MQF-Kriterium, dass $\hat{\tau}_n$ konsistent ist.

□

Beispiel (Konsistenz des Stichprobenmittels)

Für $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ folgt die Konsistenz des Stichprobenmittels als Schätzer für $\mathbb{E}(X_0)$ aus dem Bias-Varianz-Kriterium durch

$$B(\bar{X}_n) = 0 \text{ und } \lim_{n \rightarrow \infty} \mathbb{V}_\theta(\bar{X}_n) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{V}(X_0) = 0. \quad (66)$$

Konsistenz

Simulation $\bar{X}_n$ bei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit $\mu = 1, \sigma^2 = 2$

```
# Modellformulierung
mu      = 1                                # w.a.u. \mu Wert
sigsqr  = 2                                # w.a.u. \sigma^2 Wert
n       = seq(1, 1e3, by = 10)             # Stichprobengroesse n
eps     = c(0.15, 0.10, 0.05)             # \epsilon Werte
ne      = length(eps)                     # Anzahl \epsilon Werte
nn      = length(n)                       # Anzahl Stichprobengroessen
ns      = 1e3                             # Anzahl Simulationen
E       = array(rep(NaN, nn, ne, ns),      # Ereignisindikatorarray
               dim = c(nn, ne, ns))

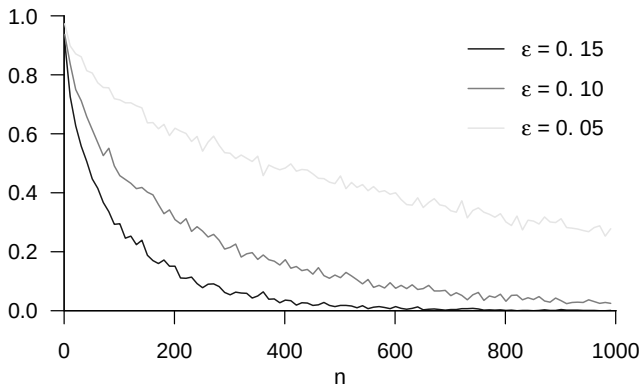
# Simulation
for(e in seq_along(eps)){                 # \epsilon Iterationen
  for(i in seq_along(n)){                 # n Iterationen
    for(s in 1:ns){                       # Simulationsiterationen

      # Stichprobenrealisationen
      x = rnorm(n[i], mu, sqrt(sigsqr))
      if(abs(mean(x) - mu) >= eps[e]){    # |X-bar - \mu| \ge \epsilon
        E[i, e, s] = 1
      } else {                             # |X-bar - \mu| < \epsilon
        E[i, e, s] = 0
      }
    }
  }
}

# Schaetzung von \mathbb{P}(|\hat{\tau}_n(X) - \tau(\theta)| \ge \epsilon)
P_hat   = apply(E, c(1, 2), mean)
```

Simulation $\bar{X}_n$ bei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit $\mu = 1, \sigma^2 = 2$

$$\hat{\mathbb{P}}(|\bar{X}_n - \mu| \geq \epsilon)$$



Asymptotische Schätzereigenschaften

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch erwartungstreu**, wenn der Erwartungswert von $\hat{\tau}_n$ für große Stichprobengrößen $n \rightarrow \infty$ gleich dem wahren, aber unbekannten, Wert $\tau(\theta)$ ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **konsistent**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Wahrscheinlichkeit dafür, dass $\hat{\tau}_n(X)$ vom wahren, aber unbekannten, Wert $\tau(\theta)$ **abweicht** beliebig klein wird.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch normalverteilt**, wenn für große Stichprobengrößen $n \rightarrow \infty$, die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung gegeben ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch effizient**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung mit Erwartungswertparameter $\tau(\theta)$ und Varianzparameter gleich der Cramér-Rao-Schranke gegeben ist.

Definition (Asymptotische Normalität)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\theta}_n$ sei ein Parameterschätzer für θ . Weiterhin sei $\tilde{\theta} \sim N(\mu, \sigma^2)$ eine normalverteilte Zufallsvariable mit Erwartungswertparameter μ und Varianzparameter σ^2 . Wenn $\hat{\theta}_n$ in Verteilung gegen $\tilde{\theta}$ konvergiert, dann heißt $\hat{\theta}_n$ *asymptotisch normalverteilt* und wir schreiben

$$\hat{\theta}_n \stackrel{a}{\sim} N(\mu, \sigma^2). \quad (67)$$

Bemerkung

- Konvergenz in Verteilung heißt $\lim_{n \rightarrow \infty} P_n(\hat{\theta}_n) = P(\tilde{\theta})$.

Asymptotische Schätzereigenschaften

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch erwartungstreu**, wenn der Erwartungswert von $\hat{\tau}_n$ für große Stichprobengrößen $n \rightarrow \infty$ gleich dem wahren, aber unbekannten, Wert $\tau(\theta)$ ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **konsistent**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Wahrscheinlichkeit dafür, dass $\hat{\tau}_n(X)$ vom wahren, aber unbekannten, Wert $\tau(\theta)$ **abweicht** beliebig klein wird.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch normalverteilt**, wenn für große Stichprobengrößen $n \rightarrow \infty$, die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung gegeben ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch effizient**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung mit Erwartungswertparameter $\tau(\theta)$ und Varianzparameter gleich der Cramér-Rao-Schranke gegeben ist,

Definition (Asymptotische Effizienz)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\theta}_n$ sei ein Parameterschätzer für θ . Weiterhin sei $J_n(\theta)$ die erwartete Fisher-Information der Stichprobe $X_1, \dots, X_n$. Wenn gilt, dass

$$\hat{\theta}_n \overset{a}{\sim} N(\theta, J_n(\theta)^{-1}), \quad (68)$$

dann heißt $\hat{\theta}_n$ *asymptotisch effizient*.

Bemerkungen

- Asymptotische Effizienz impliziert asymptotische Normalität.
- Asymptotische Effizienz impliziert asymptotische Erwartungstreue.
- Die Varianz der asymptotischen Verteilung heißt *asymptotische Varianz*.
- Die Varianz eines asymptotisch effizienten Schätzers ist gleich der Cramér-Rao-Schranke.
- Der Begriff der *Effizienz* wird in der Literatur nicht einheitlich verwendet.

Schätzereigenschaften

- Vorbemerkungen
- Schätzereigenschaften bei endlichen Stichproben
- Asymptotische Schätzereigenschaften
- **Eigenschaften von Maximum-Likelihood Schätzern**
- Selbstkontrollfragen

Theorem (Eigenschaften von ML Schätzern)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\theta}_n^{\text{ML}}$ sei ein ML Schätzer für θ . Dann gilt, dass $\hat{\theta}_n^{\text{ML}}$

- (1) nicht notwendigerweise erwartungstreu, aber
- (2) konsistent,
- (3) asymptotisch normalverteilt,
- (4) asymptotisch erwartungstreu, und
- (5) asymptotisch effizient ist.

Für einen Beweis verweisen wir auf Held and Sabanés Bové (2014, Abschnitt 3.4)

Simulation der asymptotische Effizienz des Bernoulli ML Parameterschätzers

- Es sei $X_1, \dots, X_n \sim \text{Bern}(\mu)$.
- Der ML Schätzer von μ ist $\hat{\mu}_n^{\text{ML}} = \frac{1}{n} \sum_{i=1}^n X_i$.
- Die erwartete Fisher-Information der Stichprobe $X_1, \dots, X_n$ ist

$$J_n(\mu) = \frac{n}{\mu(1-\mu)}. \quad (69)$$

- Also gilt

$$\hat{\mu}_n^{\text{ML}} \stackrel{a}{\sim} N\left(\mu, \frac{\mu(1-\mu)}{n}\right). \quad (70)$$

Eigenschaften von Maximum-Likelihood Schätzern

Simulation der asymptotische Effizienz des Bernoulli ML Parameterschätzers

```
# Modellformulierung
mu          = 0.4                                     # w.a.u. Parameterwert
n_all       = c(1e1,5e1,1e2)                         # Stichprobengroesse n
ns          = 1e4                                     # Anzahl der Simulationen
mu_hat_ML   = matrix(                                # ML Schaetzerarray
  rep(NA,
    length(n_all)*ns),
  nrow = length(n_all))

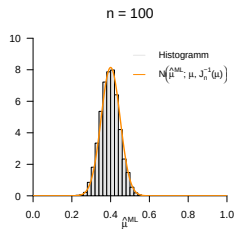
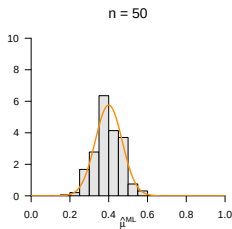
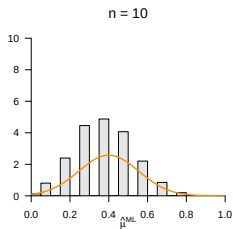
mu_hat_ML_r = 1e3                                     # ML Schaetzerraumaufloesung
mu_hat_ML_x = seq(0,1,len = mu_hat_ML_r)             # ML Schaetzerraum
mu_hat_ML_p = matrix(                                # ML WDF Array
  rep(NA, length(n_all)*mu_hat_ML_r),
  nrow = length(n_all))

# Stichprobengroesseniterationen
for(i in seq_along(n_all)){

  # Simulationsiterationen
  for(s in 1:ns){
    x          = rbinom(n_all[i],1,mu)               # Stichprobenrealisation
    mu_hat_ML[i,s] = mean(x)                         # ML Schaetzer
  }

  # WDF der asymptotischen Verteilung
  mu_hat_ML_p[i,] = dnorm(mu_hat_ML_x, mu, sqrt(mu*(1-mu)/n_all[i]))
}
```

Simulation der asymptotische Effizienz des Bernoulli ML Parameterschätzers



Schätzereigenschaften

- Vorbemerkungen
- Schätzereigenschaften bei endlichen Stichproben
- Asymptotische Schätzereigenschaften
- Eigenschaften von Maximum-Likelihood Schätzern
- **Selbstkontrollfragen**

1. Definieren und erläutern Sie den Begriff der Erwartungstreue eines Schätzers.
2. Definieren Sie die Begriffe der Varianz und des Standardfehlers eines Schätzers.
3. Definieren Sie den Begriff der Scorefunktion einer Stichprobe.
4. Definieren Sie den Begriff der Fisher-Information einer Stichprobe.
5. Definieren Sie den Begriff der erwarteten Fisher-Information einer Stichprobe.
6. Geben Sie das Theorem zur Cramér-Rao-Ungleichung wieder.
7. Erläutern Sie den Begriff der Cramér-Rao-Schranke.
8. Definieren Sie den Begriff des MQFs eines Schätzers.
9. Geben Sie das Theorem zur Zerlegung des MQFs wieder.
10. Erläutern Sie den Begriff des asymptotischen Erwartungstreue eines Schätzers
11. Erläutern Sie den Begriff der Konsistenz eines Schätzers.
12. Erläutern Sie den Begriff der asymptotischen Normalität eines Schätzers.
13. Erläutern Sie den Begriff der asymptotischen Effizienz eines Schätzers.
14. Nennen Sie fünf Eigenschaften eines ML Schätzers.

Literatur

Casella, G. and Berger, R. (2012). *Statistical Inference*. Duxbury.

DeGroot, M. H. and Schervish, M. J. (2012). *Probability and Statistics*. Addison-Wesley, Boston, 4th ed edition.

Georgii, H.-O. (2009). *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. DeGruyter-Lehrbuch. de Gruyter, Berlin, 4., überarb. und erw. aufl edition.

Held, L. and Sabanés Bové, D. (2014). *Applied Statistical Inference*. Springer Berlin Heidelberg, Berlin, Heidelberg.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, UK ; New York, NY, USA.

Wasserman, L. (2004). *All of Statistics*.



Inferenzstatistik

BSc Psychologie SoSe 2021

Prof. Dr. Dirk Ostwald

(9) Standardstatistiken und Konfidenzintervalle

Die hier gewählte Darstellung von Konfidenzintervallen basiert auf Held (2008, Abschnitte 3.2.2, 3.2.3, 3.4.4), Moeschlin (2000, Abschnitte 4.1-4.5) und Wasserman (2004, Abschnitte 6.3.2 und 9.7). Für einen verallgemeinerten Zugang zu Konfidenzintervallen, siehe Georgii (2009, Abschnitt 8.1).

(1) Parameterschätzung

Ziel der Parameterschätzung ist es, einen möglichst guten Tipp für den wahren, aber unbekannten, Parameterwert (oder eine Funktion dessen) abzugeben, typischerweise basierend auf der Beobachtung einer Realisierung von $X_1, \dots, X_n \sim p_\theta$.

(2) Konfidenzintervalle

Das Ziel der Bestimmung von Konfidenzintervallen ist es, basierend auf der Verteilung möglicher Parameterschätzwerte eine quantitative Aussage über die mit dem Schätzwert assoziierte Unsicherheit zu treffen.

(3) Hypothesentest

Das Ziel der Auswertung von Hypothesentests ist es, basierend auf der angenommenen Verteilung der Beobachtungen $X_1, \dots, X_n$ in einer möglichst sinnvollen Form zu entscheiden, ob der wahre, aber unbekannte Parameterwert, in einer von zwei sich gegenseitig ausschließenden Untermengen des Parameterraumes, welche man als Hypothesen bezeichnet, liegt.

Die Standardannahmen frequentistischer Inferenz

- $\mathcal{M}$ sei ein statistisches Modell mit $X_1, \dots, X_n \sim p_\theta$. Es wird angenommen, dass ein vorliegender Datensatz eine der möglichen Realisierungen von $X_1, \dots, X_n \sim p_\theta$ ist.
- Aus frequentistischer Sicht kann man das Experiment unendlich oft wiederholen und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. das Stichprobenmittel:

$$\text{Datensatz (1)} : x^{(1)} = (x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}) \text{ mit } \bar{x}_n^{(1)} = \frac{1}{n} \sum_{i=1}^n x_i^{(1)}$$

$$\text{Datensatz (2)} : x^{(2)} = (x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)}) \text{ mit } \bar{x}_n^{(2)} = \frac{1}{n} \sum_{i=1}^n x_i^{(2)}$$

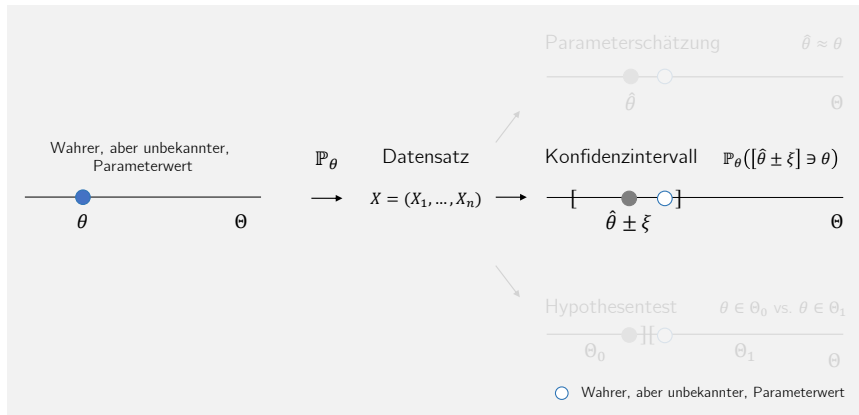
$$\text{Datensatz (3)} : x^{(3)} = (x_1^{(3)}, x_2^{(3)}, \dots, x_n^{(3)}) \text{ mit } \bar{x}_n^{(3)} = \frac{1}{n} \sum_{i=1}^n x_i^{(3)}$$

$$\text{Datensatz (4)} : x^{(4)} = (x_1^{(4)}, x_2^{(4)}, \dots, x_n^{(4)}) \text{ mit } \bar{x}_n^{(4)} = \frac{1}{n} \sum_{i=1}^n x_i^{(4)}$$

$$\text{Datensatz (5)} : x^{(5)} = \dots$$

- Um die Qualität **statistischer** Methoden zu beurteilen betrachtet die frequentistische Statistik deshalb die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken unter Annahme von $X_1, \dots, X_n \sim p_\theta$. Was zum Beispiel ist die Verteilung von $\bar{x}_n^{(1)}$, $\bar{x}_n^{(2)}$, $\bar{x}_n^{(3)}$, $\bar{x}_n^{(4)}$, ... also die Verteilung der Zufallsvariable $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$?
- Wenn eine statistische Methode im Sinne der frequentistischen Standardannahme gut ist, dann heißt das also, dass sie bei häufiger Anwendung "im Mittel" gut ist. Im Einzelfall, also bei nur einem Datensatz, kann sie auch recht schlecht sein.

Standardprobleme frequentistischer Inferenz



Standardstatistiken und Konfidenzintervalle

- Standardstatistiken
- Definition von Konfidenzintervallen
- Konfidenzintervallbeispiele
- Selbstkontrollfragen

Standardstatistiken und Konfidenzintervalle

- **Standardstatistiken**
- Definition von Konfidenzintervallen
- Konfidenzintervallbeispiele
- Selbstkontrollfragen

Z-Statistik

Theorem (Z-Statistik)

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ sei die Stichprobe eines Normalverteilungsmodells und $\bar{X}_n$ bezeichne das Stichprobenmittel. Dann ist die *Z-Statistik*

$$Z := \sqrt{n} \left(\frac{\bar{X}_n - \mu}{\sigma} \right) \quad (1)$$

eine standardnormalverteilte Zufallsvariable, es gilt also $Z \sim N(0, 1)$.

Bemerkungen

- Die WDF einer standardnormalverteilten Zufallsvariable bezeichnen wir mit ϕ .
- Ihre KVF und inverse KVF bezeichnen wir mit Φ und Φ^{-1} , respektive.

Z-Statistik

Beweis

Wir halten zunächst fest, dass mit der Mittelwertstransformation für unabhängig und identisch normalverteilte Zufallsvariablen gilt, dass $\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$ (cf. (6) Transformationen der Normalverteilungen). Die Verteilung der Z-Statistik folgt dann mit dem WDF Transformationstheorem bei linear-affinen Abbildungen (cf. ibid.) durch Transformation von $\bar{X}_n$ mit der Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R}, \bar{x}_n \mapsto f(\bar{x}_n) := \frac{\sqrt{n}}{\sigma} \bar{x}_n - \frac{\sqrt{n}}{\sigma} \mu \quad (2)$$

Für die WDF von Z gilt demnach

$$\begin{aligned} p_Z(z) &= \frac{\sigma}{\sqrt{n}} N\left(\frac{\sigma}{\sqrt{n}} \left(z + \frac{\sqrt{n}}{\sigma} \mu\right); \mu, \frac{\sigma^2}{n}\right) \\ &= \frac{\sigma}{\sqrt{n}} \frac{1}{\sqrt{2\pi \frac{\sigma^2}{n}}} \exp\left(-\frac{1}{2 \frac{\sigma^2}{n}} \left(\frac{\sigma}{\sqrt{n}} \left(z + \frac{\sqrt{n}}{\sigma} \mu\right) - \mu\right)^2\right) \\ &= \frac{\sigma}{\sqrt{n}} \frac{\sqrt{n}}{\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{n}{\sigma^2} \left(\frac{\sigma}{\sqrt{n}} z + \mu - \mu\right)^2\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2\right) \\ &= N(z; 0, 1) \end{aligned} \quad (3)$$

□

Z-Statistik

Modellformulierung

```
mu      = 10
sigsqr  = 4
n        = 12
ns       = 1e4
res      = 1e3
```

```
# w.a.u. Erwartungswertparameter
# wahrer bekannter Varianzparameter
# Stichprobengroesse
# Anzahl Stichprobenrealisierungen
# Ausgangsraumauflösung
```

analytische Definitionen und Resultate

```
x_i      = seq(3,17,len = res)
x_bar    = seq(3,17,len = res)
z        = seq(-4,4,len = res)
p_x_i    = dnorm(x_i,mu,sqrt(sigsqr))
p_x_bar  = dnorm(x_bar,mu,sqrt(sigsqr/n))
p_z      = dnorm(z,0,1)
```

```
# x_1 Raum
# x_bar Raum
# z Raum
# x_1 WDF
# x_bar WDF
# z WDF
```

Simulation

```
X_i      = rep(NaN,ns)
X_bar    = rep(NaN,ns)
Z         = rep(NaN,ns)
```

```
# X_1 Array
# \bar{X}_{12} Array
# Z Array
```

```
for(s in 1:ns){
  X      = rnorm(n,mu,sqrt(sigsqr))
  X_i[s] = X[1]
  X_bar[s] = mean(X)
```

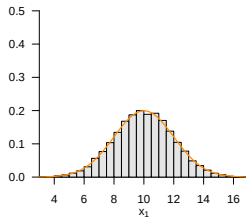
```
# Simulationsiterationen
# Stichprobenrealisierung
# X_i
# Stichprobenmittelrealisierung
```

Z-Statistik Realisation

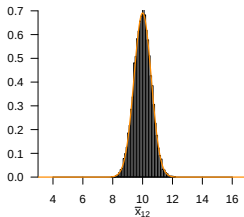
```
Z[s]     = sqrt(n)*((X_bar[s] - mu)/sqrt(sigsqr))
}
```

Z-Statistik

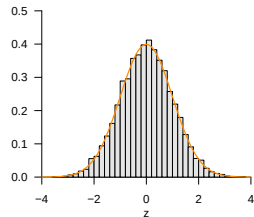
$X_i \sim N(10, 4)$, $i = 1, \dots, 12$



$\bar{X}_{12} \sim N(10, 1/3)$



$Z \sim N(0, 1)$



U -Statistik

Theorem (U -Statistik)

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ sei die Stichprobe eines Normalverteilungsmodells, und S_n^2 bezeichne die Stichprobenvarianz. Dann ist die U -Statistik

$$U := \frac{n-1}{\sigma^2} S_n^2 \quad (4)$$

eine χ^2 -verteilte Zufallsvariable mit $n - 1$ Freiheitsgraden

Bemerkungen

- Die WDF einer χ^2 -verteilten Zufallsvariable bezeichnen wir mit χ^2 .
- Ihre KVF und inverse KVF bezeichnen wir mit Ξ und Ξ^{-1} , respektive.
- Die U -Statistik ist eine normalisierte Stichprobenvarianz.
- Für einen Beweis verweisen wir auf Casella and Berger (2012, Abschnitt 5.3).

U-Statistik

Modellformulierung

```
mu      = 10
sigsqr  = 4
n       = 12
ns      = 1e4
res     = 1e3
```

```
# w.a.u. Erwartungswertparameter
# wahrer bekannter Varianzparameter
# Stichprobengroesse
# Anzahl Stichprobenrealisierungen
# Ausgangsraumauflösung
```

analytische Definitionen und Resultate

```
x_i     = seq(3,17,len = res)
x_bar   = seq(3,17,len = res)
u       = seq(0,30,len = res)
p_x_i   = dnorm(x_i,mu,sqrt(sigsqr))
p_x_bar = dnorm(x_bar,mu,sqrt(sigsqr/n))
p_u     = dchisq(u,n-1)
```

```
# x_1 Raum
# x_bar Raum
# normalized s_n^2 Raum
# x_1 WDF
# x_bar WDF
# u WDF
```

Simulation

```
X_i     = rep(NaN,ns)
X_bar   = rep(NaN,ns)
S_sqr   = rep(NaN,ns)
U       = rep(NaN,ns)
```

```
# X_1 Array
# \bar{X} Array
# S^2 Array
# U Array
```

```
for(s in 1:ns){
  X      = rnorm(n,mu,sqrt(sigsqr))
  X_i[s] = X[1]
  X_bar[s] = mean(X)
  S_sqr[s] = var(X)
```

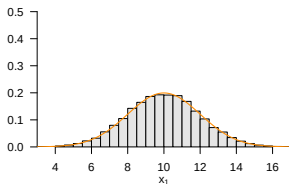
```
# Simulationsiterationen
# Stichprobenrealisierung
# X_i
# Stichprobenmittelrealisierung
# Stichprobenvarianzrealisierung
```

U-Statistik Realisation

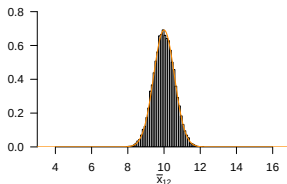
```
U[s]    = ((n-1)/sigsqr)*S_sqr[s]
}
```


U-Statistik

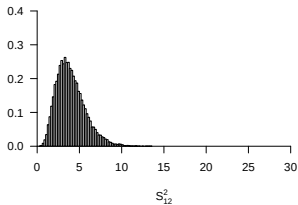
$$X_i \sim N(10, 4), i = 1, \dots, 12$$



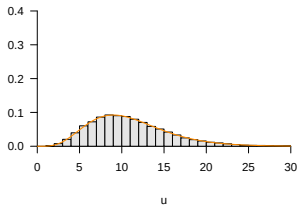
$$\bar{X}_{12} \sim N(10, 1/3)$$



$$S_{12}^2$$



$$U \sim \frac{11}{4} S_{12}^2$$



Theorem (T -Statistik)

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ sei die Stichprobe eines Normalverteilungsmodells, $\bar{X}_n$ bezeichne das Stichprobenmittel und S_n bezeichne die Stichprobenstandardabweichung. Dann ist die T -Statistik

$$T := \sqrt{n} \left(\frac{\bar{X}_n - \mu}{S_n} \right) \quad (5)$$

eine t -verteilte Zufallsvariable mit $n - 1$ Freiheitsgraden.

Bemerkungen

- Die WDF einer t -verteilten Zufallsvariable bezeichnen wir mit t .
- Ihre KVF und inverse KVF bezeichnen wir mit ψ und ψ^{-1} , respektive.

T -Statistik

Beweisskizze

Wir halten zunächst fest, dass

$$T = \sqrt{n} \left(\frac{\bar{X}_n - \mu}{S_n} \right) = \frac{\bar{X}_n - \mu}{S_n / \sqrt{n}} = \frac{\sigma}{\sigma} \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} = \frac{\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)}{S_n / \sigma} = \frac{\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)}{\sqrt{\frac{1}{\sigma^2} S_n^2}} \quad (6)$$

An der rechten Seite obiger Gleichung sehen wir also, dass die T -Statistik als Quotient der Zufallsvariablen

$$\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \text{ und } \sqrt{\frac{1}{\sigma^2} S_n^2} \quad (7)$$

geschrieben werden kann. Dabei hat die Zufallsvariable im Zähler die Form der Z -Statistik, kann also als

$$\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) =: Z \quad (8)$$

geschrieben werden. Die Zufallsvariable im Nenner dagegen hat die Form der Quadratwurzel einer mit $n - 1$ multiplizierten U -Statistik, kann also als

$$\sqrt{\frac{1}{\sigma^2} S_n^2} = \sqrt{\frac{1}{n-1} \frac{n-1}{\sigma^2} S_n^2} := \sqrt{\frac{1}{n-1} U} \quad (9)$$

geschrieben werden.

T -Statistik

Beweisskizze (fortgeführt)

Insgesamt kann die T Statistik also als

$$T = \frac{Z}{\sqrt{\frac{U}{n-1}}} \text{ mit } Z \sim N(0, 1) \text{ und } U \sim \chi^2(n-1) \quad (10)$$

geschrieben werden. Casella and Berger (2012, Abschnitt 5.3) zeigen, dass die Zufallsvariablen im Zähler und Nenner dieser Formulierung der T -Statistik unabhängig sind.

Es ergibt sich also, dass die T -Statistik **als** Quotient zweier unabhängiger Zufallsvariablen geschrieben werden kann, wobei die Zählerzufallsvariable eine standardnormalverteilte Zufallsvariable ist und die Nennerzufallsvariable die Quadratwurzel einer durch die Anzahl ihrer Freiheitsgrade **geteilten** χ^2 -**verteilten** Zufallsvariable ist.

Dann aber folgt mit dem T -Transformationstheorem (cf. (6) Transformationen der Normalverteilung), dass die T -Statistik eine t -verteilte Zufallsvariable mit Freiheitsgraden entsprechend den Freiheitsgraden der χ^2 -verteilten Zufallsvariable, hier also $n-1$, ist. Es folgt demnach $T \sim t(n-1)$.



T-Statistik

Modellformulierung

```
mu      = 10
sigsqr  = 4
n       = 12
ns      = 1e4
res     = 1e3
```

```
# w.a.u. Erwartungswertparameter
# wahrer bekannter Varianzparameter
# Stichprobengroesse
# Anzahl Stichprobenrealisierungen
# Ausgangsraumaufloesung
```

analytische Definitionen und Resultate

```
x_i     = seq(3,17,len = res)
x_bar   = seq(3,17,len = res)
u       = seq(0,30,len = res)
t       = seq(-4,4,len = res)
p_x_i   = dnorm(x_i,mu,sqrt(sigsqr))
p_x_bar = dnorm(x_bar,mu,sqrt(sigsqr/n))
p_u     = dchisq(u,n-1)
p_t     = dt(t,n-1)
```

```
# x_1 Raum
# x_bar Raum
# normalized s_{12}^2 Raum
# t Raum
# x_1 WDF
# x_bar WDF
# u WDF
# t WDF
```

Simulation

```
X_i     = rep(NaN,ns)
X_bar   = rep(NaN,ns)
U       = rep(NaN,ns)
Tee     = rep(NaN,ns)
for(s in 1:ns){
  X      = rnorm(n,mu,sqrt(sigsqr))
  X_i[s] = X[1]
  X_bar[s] = mean(X)
  U[s]    = (n-1)/(sigsqr)*var(X)
```

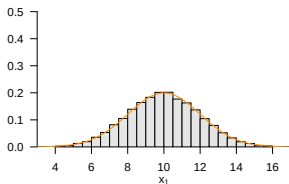
```
# X_1 Array
# \bar{X} Array
# U Array
# T Array
# Simulationsiterationen
# Stichprobenrealisierung
# X_i
# Stichprobenmittelrealisierung
# U Statistik Realisierung
```

T-Statistik Realisierung

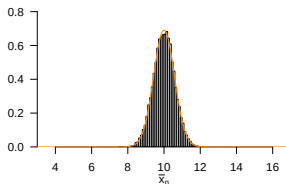
```
Tee[s] = sqrt(n)*((X_bar[s] - mu)/sqrt(var(X)))
}
```

T-Statistik

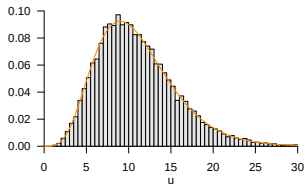
$$X_i \sim N(10, 4), i = 1, \dots, 12$$



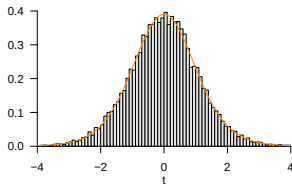
$$\bar{X}_n \sim N(10, 1/3)$$



$$U \sim \chi^2(11)$$



$$T \sim t(11)$$



Standardstatistiken und Konfidenzintervalle

- Standardstatistiken
- **Definition von Konfidenzintervallen**
- Konfidenzintervallbeispiele
- Selbstkontrollfragen

Definition (δ -Konfidenzintervall (δ -KI))

Es sei $X = X_1, \dots, X_n \sim p_\theta$ eine Stichprobe, $\delta \in]0, 1[$, und $b_l(X)$ und $b_u(X)$ seien zwei Statistiken. Dann ist ein δ -Konfidenzintervall ein Intervall der Form $C_n := [b_l, b_u]$, so dass

$$\mathbb{P}_\theta(C_n \ni \theta) = \mathbb{P}_\theta(b_l(X) \leq \theta \leq b_u(X)) = \delta \text{ für alle } \theta \in \Theta \text{ gilt.} \quad (11)$$

δ heißt das *Konfidenzniveau* oder die *Überdeckungswahrscheinlichkeit* des Konfidenzintervalls. Die Statistiken $b_l(X)$ und $b_u(X)$ sind die unteren und oberen Grenzen des Konfidenzintervalls.

Bemerkungen

- θ ist fest, nicht zufällig, und unbekannt.
- C_n ist ein zufälliges Intervall, weil $b_l(X)$ und $b_u(X)$ Zufallsvariablen sind.
- $C_n \ni \theta$ bedeutet $\theta \in C_n$, aber C_n ist zufällig und steht deshalb vorn (cf. $\mathbb{P}(X = x)$).
- Ein δ -Konfidenzintervall überdeckt den wahren Wert θ mit Wahrscheinlichkeit δ .
- Oft wird $\delta = 0.95$ gewählt, also 95%-Konfidenzintervalle betrachtet.

Zwei Interpretationen von δ -Konfidenzintervallen

- (1) Wird ein Experiment unter gleichen Umständen häufig wiederholt, so überdeckt das zugehörige δ -Konfidenzintervall den wahren, aber unbekannten Parameterwert in $\delta \cdot 100\%$ der Fälle. Technischer ausgedrückt, für unabhängig gezogene Stichproben einer Verteilung mit wahren, aber unbekannten, Parameter θ überdeckt ein entsprechendes δ -Konfidenzintervall θ in $\delta \cdot 100\%$ aller Fälle.
- (2) Man betrachte eine Folge von Experimenten mit wahren, aber unbekannten, Parametern $\theta_1, \theta_2, \dots$ und stelle sich vor δ -Konfidenzintervalle für eben jene Folge von wahren, aber unbekannten Parametern $\theta_1, \theta_2, \dots$ zu konstruieren. Dann überdecken $\delta \cdot 100\%$ der Konfidenzintervalle den wahren, aber unbekannten, Wert $\theta_i, i = 1, 2, \dots$.

Definition (Pivot)

Ein *Pivot* ist eine Funktion einer Stichprobe $X_1, \dots, X_n$ (also eine Statistik) und eines wahren, aber unbekannten, Parameters θ , deren Verteilung nicht von θ abhängt.

Bemerkungen

- Ein Pivot ist eine Statistik, die vom wahren, aber unbekannten Parameter θ abhängt, deren Verteilung aber nicht von vom wahren, aber unbekannten Parameter θ abhängt.
- Mithilfe von Pivots kann man Konfidenzintervalle für alle Werte von θ konstruieren.

Die typische Konstruktion eines Konfidenzintervalls umfasst

- (1) Definition eines statistischen Modells
- (2) Definition einer Statistik und Analyse ihrer Verteilung
- (3) Verifikation der Konfidenzbedingung
- (4) Definition des Konfidenzintervalls

Beispiele für die Konstruktion von Konfidenzintervallen

Erwartungswert einer Normalverteilung mit bekannter Varianz

Erwartungswert einer Normalverteilung mit unbekannter Varianz

Varianz einer Normalverteilung

Beispiele für die Konstruktion von Konfidenzintervallen

Erwartungswert einer Normalverteilung mit bekannter Varianz

Erwartungswert einer Normalverteilung mit unbekannter Varianz

Varianz einer Normalverteilung

Erwartungswert einer Normalverteilung mit bekannter Varianz

(1) Statistisches Modell

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ eine Stichprobe, wobei μ unbekannt sei und $\sigma^2 > 0$ bekannt sei. Wir entwickeln ein δ -Konfidenzintervall für μ .

(2) Statistik und ihre Verteilung

Wir betrachten die Statistik

$$Z := \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu). \quad (12)$$

Wir haben bereits gesehen, dass $Z \sim N(0, 1)$. Z ist eine Funktion der Stichprobe $X_1, \dots, X_n$ (via $\bar{X}_n$) und μ , während die Verteilung von Z nicht von μ abhängt. Die Statistik Z ist also ein exaktes Pivot.

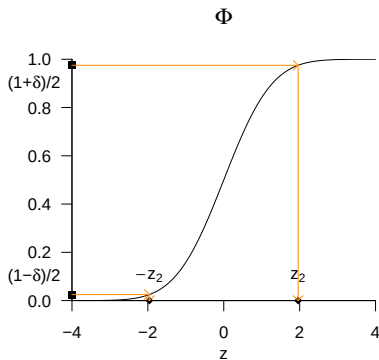
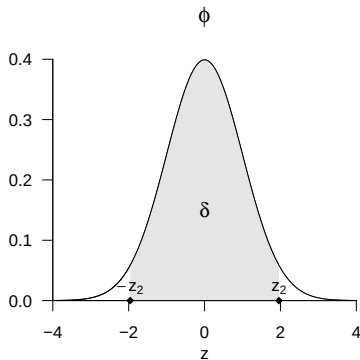
(3) Konfidenzbedingung

Für $\delta \in]0, 1[$, seien $z_1 := \Phi^{-1}\left(\frac{1-\delta}{2}\right)$ und $z_2 := \Phi^{-1}\left(\frac{1+\delta}{2}\right)$ die entsprechenden Quantile von $N(0, 1)$. Es gilt $(1 + \delta)/2 - (1 - \delta)/2 = \delta$. Zum Beispiel gilt für $\delta = 0.95$, dass $z_1 = \Phi^{-1}(0.025) = -1.96$ und $z_2 = \Phi^{-1}(0.975) = 1.96$. Weiterhin gilt mit der Symmetrie von $N(0, 1)$, $z_1 = -z_2$. Es gilt also per definitionem $\mathbb{P}(-z_2 \leq Z \leq z_2) = \delta$.

Erwartungswert einer Normalverteilung mit bekannter Varianz

(3) Konfidenzbedingung (fortgeführt)

$$\mathbb{P}(-z_2 \leq Z \leq z_2) = \delta$$



Erwartungswert einer Normalverteilung mit bekannter Varianz

(3) Konfidenzbedingung (fortgeführt)

Damit folgt

$$\begin{aligned}\delta &= \mathbb{P}(-z_2 \leq Z \leq z_2) \\&= \mathbb{P}\left(-z_2 \leq \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \leq z_2\right) \\&= \mathbb{P}\left(-\frac{\sigma}{\sqrt{n}}z_2 \leq \bar{X}_n - \mu \leq \frac{\sigma}{\sqrt{n}}z_2\right) \\&= \mathbb{P}\left(-\bar{X}_n - \frac{\sigma}{\sqrt{n}}z_2 \leq -\mu \leq -\bar{X}_n + \frac{\sigma}{\sqrt{n}}z_2\right) \\&= \mathbb{P}\left(\bar{X}_n + \frac{\sigma}{\sqrt{n}}z_2 \geq \mu \geq \bar{X}_n - \frac{\sigma}{\sqrt{n}}z_2\right) \\&= \mathbb{P}\left(\bar{X}_n - \frac{\sigma}{\sqrt{n}}z_2 \leq \mu \leq \bar{X}_n + \frac{\sigma}{\sqrt{n}}z_2\right) \\&= \mathbb{P}\left(\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}}z_2, \bar{X}_n + \frac{\sigma}{\sqrt{n}}z_2\right] \ni \mu\right).\end{aligned}\tag{13}$$

Erwartungswert einer Normalverteilung mit bekannter Varianz

(4) Definition des Konfidenzintervalls

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ eine Stichprobe bei bekanntem Varianzparameter σ^2 und unbekanntem Erwartungswertparameter μ , es sei $\delta \in]0, 1[$, und es sei $z_\delta := \Phi^{-1}\left(\frac{1+\delta}{2}\right)$. Definiere

$$C_n := \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_\delta, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_\delta \right]. \quad (14)$$

Dann gilt wie oben gezeigt, dass

$$\mathbb{P}(C_n \ni \mu) = \delta. \quad (15)$$

Damit ist C_n ein δ -Konfidenzintervall für μ . Weil $\bar{X}_n$ eine Zufallsvariable ist, ist C_n ein zufälliges Intervall.

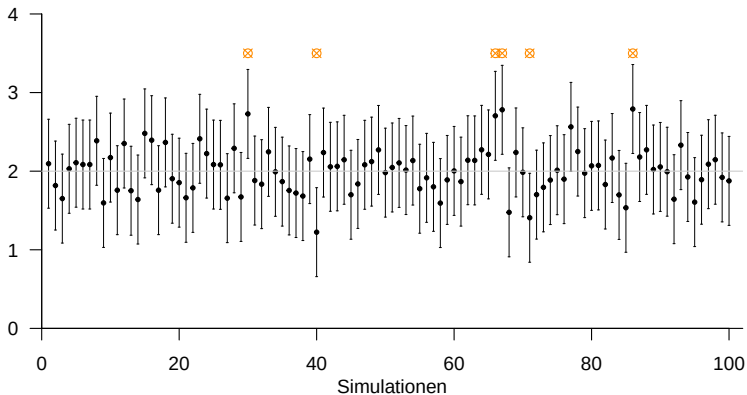
Erwartungswert einer Normalverteilung mit bekannter Varianz

```
# Modellformulierung
mu      = 2                                # w.a.u. Erwartungswertparameter
sigsqr  = 1                                # w.a.u. Varianzparameter
sigma   = sqrt(sigsqr)                     # w.a.u. St. Abweichungsparameter
n        = 12                              # Stichprobengroesse
delta   = 0.95                             # Konfidenzbedingung
phi_inv = qnorm((1+delta)/2)               # \Phi^{-1}((\delta + 1)/2)

# Simulation
ns       = 1e2                             # Anzahl Simulationen
X_bar    = rep(NaN, ns)                     # Stichprobenmittelarray
C        = matrix(rep(NaN, 2*ns), ncol = 2) # Konfidenzintervallarray
for(i in 1:ns){
  X      = rnorm(n, mu, sigma)              # Stichprobenrealisierung
  X_bar[i] = mean(X)                       # Stichprobenmittel
  C[i,1] = X_bar[i] - (sigma/sqrt(n))*phi_inv # untere KI Grenze
  C[i,2] = X_bar[i] + (sigma/sqrt(n))*phi_inv # obere KI Grenze
}
```

Erwartungswert einer Normalverteilung mit bekannter Varianz

$$\mu = 2, \sigma^2 = 2, n = 12, \delta = 0.95$$



Beispiele für die Konstruktion von Konfidenzintervallen

Erwartungswert einer Normalverteilung mit bekannter Varianz

Erwartungswert einer Normalverteilung mit unbekannter Varianz

Varianz einer Normalverteilung

Erwartungswert einer Normalverteilung mit unbekannter Varianz

1. Statistisches Modell

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ eine Stichprobe, wobei μ und $\sigma^2 > 0$ unbekannt seien. Wir entwickeln ein δ -Konfidenzintervall für μ .

2. Statistik und ihre Verteilung

Wir betrachten die Statistik

$$T := \frac{\sqrt{n}}{S_n}(\bar{X}_n - \mu) \text{ mit } S_n := \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2}. \quad (16)$$

Wir haben bereits gesehen, dass $T \sim t(n-1)$. T ist eine Funktion der Stichprobe $X_1, \dots, X_n$ (via $\bar{X}_n$ und S_n) während die Verteilung von T weder von μ noch von σ^2 abhängt. Die T Statistik ist also ein exaktes Pivot.

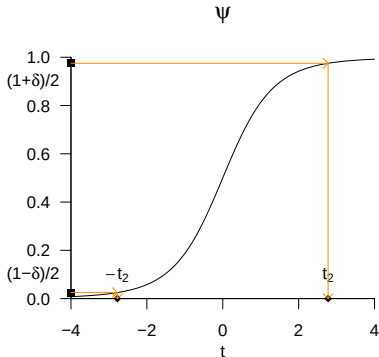
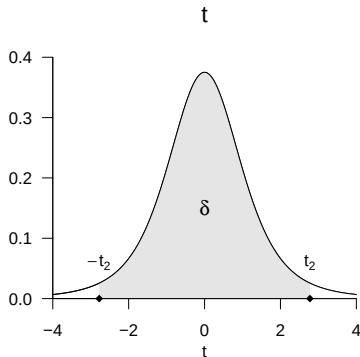
3. Konfidenzbedingung

Für $\delta \in]0, 1[$ seien $t_1 := \psi^{-1}\left(\frac{1-\delta}{2}; n-1\right)$ und $t_2 := \psi^{-1}\left(\frac{1+\delta}{2}; n-1\right)$ die entsprechenden Quantile von $t(n-1)$, so dass $(1+\delta)/2 - (1-\delta)/2 = \delta$ gilt. Zum Beispiel gilt für $n=5$ und $\delta=0.95$, $t_1 = \psi^{-1}(0.025; 4) = -2.57$ und $t_2 = \psi^{-1}(0.975; 4) = 2.57$. Weiterhin gilt mit der Symmetrie von $t(n-1)$, $t_1 = -t_2$. Es gilt also per definitionem $\mathbb{P}(-t_2 \leq T \leq t_2) = \delta$.

Erwartungswert einer Normalverteilung mit unbekannter Varianz

(3) Konfidenzbedingung (fortgeführt)

$$\mathbb{P}(-t_2 \leq T \leq t_2) = \delta, n = 12$$



Erwartungswert einer Normalverteilung mit unbekannter Varianz

3. Konfidenzbedingung (fortgeführt)

Damit folgt

$$\begin{aligned}\delta &= \mathbb{P}(-t_2 \leq T \leq t_2) \\&= \mathbb{P}\left(-t_2 \leq \frac{\sqrt{n}}{S_n}(\bar{X}_n - \mu) \leq t_2\right) \\&= \mathbb{P}\left(-\frac{S_n}{\sqrt{n}}t_2 \leq \bar{X}_n - \mu \leq \frac{S_n}{\sqrt{n}}t_2\right) \\&= \mathbb{P}\left(-\bar{X}_n - \frac{S_n}{\sqrt{n}}t_2 \leq -\mu \leq -\bar{X}_n + \frac{S_n}{\sqrt{n}}t_2\right) \\&= \mathbb{P}\left(\bar{X}_n + \frac{S_n}{\sqrt{n}}t_2 \geq \mu \geq \bar{X}_n - \frac{S_n}{\sqrt{n}}t_2\right) \\&= \mathbb{P}\left(\bar{X}_n - \frac{S_n}{\sqrt{n}}t_2 \leq \mu \leq \bar{X}_n + \frac{S_n}{\sqrt{n}}t_2\right) \\&= \mathbb{P}\left(\left[\bar{X}_n - \frac{S_n}{\sqrt{n}}t_2, \bar{X}_n + \frac{S_n}{\sqrt{n}}t_2\right] \ni \mu\right).\end{aligned}\tag{17}$$

Erwartungswert einer Normalverteilung mit unbekannter Varianz

4. Definition des Konfidenzintervalls

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit unbekannten Parametern μ und σ^2 , es sei $\delta \in]0, 1[$, und es sei $t_\delta := \psi^{-1}\left(\frac{1+\delta}{2}; n-1\right)$. Definiere

$$C_n := \left[\bar{X}_n - \frac{S_n}{\sqrt{n}} t_\delta, \bar{X}_n + \frac{S_n}{\sqrt{n}} t_\delta \right]. \quad (18)$$

Dann gilt wie oben gezeigt

$$\mathbb{P}(C_n \ni \mu) = \delta. \quad (19)$$

Damit ist C_n ein δ -Konfidenzintervall für μ . Weil $\bar{X}_n$ und S_n Zufallsvariablen sind, ist C_n ein zufälliges Intervall.

Erwartungswert einer Normalverteilung mit unbekannter Varianz

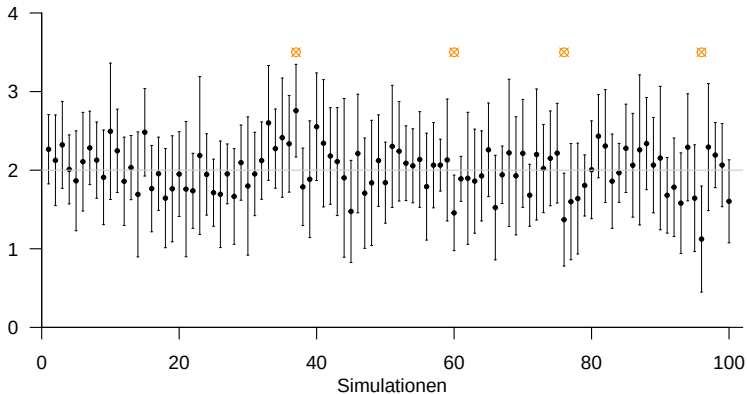
```
# Modellformulierung
mu      = 2                                # w.a.u. Erwartungswertparameter
sigsqr  = 1                                # w.a.u. Varianzparameter
sigma   = sqrt(sigsqr)                     # w.a.u. St.Abweichungsparameter
n       = 12                               # Stichprobengroesse
delta   = 0.95                             # Konfidenzbedingung
psi_inv = qt((1+delta)/2,n-1)              # \psi^{-1}((\delta + 1)/2, n-1)

# Simulation
ns      = 1e2                              # Anzahl Simulationen
X_bar   = rep(NaN, ns)                     # Stichprobenmittelarray
S       = rep(NaN, ns)                     # St.Abweichungsarray
C       = matrix(rep(NaN,2*ns), ncol = 2)  # Konfidenzintervallarray

for(i in 1:ns){
  X      = rnorm(n,mu,sigma)               # Stichprobenrealisierung
  X_bar[i] = mean(X)                       # Stichprobenmittel
  S[i]    = sd(X)                           # Stichprobenstandardabweichung
  C[i,1]  = X_bar[i] - (S[i]/sqrt(n))*psi_inv # untere KI Grenze
  C[i,2]  = X_bar[i] + (S[i]/sqrt(n))*psi_inv # obere KI Grenze
}
```

Erwartungswert einer Normalverteilung mit unbekannter Varianz

$$\mu = 2, \sigma^2 = 2, n = 12, \delta = 0.95$$



Beispiele für die Konstruktion von Konfidenzintervallen

Erwartungswert einer Normalverteilung mit bekannter Varianz

Erwartungswert einer Normalverteilung mit unbekannter Varianz

Varianz einer Normalverteilung

Varianz einer Normalverteilung

1. Statistisches Modell

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ eine Stichprobe mit unbekanntem Varianzparameter $\sigma^2 > 0$ und bekanntem oder unbekanntem Erwartungswertparameter μ . Wir entwickeln ein δ -Konfidenzintervall für σ^2 .

2. Statistik und ihre Verteilung

Wir betrachten die Statistik

$$U := \frac{n-1}{\sigma^2} S_n^2, \text{ wobei } S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2. \quad (20)$$

Wir haben bereits gesehen, dass $U \sim \chi^2(n-1)$. U ist eine Funktion von $X_1, \dots, X_n$ (via S_n^2) und σ^2 , während die Verteilung von U nicht von σ^2 abhängt. U ist also ein exaktes Pivot.

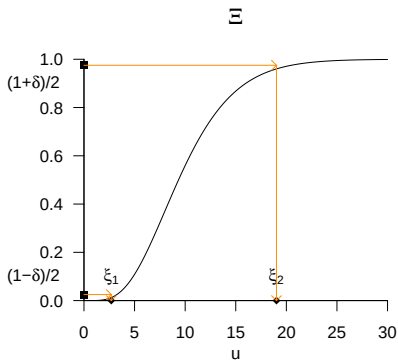
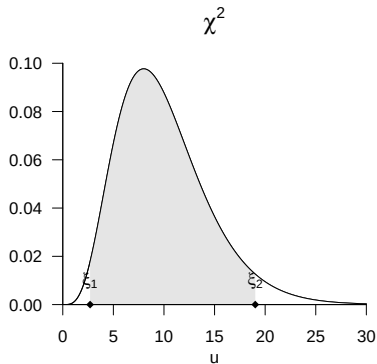
3. Konfidenzbedingung

Für $\delta \in]0, 1[$ seien $\xi_1 := \Xi^{2^{-1}}\left(\frac{1-\delta}{2}; n-1\right)$ und $\xi_2 := \Xi^{2^{-1}}\left(\frac{1+\delta}{2}; n-1\right)$ die entsprechenden Quantile von $\chi^2(n-1)$, so dass $(1+\delta)/2 - (1-\delta)/2 = \delta$ gilt. Zum Beispiel gilt für $n = 10$ und $\delta = 0.95$, $\xi_1 := \Xi^{2^{-1}}(0.025; 9) = 2.70$ und $\xi_2 := \Xi^{2^{-1}}(0.975; 9) = 19.0$. Es gilt also per definitionem $\mathbb{P}(\xi_1 \leq U \leq \xi_2) = \delta$.

Varianz einer Normalverteilung

(3) Konfidenzbedingung (fortgeführt)

$$\mathbb{P}(\xi_1 \leq U \leq \xi_2) = \delta, n = 12$$



Varianz einer Normalverteilung

3. Konfidenzbedingung (fortgeführt)

Damit folgt

$$\begin{aligned}\delta &= \mathbb{P}(\xi_1 \leq U \leq \xi_2) \\&= \mathbb{P}\left(\xi_1 \leq \frac{n-1}{\sigma^2} S_n^2 \leq \xi_2\right) \\&= \mathbb{P}\left(\xi_1^{-1} \geq \frac{\sigma^2}{(n-1)S_n^2} \geq \xi_2^{-1}\right) \\&= \mathbb{P}\left(\frac{(n-1)S_n^2}{\xi_1} \geq \sigma^2 \geq \frac{(n-1)S_n^2}{\xi_2}\right) \\&= \mathbb{P}\left(\frac{(n-1)S_n^2}{\xi_2} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{\xi_1}\right) \\&= \mathbb{P}\left(\left[\frac{(n-1)S_n^2}{\xi_2}, \frac{(n-1)S_n^2}{\xi_1}\right] \ni \sigma^2\right).\end{aligned}\tag{21}$$

Varianz einer Normalverteilung

4. Definition des Konfidenzintervalls

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ eine Stichprobe mit unbekanntem Varianzparameter σ^2 und bekanntem oder unbekanntem Erwartungswertparameter μ , und es seien weiterhin $\delta \in]0, 1[$, $\xi_1 := \Xi^{2^{-1}}\left(\frac{1-\delta}{2}; n-1\right)$ und $\xi_2 := \Xi^{2^{-1}}\left(\frac{1+\delta}{2}; n-1\right)$. Definiere

$$C_n := \left[\frac{(n-1)S_n^2}{\xi_2}, \frac{(n-1)S_n^2}{\xi_1} \right]. \quad (22)$$

Dann gilt wie oben gezeigt

$$\mathbb{P}(C_n \ni \sigma^2) = \delta. \quad (23)$$

Damit ist C_n ein δ -Konfidenzintervall für σ^2 . Weil S_n^2 eine Zufallsvariable ist, ist C_n ein zufälliges Intervall.

Varianz einer Normalverteilung

Modellformulierung

```
mu      = 2                                # w.a.u. Erwartungswertparameter
sigsqr  = 2                                # w.a.u. Varianzparameter
n       = 12                               # Stichprobengroesse
delta   = 0.95                             # Konfidenzbedingung
xi_1    = qchisq((1-delta)/2, n - 1)       # \(\chi^2_{(1-\delta)/2; n-1}\)
xi_2    = qchisq((1+delta)/2, n - 1)       # \(\chi^2_{(1+\delta)/2; n-1}\)
```

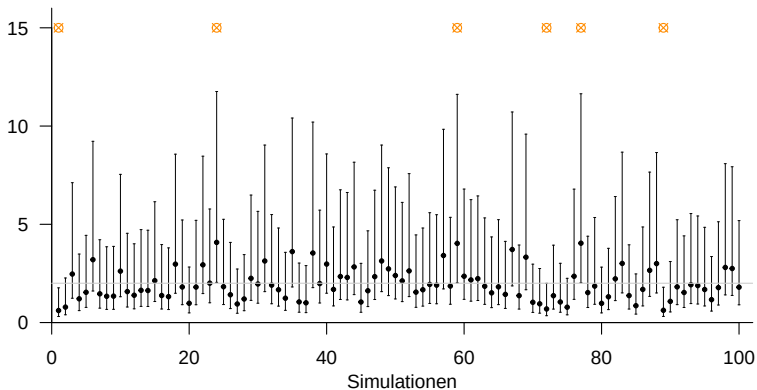
Simulation

```
ns      = 1e2                              # Anzahl Simulationen
X_bar   = rep(NaN, ns)                     # Stichprobenmittelarray
S2      = rep(NaN, ns)                     # Stichprobenvarianzarray
C       = matrix(rep(NaN, 2*ns), ncol = 2) # Konfidenzintervallarray

for(i in 1:ns){
  X      = rnorm(n, mu, sqrt(sigsqr))      # Stichprobenrealisierung
  S2[i]  = var(X)                          # Stichprobenvarianz
  C[i,1] = (n-1)*S2[i]/xi_2                # untere KI Grenze
  C[i,2] = (n-1)*S2[i]/xi_1                # obere KI Grenze
}
```


Varianz einer Normalverteilung

$$\mu = 2, \sigma^2 = 2, n = 12, \delta = 0.95$$



Anwendungsbeispiel

Beck Depressions-Inventar II (BDI II)

VP	BDI II Σ	Modellannahmen
1	11	Wir legen das Normalverteilungsmodell zugrunde, d.h. wir nehmen an, dass die BDI II Σ Werte Realisierungen von u.i.v. Zufallsvariablen $X_1, \dots, X_{12} \sim N(\mu, \sigma^2)$ sind. Dies entspricht der Annahme, dass sich der BDI II Σ Wert einer VP durch Addition einer normalverteilten Fehlervariable mit Erwartungswertparameter 0 und Varianzparameter σ^2 zu dem über VP identischen Wert μ ergibt.
2	14	
3	10	
4	9	
5	8	
6	17	
7	3	Fragestellungen (1) Was sind sinnvolle Tipps für die wahren, aber unbekannten, Parameterwerte μ und σ^2 ? (2) Wie hoch ist im frequentistischen Sinne die mit diesen Tipps assoziierte Unsicherheit? (3) Entscheiden wir uns sinnvollerweise für die Hypothese, dass $\mu \leq 8$ ist oder dass $\mu > 8$ ist?
8	20	
9	13	
10	3	
11	6	
12	2	

Anwendungsbeispiel

Beck Depressions-Inventar II (BDI II)

Fragestellungen

- (1) Was sind sinnvolle Tipps für die wahren, aber unbekannten, Parameterwerte μ und σ^2 ?

Im Sinne einer erwartungstreuen Schätzung sind sinnvolle Tipps durch das Stichprobenmittel und die Stichprobenvarianz, also durch

$$\bar{x}_{12} = 9.\bar{6} \text{ und } s_{12}^2 = 32.4\bar{2}$$

gegeben.

- (2) Wie hoch ist im frequentistischen Sinne die mit diesen Tipps assoziierte Unsicherheit?

Die 95%-Konfidenzintervalle dieser Tipps ergeben sich für μ zu

$$\left[\bar{x}_{12} - \frac{s_{12}}{\sqrt{12}} \psi^{-1} \left(\frac{1 + 0.95}{2}; 11 \right), \bar{x}_{12} + \frac{s_{12}}{\sqrt{12}} \psi^{-1} \left(\frac{1 + 0.95}{2}; 11 \right) \right] \approx [6.04, 13.28]$$

und für σ^2 zu

$$\left[11s_n^2/\Xi^{2^{-1}} \left(\frac{1 + 0.95}{2}; 11 \right), 11s_n^2/\Xi^{2^{-1}} \left(\frac{1 - 0.95}{2}; 11 \right) \right] \approx [16.27, 93.47]$$

Anwendungsbeispiel

Beck Depressions-Inventar II (BDI II)

Fragestellungen

- (1) Was sind sinnvolle Tipps für die wahren, aber unbekannten, Parameterwerte μ und σ^2 ?
- (2) Wie hoch ist im frequentistischen Sinne die mit diesen Tipps assoziierte Unsicherheit?

```
# Daten- und Analysedefinitionen
x      = c(11,14,10,9,8,17,3,20,13,3,6,2)      # BDI Datensatz
n      = length(x)                             # Stichprobengroesse
delta  = 0.95                                  # Konfidenzniveau

# Stichprobenstatistiken
x_bar  = mean(x)                               # Stichprobenmittel
s2     = var(x)                                # Stichprobenvarianz
s      = sd(x)                                 # Stichprobenstandardabweichung

# 95% Prozent KI fuer \mu
C_m_u  = x_bar - (s/sqrt(n))*qt((1+delta)/2, n - 1) # untere KI Grenze
C_m_o  = x_bar + (s/sqrt(n))*qt((1+delta)/2, n - 1) # obere KI Grenze

# 95% KI fuer \sigma^2
C_s_u  = (n-1)*s2/qchisq((1+delta)/2, n - 1)      # untere KI Grenze
C_s_o  = (n-1)*s2/qchisq((1-delta)/2, n - 1)      # obere KI Grenze
```

Alternative Visualisierung von Konfidenzintervallen (Georgii, 2009, Kapitel 8)

- Man beachte, dass hier die Notation $\vartheta = \theta$, $P_{\vartheta} = \mathbb{P}_{\theta}$, $\delta = 1 - \alpha$ benutzt wird.

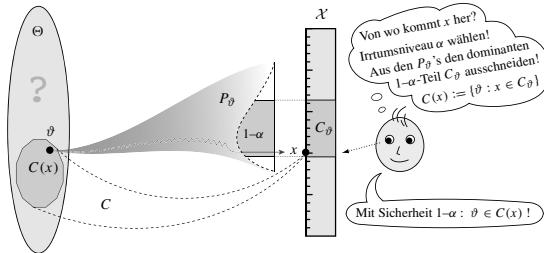


Abbildung 8.2: Schätzen mit Irrtumsangabe: Eine unbekannte Größe ϑ soll ermittelt werden; das tatsächlich vorliegende ϑ bestimmt ein Zufallsgesetz P_{ϑ} , welches im konkreten Fall zum Messwert x führt. Der Statistiker muss zunächst für jedes ϑ das Zufallsgesetz P_{ϑ} analysieren und dann, zu vorgegebener Irrtumswahrscheinlichkeit α , eine möglichst kleine Menge $C_{\vartheta} \subset \mathcal{X}$ bestimmen mit $P_{\vartheta}(C_{\vartheta}) \geq 1 - \alpha$. Damit kann er eine Menge $C(x) \subset \Theta$ angeben, in der ϑ mit Sicherheit von mindestens $1 - \alpha$ liegt – ganz egal, welches ϑ wirklich vorliegt.

Konfidenzintervalle

- Definition von Konfidenzintervallen
- Exakte Konfidenzintervalle
- Approximative Konfidenzintervalle
- **Selbstkontrollfragen**

1. Definieren Sie die Z -Statistik und geben Sie ihre Verteilung an.
2. Definieren Sie die U -Statistik und geben Sie ihre Verteilung an.
3. Definieren Sie die T -Statistik und geben Sie ihre Verteilung an.
4. Definieren Sie den Begriff des δ -Konfidenzintervalls (δ -KIs).
5. Geben Sie zwei Interpretationen eines δ -KIs.
6. Erläutern Sie den Begriff des Pivots.
7. Erläutern Sie die typischen Schritte zur Konstruktion eines δ -KIs.
8. Geben Sie das δ -KI für den Erwartungswert einer Normalverteilung bei bekannter Varianz an.
9. Geben Sie das δ -KI für den Erwartungswert einer Normalverteilung bei unbekannter Varianz an.
10. Geben Sie das δ -KI für den Varianzparameter einer Normalverteilung an.

Literatur

Casella, G. and Berger, R. (2012). *Statistical Inference*. Duxbury.

Georgii, H.-O. (2009). *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. De-Gruyter-Lehrbuch. de Gruyter, Berlin, 4., überarb. und erw. aufl edition.

Held, L. (2008). *Methoden Der Statistischen Inferenz*. Spektrum Akademischer Verlag, Heidelberg.

Moeschlin, O. (2000). *Angewandte Statistik*.

Wasserman, L. (2004). *All of Statistics*.



Inferenzstatistik

BSc Psychologie SoSe 2021

Prof. Dr. Dirk Ostwald

(10) Hypothesentests

Die Darstellung der basiert im Wesentlichen auf DeGroot and Schervish (2012), Section 9.1 und Czado and Schmidt (2011), Abschnitt 5.3.

Hypothesentests

- Vorbemerkungen
- Grundlegende Definitionen
- Z-Test
- p-Wert
- Konfidenzintervalle und Hypothesentests
- Selbstkontrollfragen

Hypothesentests

- **Vorbemerkungen**
- Grundlegende Definitionen
- Z-Test
- p-Wert
- Konfidenzintervalle und Hypothesentests
- Selbstkontrollfragen

Standardprobleme frequentistischer Inferenz

(1) Parameterschätzung

Ziel der Parameterschätzung ist es, einen möglichst guten Tipp für den wahren, aber unbekannten, Parameterwert (oder eine Funktion dessen) abzugeben, typischerweise basierend auf der Beobachtung einer Realisierung von $X_1, \dots, X_n \sim p_\theta$.

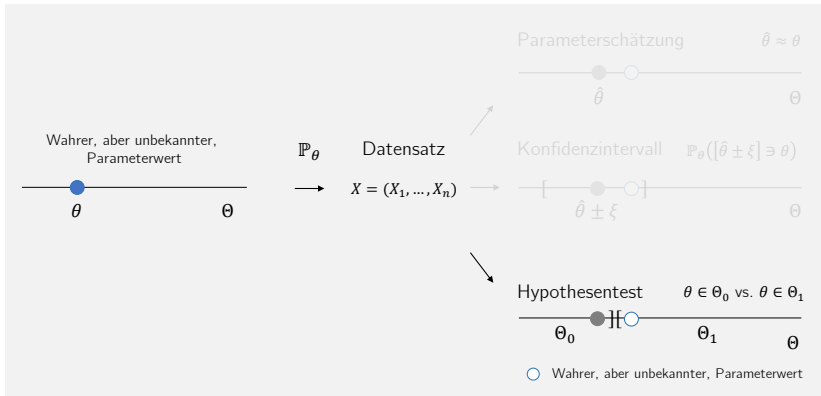
(2) Konfidenzintervalle

Das Ziel der Bestimmung von Konfidenzintervallen ist es, basierend auf der Verteilung möglicher Parameterschätzwerte eine quantitative Aussage über die mit dem Schätzwert assoziierte Unsicherheit zu treffen.

(3) Hypothesentest

Das Ziel der Auswertung von Hypothesentests ist es, basierend auf der angenommenen Verteilung der Beobachtungen $X_1, \dots, X_n$ in einer möglichst sinnvollen Form zu entscheiden, ob der wahre, aber unbekannte Parameterwert, in einer von zwei sich gegenseitig ausschließenden Untermengen des Parameterraumes, welche man als Hypothesen bezeichnet, liegt.

Standardprobleme frequentistischer Inferenz



Grundlegende Logik statistischer Hypothesentests

- Man hat einen Datensatz $x_1, \dots, x_n$ vorliegen und nimmt an, dass es sich dabei um die Realisation einer Stichprobe handelt, zum Beispiel von $X_1, \dots, X_n \sim N(\mu, \sigma^2)$.
- Man berechnet basierend auf dem Datensatz eine *Teststatistik*, zum Beispiel das anhand der Stichprobenvarianz und der Stichprobengröße normalisierte Stichprobenmittel $\sqrt{n}\bar{x}_n/s$.
- Man fragt sich, wie wahrscheinlich es wäre, den beobachteten oder einen extremeren Wert der Teststatistik unter der Annahme eines *Nullmodels* zu observieren. Dabei meint man mit *Nullmodell* intuitiv ein Verteilungsmodell bei dem kein "interessanter Effekt" vorliegt, also zum Beispiel $\mu = 0$ gilt. Die Wahrscheinlichkeit ist wie immer frequentistisch zu verstehen, d.h. als idealisierte relative Häufigkeit, wenn man viele Stichprobenrealisationen des Nullmodels generieren würde.
- Ist die betrachtete Wahrscheinlichkeit dafür, den beobachteten oder einen extremeren Wert der Teststatistik unter Annahme des Nullmodells zu observieren groß, so sagt man sich "Nunja, dann ist es wohl ganz plausibel, dass das Nullmodell die Daten generiert hat". Im Wissenschaftsjargon spricht man von einem "nicht-signifikanten Ergebnis".
- Ist die betrachtete Wahrscheinlichkeit dafür, den beobachteten oder einen extremeren Wert der Teststatistik unter Annahme des Nullmodells zu observieren dagegen klein, so sagt man sich "Aha, dann ist es wohl nicht so plausibel, dass das Nullmodell die Daten generiert hat". Im Wissenschaftsjargon spricht man von einem "signifikanten Ergebnis".
- Wie immer in der frequentistischen Statistik weiß man nach Durchführung dieser Prozedur nicht, ob im vorliegenden Fall nun wirklich das Nullmodell oder ein anderes Modell die Daten generiert hat, sondern man weiß nur, wie oft man bei dieser Prozedur im Mittel richtig oder falsch liegen würde, wenn alle Annahmen zuträfen und man diese Prozedur sehr oft wiederholen würde.

Hypothesentests

- Vorbemerkungen
- **Grundlegende Definitionen**
- Z-Test
- p-Wert
- Konfidenzintervalle und Hypothesentests
- Selbstkontrollfragen

Definition (Statistische Hypothesen und Testszenario)

$X_1, \dots, X_n \sim p_\theta$ sei eine Stichprobe mit WMF oder WDF p_θ , $\mathcal{X}$ sei der Ergebnisraum des Zufallsvektors $X := (X_1, \dots, X_n)$, und Θ sei der Parameterraum des zugrundeliegenden statistischen Modells. Weiterhin sei $\{\Theta_0, \Theta_1\}$ eine Partition des Parameterraumes, so dass $\Theta = \Theta_0 \cup \Theta_1$ und $\Theta_0 \cap \Theta_1 = \emptyset$ gelten. Eine *statistische Hypothese* ist dann eine Aussage über den wahren, aber unbekannten, Parameterwert θ in Hinblick auf die Untermengen Θ_0 und Θ_1 des Parameterraums. Speziell werden die Aussagen

- $\theta \in \Theta_0$ als *Nullhypothese* H_0
- $\theta \in \Theta_1$ als *Alternativhypothese* H_1

bezeichnet. Die Einheit aus Stichprobe, Ergebnisraum, Parameterraum, und Hypothesen wird im Folgenden als *Testszenario* bezeichnet.

Definition (Einfache und zusammengesetzte Hypothesen)

Für statistische Hypothesen $\Theta_i, i = 0, 1$ gilt:

- Enthält Θ_i nur ein einziges Element, so heißt Θ_i *einfach*.
- Enthält Θ_i mehr als ein Element, so heißt Θ_i *zusammengesetzt*.

Bemerkungen

- Die Nullhypothese $\Theta_0 = \{0\}$ ist ein Beispiel für eine einfache Hypothese.
- Bei einer einfachen Hypothese ist die Verteilung von X genau festgelegt.
- Bei einer zusammengesetzten Hypothese ist nur Verteilungsklasse von X festgelegt.

Definition (Einseitige und zweiseitige Hypothesen)

$\Theta := \mathbb{R}$ sei ein eindimensionaler Parameterraum und θ_0 sei ein Element von Θ . Dann werden zusammengesetzte Nullhypothesen der Form

$$\Theta_0 :=]-\infty, \theta_0] \text{ oder } \Theta_0 := [\theta_0, \infty[\quad (1)$$

einseitige Nullhypothesen genannt und auch in der Form

$$H_0 : \theta \leq \theta_0 \text{ oder } H_0 : \theta \geq \theta_0 \quad (2)$$

geschrieben. Die entsprechenden Alternativhypothesen haben dabei die Form

$$\Theta_1 :=]\theta_0, \infty[\text{ oder } \Theta_1 :=]-\infty, \theta_0[\text{ bzw. } H_1 : \theta > \theta_0 \text{ oder } H_1 : \theta < \theta_0. \quad (3)$$

Bei einer einfachen Nullhypothese der Form

$$\Theta_0 := \{\theta_0\} \text{ bzw. } H_0 : \theta = \theta_0 \quad (4)$$

wird die Alternativhypothese

$$\Theta_1 := \Theta \setminus \{\theta_0\} \text{ bzw. } H_1 : \theta \neq \theta_0 \Leftrightarrow \Theta_1 :=]-\infty, \theta_0[\cup]\theta_0, \infty[\quad (5)$$

zweiseitige Alternativhypothese genannt.

Definition (Test)

In einem Testszenario ist ein *Test* ϕ eine Abbildung aus dem Ergebnisraum $\mathcal{X}$ nach $\{0, 1\}$,

$$\phi : \mathcal{X} \rightarrow \{0, 1\}, x \mapsto \phi(x). \quad (6)$$

Dabei repräsentiert

- $\phi(x) = 0$ den Vorgang des Nichtablehnens der Nullhypothese.
- $\phi(x) = 1$ den Vorgang des Ablehnens der Nullhypothese.

Bemerkungen

- Weil die Stichprobe X ein Zufallsvektor ist, ist ein Test ϕ eine Zufallsvariable.

Definition (Standardtest)

Ein *Standardtest* ist definiert durch die Verkettung einer *Teststatistik*

$$\gamma : \mathcal{X} \rightarrow \mathbb{R} \quad (7)$$

und einer *Entscheidungsregel*

$$\delta : \mathbb{R} \rightarrow \{0, 1\}. \quad (8)$$

Ein Standardtest kann also geschrieben werden als

$$\phi := \delta \circ \gamma : \mathcal{X} \rightarrow \{0, 1\}. \quad (9)$$

- Weil die Stichprobe X ein Zufallsvektor ist, sind sowohl γ als auch δ Zufallsvariablen.
- Wir betrachten in der Folge nur Standardtests.

Definition (Kritischer Bereich)

Die Untermenge K des Ergebnisraums des Zufallsvektors $X := (X_1, \dots, X_n)$, für die ein Test den Wert 1 annimmt, heißt *kritischer Bereich* des Tests,

$$K := \{ \mathbf{x} \in \mathcal{X} \mid \phi(\mathbf{x}) = 1 \} \subset \mathcal{X}. \quad (10)$$

Bemerkungen

- Die Ereignisse $\phi(X) = 1$ und $X \in K$ sind äquivalent.
- Die Ereignisse $\phi(X) = 1$ und $X \in K$ haben die gleiche Wahrscheinlichkeit.

Definition (Ablehnungsbereich)

Die Untermenge A des Ergebnisraums einer Teststatistik, für die der Test den Wert 1 annimmt, heißt *Ablehnungsbereich* des Tests,

$$A := \{\gamma(\mathbf{x}) \in \mathbb{R} \mid \phi(\mathbf{x}) = 1\} \subset \mathbb{R}. \quad (11)$$

Bemerkungen

- Die Ereignisse $\phi(X) = 1$ und $\gamma(X) \in A$ sind äquivalent.
- Die Ereignisse $\phi(X) = 1$ und $\gamma(X) \in A$ haben die gleiche Wahrscheinlichkeit.

Definition (Kritischer Wert-basierte Tests)

Ein *kritischer Wert-basierter Test* ist ein Standardtest, bei dem die Entscheidungsregel δ von einem kritischen Wert $k \in \mathbb{R}$ abhängt. Speziell ist

- ein *einseitiger kritischer Wert-basierter Test* von der Form

$$\phi : \mathcal{X} \rightarrow \{0, 1\}, x \mapsto \phi(x) := 1_{\{\gamma(x) \geq k\}} = \begin{cases} 1 & \gamma(x) \geq k \\ 0 & \gamma(x) < k \end{cases} \quad (12)$$

- ein *zweiseitiger kritischer Wert-basierter Test* von der Form

$$\phi : \mathcal{X} \rightarrow \{0, 1\}, x \mapsto \phi(x) := 1_{\{|\gamma(x)| \geq k\}} = \begin{cases} 1 & |\gamma(x)| \geq k \\ 0 & |\gamma(x)| < k \end{cases} \quad (13)$$

- Wir betrachten in der Folge nur kritischer Wert-basierte Tests.

Definition (Richtige Testentscheidungen und Testfehler)

Das Nichtablehnen der Nullhypothese, wenn die Nullhypothese zutrifft, sowie das Ablehnen der Nullhypothese, wenn die Nullhypothese nicht zutrifft, werden *richtige Testentscheidungen* genannt. Es können weiterhin zwei Arten von Testfehlern auftreten: das Ablehnen der Nullhypothese, wenn die Nullhypothese zutrifft, heißt *Typ I Fehler*, das Nichtablehnen der Nullhypothese, wenn die Alternativhypothese zutrifft, heißt *Typ II Fehler*.

Die untenstehende Graphik gibt eine Übersicht.

	Testentscheidung	
	$\phi(X) = 0$	$\phi(X) = 1$
Wahrer Parameterwert	$\theta \in \Theta_0$	Richtige Entscheidung Typ I Fehler
	$\theta \in \Theta_1$	Typ II Fehler Richtige Entscheidung

Definition (Testgütefunktion)

Für einen Test ϕ ist die *Testgütefunktion* definiert als

$$q_\phi : \Theta \rightarrow [0, 1], \theta \mapsto q_\phi(\theta) := \mathbb{P}_\theta(\phi = 1). \quad (14)$$

Für $\theta \in \Theta_1$ heißt q_ϕ auch *Powerfunktion* oder *Trennschärfefunktion*.

Bemerkungen

- $\mathbb{P}_\theta$ bezeichnet die Verteilung von ϕ unter der Annahme $X_1, \dots, X_n \sim p_\theta$.
- Es gilt $\mathbb{P}_\theta(\phi = 1) = \mathbb{P}_\theta(X \in K) = \mathbb{P}_\theta(\gamma \in A)$
- Für jedes $\theta \in \Theta$ liefert q_ϕ die Wahrscheinlichkeit, dass H_0 durch ϕ abgelehnt wird.
- Bei Poweranalysen betrachtet man q_ϕ als Funktion aller Testszenario und Testparameter.
- Ändert sich ϕ , z.B. weil sich der kritische Wert von ϕ ändert, dann ändert sich $q_\phi(\theta)$.
- Im Idealfall hätte man einen Test ϕ mit

$$q_\phi(\theta) = \mathbb{P}_\theta(\phi = 1) = 0 \text{ für } \theta \in \Theta_0 \text{ und } q_\phi(\theta) = \mathbb{P}_\theta(\phi = 1) = 1 \text{ für } \theta \in \Theta_1. \quad (15)$$

Die Testentscheidung eines solchen ϕ wäre mit Wahrscheinlichkeit 1 richtig.

Intuition zur Testkonstruktion

- Im Idealfall hätte man einen Test ϕ mit

$$q_\phi(\theta) = \mathbb{P}_\theta(\phi = 1) = 0 \text{ für } \theta \in \Theta_0 \text{ und } q_\phi(\theta) = \mathbb{P}_\theta(\phi = 1) = 1 \text{ für } \theta \in \Theta_1. \quad (16)$$

⇒ Gut sind kleine Werte von q_ϕ für $\theta \in \Theta_0$ und große Werte von q_ϕ für $\theta \in \Theta_1$.

- Generell gibt es Abhängigkeiten zwischen den Werten von q_ϕ für $\theta \in \Theta_0$ und $\theta \in \Theta_1$:
 - Sei zum Beispiel ϕ_a der Test definiert durch $\phi_a(X) := 0$ für alle $X \in \mathcal{X}$, also der Test, der die Nullhypothese, unabhängig von den beobachteten Daten, *niemals ablehnt*. Für diesen Test gilt $q_{\phi_a}(\theta) = 0$ für $\theta \in \Theta_0$. Allerdings gilt für diesen Test auch $q_{\phi_a}(\theta) = 0$ für $\theta \in \Theta_1$.
 - Andersherum sei ϕ_b der Test definiert durch $\phi_b(X) := 1$ für alle $X \in \mathcal{X}$, also ein Test, der die Nullhypothese, unabhängig von den beobachteten Daten, *immer ablehnt*. Für diesen Test gilt $q_{\phi_b}(\theta) = 1$ für $\theta \in \Theta_1$. Allerdings gilt für diesen Test auch $q_{\phi_b}(\theta) = 1$ für $\theta \in \Theta_0$.
- In der Konstruktion eines Tests muss also eine angemessene Balance zwischen kleinen Werten von q_ϕ für $\theta \in \Theta_0$ und großen Werten von q_ϕ für $\theta \in \Theta_1$ gefunden werden.

Intuition zur Testkonstruktion

- Die populärste Methode, eine Balance zwischen kleinen Werten von q für $\theta \in \Theta_0$ und großen Werten von q für $\theta \in \Theta_1$ zu finden, ist in einem ersten Schritt ein $\alpha_0 \in [0, 1]$ zu wählen und sicher zu stellen, dass

$$q_\phi(\theta) \leq \alpha_0 \text{ für alle } \theta \in \Theta_0. \quad (17)$$

- Eine konventionelle Wahl für ein solches α_0 ist zum Beispiel $\alpha_0 := 0.05$.
- Unter allen Tests und statistischen Modellen, die Ungleichung (17) erfüllen, wird man dann einen Test oder ein statistisches Modell auswählen, so dass $q_\phi(\theta)$ für $\theta \in \Theta_1$ so groß wie möglich ist.
- Dieses Vorgehen ist nicht alternativlos, man kann zum Beispiel auch lineare Kombinationen verschiedener Fehlerwahrscheinlichkeiten minimieren. Es ist aber das in der Anwendung populärste Vorgehen. Wir werden uns deshalb in der Folge auf dieses Vorgehen beschränken.
- Das beschriebene Vorgehen motiviert die folgenden Definitionen der Begriffe des Level- α_0 -Tests, des Signifikanzlevels α_0 (oft auch als *Signifikanzniveau* bezeichnet) und des Testumfangs α (auch als *effektives Niveau* bezeichnet).

Definition (Level- α_0 -Test, Signifikanzlevel α_0 , Testumfang α)

q_ϕ sei die Testgütefunktion eines Tests ϕ und es sei $\alpha_0 \in [0, 1]$. Dann heißt ein Test ϕ , für den gilt, dass

$$q_\phi(\theta) \leq \alpha_0 \text{ für alle } \theta \in \Theta_0 \quad (18)$$

ein *Level- α_0 -Test* und man sagt, dass der Test das *Signifikanzlevel* α_0 hat. Die Zahl

$$\alpha := \max_{\theta \in \Theta_0} q_\phi(\theta) \in [0, 1] \quad (19)$$

heißt der *Testumfang* von ϕ .

Bemerkungen

- α ist die größtmögliche Wahrscheinlichkeit für einen Typ I Fehler.
- Ein Test ist dann, und nur dann, ein Level- α_0 -Test, wenn $\alpha \leq \alpha_0$ gilt.
- Bei einer einfachen Nullhypothese gilt für den Testumfang, dass $\alpha = q_\phi(\theta_0) = \mathbb{P}_{\theta_0}(\phi = 1)$.

Typ I Fehlerwahrscheinlichkeit vs. Testumfang vs. Signifikanzlevel

- Bei einfacher Θ_0 ist der Testumfang gleich der Wahrscheinlichkeit eines Typ I Fehlers

$$\alpha := \max_{\theta \in \Theta_0} q_\phi(\theta) = \max_{\theta \in \{\theta_0\}} q_\phi(\theta) = q_\phi(\theta_0) = \mathbb{P}_{\theta_0}(\phi = 1). \quad (20)$$

- Bei zusammengesetzter Θ_0 gibt es je nach Wert von $\theta \in \Theta_0$ verschiedene Wahrscheinlichkeiten für einen Typ I Fehler. Die größte dieser Wahrscheinlichkeiten ist der Testumfang

$$\alpha := \max_{\theta \in \Theta_0} q_\phi(\theta) = \max_{\theta \in \Theta_0} \mathbb{P}_\theta(\phi = 1). \quad (21)$$

- Ein Test hat Signifikanzlevel α_0 , wenn der Testumfang kleiner oder gleich α_0 ist.

$$\alpha = \max_{\theta \in \Theta_0} q_\phi(\theta) \leq \alpha_0 \quad (22)$$

- Ein Test, bei dem das Signifikanzlevel größer als der Testumfang ist, heißt *konservativ*.
- Ein Test, bei dem das Signifikanzlevel gleich dem Testumfang ist, heißt *exakt*.

Zur Wahl von Nullhypothese und Alternativhypothese

- Das Vorgehen in der Testkonstruktion zunächst durch die Wahl eines Signifikanzniveaus den Testumfang zu begrenzen und erst in einem zweiten Schritt dafür zu sorgen, dass die Wahrscheinlichkeit von $\phi = 1$ bei $\theta \in \Theta_1$ bei diesem Signifikanzniveau möglichst groß ist, induziert eine Asymmetrie in der Behandlung von Null- und Alternativhypothese. Implizit wichtet man mit diesem Vorgehen Typ I Fehler als schwerwiegender als Typ II Fehler.
- Dies wiederum impliziert eine mögliche Strategie zur Festlegung von Null- und Alternativhypothese: Die Nullhypothese ist die Hypothese, hinsichtlich deren assoziierter Testentscheidung man eher keinen Fehler machen möchte bzw. deren Fehlerwahrscheinlichkeit man primär kontrollieren möchte.
- In der wissenschaftlichen Anwendung ist es Standard, die falsche Konfirmation der eigenen Theorie als einen schwerwiegenderen Fehler als die falsche Ablehnung der eigenen Theorie zu werten.
- Die falsche Konfirmation der eigenen Theorie sollte also ein Typ I Fehler, das falsche Ablehnen der eigenen Theorie ein Typ II Fehler sein.
- Damit die falsche Konfirmation der eigenen Theorie einen Typ I Fehler, also das Ablehnen von H_0 bei Zutreffen von H_0 , darstellt, muss die eigene Theorie als Alternativhypothese aufgestellt werden. Die Alternativhypothese fälschlicherweise Abzulehnen wird damit ein Typ II Fehler.

Dirk Ostwalds Meinung zum Hypothesentesten in der Wissenschaft

- Frequentistisches Hypothesentesten ist als Entscheidungsproblem ohne klar und explizit definierte Entscheidungsnutzenfunktion formuliert und deshalb recht mühselig zu analysieren und zu studieren. Es gibt sehr viel zugänglichere Theorien zu Entscheidungen unter Unsicherheit (vgl. z.B. Pratt, Raiffa, and Schlaifer (1995), Puterman (2005), Ostwald, Starke, and Hertwig (2015), Horvath et al. (2020))
- Oberflächlich betrachtet liefern Hypothesentests einfache binäre Aussagen der Form “Die Hypothese (Theorie) ist gegeben die Evidenz abzulehnen oder zu akzeptieren.” Solche Aussagen sind im Entscheidungskontext hilfreich, denn es muss etwas passieren, also eine Entscheidung getroffen werden. In der Wissenschaft, also der menschlichen Kommunikationsstruktur über die Beschaffenheit der Welt, muss aber nichts final entschieden, sondern nur das Maß an Unsicherheit über den gerade vorherrschenden Theoriestand quantifiziert und kommuniziert werden. Generell sollten Fragestellungen der Grundlagenwissenschaften deshalb gerade nicht als Entscheidungsprobleme formuliert werden.
- Trotz landläufiger Meinung das Bayesianische Herangehensweisen wie Positive Predictive Values oder Bayes Factors hier irgendwie besser sind, ist dem nicht so, so lange die mit einer gewissen Modellpräferenz assoziierte Unsicherheit nicht klar mitkommuniziert wird.
- Und trotz alledem ist frequentistisches Hypothesentesten in der Wissenschaftscommunity weiterhin sehr populär (wenn auch vermutlich nicht immer ganz verstanden. . .) und sollte deshalb im Rahmen eines wissenschaftlichen Studiums wie der Psychologie intellektuell durchdrungen werden.

Hypothesentests

- Vorbemerkungen
- Grundlegende Definitionen
- **Z-Test**
- p-Wert
- Konfidenzintervalle und Hypothesentests
- Selbstkontrollfragen

- (1) Statistisches Modell, Hypothesen, Teststatistik, Test
- (2) Analyse der Testgütefunktion
- (3) Testumfangkontrolle
- (4) Analyse der Powerfunktion

Z-Test (1) Statistisches Modell, Hypothesen, Teststatistik, Test

Statistisches Modell

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ sei die Stichprobe eines **Normalverteilungsmodells** mit **unbekanntem Erwartungswertparameter** μ und **bekanntem Varianzparameter** $\sigma^2 > 0$. Der Parameter dieses Modells ist $\theta = \mu$ und der Parameterraum dieses Modells ist $\Theta = \mathbb{R}$.

Testhypothesen, Teststatistik, Test

Für ein $\mu_0 \in \mathbb{R}$ betrachten wir die einfache Nullhypothese und die zusammengesetzte Alternativhypothese

$$H_0 : \mu = \mu_0 \Leftrightarrow \Theta_0 := \{\mu_0\} \text{ und } H_1 : \mu \neq \mu_0 \Leftrightarrow \Theta_1 := \mathbb{R} \setminus \{\mu_0\}, \quad (23)$$

respektive. Weiterhin betrachten wir die *Z-Teststatistik*

$$Z := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{\sigma} \right) \quad (24)$$

und definieren den zweiseitigen kritischen Wert-basierten Test

$$\phi(X) := 1_{\{|Z| \geq k\}} = \begin{cases} 1 & |Z| \geq k \\ 0 & |Z| < k \end{cases}. \quad (25)$$

Theorem (Testgütefunktion bei zweiseitigem Z-Test, H_0 einfach)

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ die Stichprobe eines Normalverteilungsmodells mit unbekanntem Erwartungswertparameter μ und bekanntem Varianzparameter $\sigma^2 > 0$. Für $\Theta := \mathbb{R}$ und ein $\mu_0 \in \mathbb{R}$ sei ein Testszenario durch

$$H_0 : \mu = \mu_0 \Leftrightarrow \Theta_0 := \{\mu_0\} \text{ und } H_1 : \mu \neq \mu_0 \Leftrightarrow \Theta_1 := \mathbb{R} \setminus \{\mu_0\}, \quad (26)$$

gegeben. Weiterhin sei mit der Z-Teststatistik

$$Z := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{\sigma} \right) \quad (27)$$

in diesem Szenario der zweiseitige kritischer Wert-basierte Test

$$\phi(X) := 1_{\{|Z| \geq k\}} = \begin{cases} 1 & |Z| \geq k \\ 0 & |Z| < k \end{cases}. \quad (28)$$

definiert.

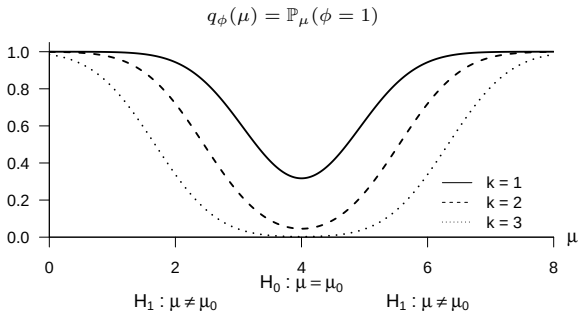
Dann ist die Testgütefunktion von ϕ gegeben durch

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \Phi \left(k - \frac{\sqrt{n}}{\sigma} (\mu - \mu_0) \right) + \Phi \left(-k - \frac{\sqrt{n}}{\sigma} (\mu - \mu_0) \right), \quad (29)$$

wobei Φ die KVF der Standardnormalverteilung bezeichnet.

Z-Test (2) Analyse der Testgütefunktion

Testgütefunktion q_ϕ für $\mu_0 = 4$, $n = 15$, $\sigma^2 = 9$ und drei zweiseitigen kritischen Wert-basierten Tests ϕ mit kritischen Werten $k = 1, 2, 3$ (vgl. DeGroot and Schervish (2012), Abbildung 9.1).



Z-Test (2) Analyse der Testgütefunktion

Beweis

Die Testgütefunktion des betrachteten Test im vorliegenden Testszenario ist definiert als

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := \mathbb{P}_\mu(\phi = 1). \quad (30)$$

Wir sind also an der Wahrscheinlichkeit $\mathbb{P}_\mu(\phi = 1)$ mit $\mu \in \mathbb{R}$ interessiert, d.h. an der Wahrscheinlichkeit dafür, dass der betrachtete Test den Wert 1 annimmt, wenn der wahre, aber unbekannte, Parameterwert μ ist. Wir hatten oben festgehalten, dass die Wahrscheinlichkeiten dafür, dass ein Test den Wert 1 annimmt und dafür, dass die zugehörige Teststatistik im Ablehnungsbereich des Tests liegt, gleich sind.

Um $\mathbb{P}_\mu(\phi = 1)$ auswerten zu können, benötigen wir also zunächst einmal die Verteilung der Teststatistik für $\mu \in \mathbb{R}$. Sei also $\mu \in \mathbb{R}$. Dann betrachten wir im vorliegenden Testszenario die Stichprobe $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit unbekanntem Erwartungswertparameter μ und bekanntem Varianzparameter $\sigma^2 > 0$ und die Z -Teststatistik $Z := \sqrt{n}((\bar{X}_n - \mu_0)/\sigma)$. Man beachte, dass hier sowohl $\mu = \mu_0$ als auch $\mu \neq \mu_0$ zugelassen sind. Wir befinden uns also nicht nur im in (9) Standardstatistiken und Konfidenzintervalle betrachteten Szenario $\mu = \mu_0$. Es gilt also zunächst, die Verteilung der Z -Teststatistik für ein generelles $\mu \in \mathbb{R}$ zu bestimmen.

Dazu folgen wir der in (9) Standardstatistiken und Konfidenzintervalle etablierten Herangehensweise: Wir halten zunächst fest, dass mit der Mittelwertstransformation für unabhängig und identisch normalverteilte Zufallsvariablen gilt, dass $\bar{X}_n \sim N(\mu, \sigma^2/n)$ (cf. (6) Transformationen der Normalverteilungen). Die Verteilung der Z -Teststatistik folgt dann mit dem WDF Transformationstheorem bei linear-affinen Abbildungen (cf. *ibid.*) durch Transformation von $\bar{X}_n$ mit der Funktion

$$f : \mathbb{R} \rightarrow \mathbb{R}, \bar{x}_n \mapsto f(\bar{x}_n) := \frac{\sqrt{n}}{\sigma} \bar{x}_n - \frac{\sqrt{n}}{\sigma} \mu_0 \quad (31)$$

Z-Test (2) Analyse der Testgütefunktion

3. Analyse der Testgütefunktion (fortgeführt)

Beweis (fortgeführt)

Wir erhalten

$$\begin{aligned}p_Z(z) &= \frac{\sigma}{\sqrt{n}} N\left(\frac{\sigma}{\sqrt{n}}\left(z + \frac{\sqrt{n}}{\sigma}\mu_0\right); \mu, \frac{\sigma^2}{n}\right) \\&= \frac{\sigma}{\sqrt{n}} \frac{1}{\sqrt{2\pi \frac{\sigma^2}{n}}} \exp\left(-\frac{1}{2 \frac{\sigma^2}{n}} \left(\frac{\sigma}{\sqrt{n}}\left(z + \frac{\sqrt{n}}{\sigma}\mu_0\right) - \mu\right)^2\right) \\&= \frac{\sigma}{\sqrt{n}} \frac{\sqrt{n}}{\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{n}{\sigma^2} \left(\frac{\sigma}{\sqrt{n}}z - (\mu - \mu_0)\right)^2\right) \\&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{n}{\sigma^2} \left(\frac{\sigma^2}{n}z^2 - 2\frac{\sigma}{\sqrt{n}}z(\mu - \mu_0) + (\mu - \mu_0)^2\right)\right) \\&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{n}{\sigma^2} \frac{\sigma^2}{n}z^2 - 2\frac{n}{\sigma^2} \frac{\sigma}{\sqrt{n}}z(\mu - \mu_0) + \frac{n}{\sigma^2}(\mu - \mu_0)^2\right)\right) \\&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(z^2 - 2z \frac{\sigma}{\sqrt{n}} \frac{\sqrt{n}}{\sigma}(\mu - \mu_0) + \left(\frac{\sqrt{n}}{\sigma}\right)^2(\mu - \mu_0)^2\right)\right) \\&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(z^2 - 2z \frac{\sqrt{n}}{\sigma}(\mu - \mu_0) + \left(\frac{\sqrt{n}}{\sigma}(\mu - \mu_0)\right)^2\right)\right) \\&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(z - \frac{\sqrt{n}}{\sigma}(\mu - \mu_0)\right)^2\right)\end{aligned}\tag{32}$$

Z-Test (2) Analyse der Testgütefunktion

Beweis (fortgeführt)

Wir haben also gefunden, dass für die Z -Teststatistik für beliebiges $\mu \in \mathbb{R}$ gilt, dass

$$Z \sim N \left(\frac{\sqrt{n}}{\sigma} (\mu - \mu_0), 1 \right). \quad (33)$$

Für den Fall, dass $\mu_0 = \mu$ ist, die Nullhypothese also zutrifft, folgt $Z \sim N(0, 1)$.

Als nächstes bestimmen wir den Ablehnungsbereich des betrachteten Tests. Per Definition gilt

$$A := \{Z \in \mathbb{R} | \phi(X) = 1\} \quad (34)$$

Es ergibt sich also

$$\begin{aligned} A &= \{Z \in \mathbb{R} | \phi(X) = 1\} \\ &= \{Z \in \mathbb{R} | |Z| \geq k\} \\ &= \left\{ Z \in \mathbb{R} \left| \begin{cases} Z \geq k, & Z \geq 0 \\ -Z \geq k, & Z < 0 \end{cases} \right. \right\} \\ &= \left\{ Z \in \mathbb{R} \left| \begin{cases} Z \geq k, & Z \geq 0 \\ Z \leq -k, & Z < 0 \end{cases} \right. \right\} \\ &= \{Z \in \mathbb{R} | Z < 0, Z \leq -k\} \cup \{Z \in \mathbb{R} | Z > 0, Z \geq k\} \\ &=] - \infty, -k] \cup]k, \infty [. \end{aligned} \quad (35)$$

Z-Test (2) Analyse der Testgütefunktion

Beweis (fortgeführt)

Mit diesem Ablehnungsbereich des betrachteten Tests ergibt sich dann

$$\begin{aligned}q_{\phi}(\mu) &= \mathbb{P}_{\mu}(\phi = 1) \\&= \mathbb{P}_{\mu}\left(Z \in]-\infty, -k] \cup]k, \infty [\right) \\&= \mathbb{P}_{\mu}(Z \in]-\infty, -k]) + \mathbb{P}_{\mu}(Z \in [k, \infty[) \\&= \mathbb{P}_{\mu}(Z \leq -k) + \mathbb{P}_{\mu}(Z \geq k) \\&= \mathbb{P}_{\mu}(Z \leq -k) + (1 - \mathbb{P}_{\mu}(Z \leq k)) \\&= 1 - \mathbb{P}_{\mu}(Z \leq k) + \mathbb{P}_{\mu}(Z \leq -k)\end{aligned}\tag{36}$$

Wir sind also an der Wahrscheinlichkeiten $\mathbb{P}_{\mu}(Z \leq -k)$ und $\mathbb{P}_{\mu}(Z \leq k)$ interessiert. Dazu halten wir zunächst fest, dass für jede normalverteilte Zufallsvariable $X \sim N(m, s^2)$ gilt, dass

$$\mathbb{P}(X \leq x) = \mathbb{P}\left(\frac{X - m}{s} \leq \frac{x - m}{s}\right) = \mathbb{P}\left(\xi \leq \frac{x - m}{s}\right) = \Phi\left(\frac{x - m}{s}\right)\tag{37}$$

gilt, wobei ξ hier die durch $(X - m)/s$ definierte Z -verteilte Zufallsvariable bezeichnet.

Im vorliegenden Fall folgt mit $m = \frac{\sqrt{n}}{\sigma}(\mu - \mu_0)$, $s = 1$ und $x = k$ bzw. $x = -k$ also

$$q_{\phi}(\mu) = 1 - \mathbb{P}_{\mu}(Z \leq k) + \mathbb{P}_{\mu}(Z \leq -k) = 1 - \Phi\left(k - \frac{\sqrt{n}}{\sigma}(\mu - \mu_0)\right) + \Phi\left(-k - \frac{\sqrt{n}}{\sigma}(\mu - \mu_0)\right)$$

und wir haben alles gezeigt. $\square$

Theorem (Testumfangankontrolle bei zweiseitigem Z-Test, H_0 einfach)

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ die Stichprobe eines Normalverteilungsmodells mit unbekanntem Erwartungswertparameter μ und bekanntem Varianzparameter $\sigma^2 > 0$. Für $\Theta := \mathbb{R}$ und ein $\mu_0 \in \mathbb{R}$ sei ein Testszenario durch

$$H_0 : \mu = \mu_0 \Leftrightarrow \Theta_0 := \{\mu_0\} \text{ und } H_1 : \mu \neq \mu_0 \Leftrightarrow \Theta_1 := \mathbb{R} \setminus \{\mu_0\}, \quad (38)$$

gegeben. Weiterhin sei mit der Z -Teststatistik

$$Z := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{\sigma} \right) \quad (39)$$

in diesem Szenario der zweiseitige kritische Wert-basierte Test

$$\phi(X) := 1_{\{|Z| \geq k\}} = \begin{cases} 1 & |Z| \geq k \\ 0 & |Z| < k \end{cases}. \quad (40)$$

definiert.

Dann ist ϕ ein Level- α_0 -Test mit Testumfang α_0 , wenn der kritische Wert k definiert ist durch

$$k := \Phi^{-1} \left(1 - \frac{\alpha_0}{2} \right) \quad (41)$$

Wir bezeichnen diesen kritischen Wert mit k_{α_0} .

Z-Test (3) Testumfangkontrolle

Beweis

Damit der betrachtete Test ein Level- α_0 -Test ist, muss bekanntlich $q_\phi(\mu) \leq \alpha_0$ für alle $\mu \in \{\mu_0\}$, also hier $q_\phi(\mu_0) \leq \alpha_0$, gelten.

Weiterhin ist der Testumfang des betrachteten Tests durch $\alpha = \max_{\mu \in \{\mu_0\}} q_\phi(\mu)$, also hier durch $\alpha = q_\phi(\mu_0)$ gegeben.

Wir müssen also zeigen, dass die Wahl von

$$k_{\alpha_0} := \Phi^{-1}(1 - \alpha_0/2) \quad (42)$$

garantiert, dass ϕ ein Level- α_0 -Test mit Testumfang α_0 ist.

Sei also $k := k_{\alpha_0}$. Dann gilt

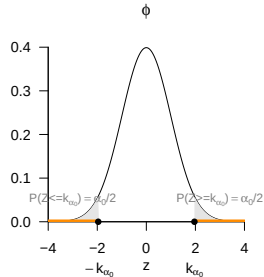
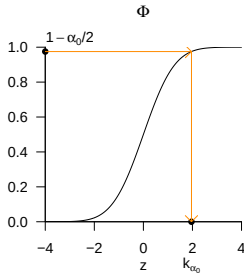
$$\begin{aligned} q_\phi(\mu_0) &= 1 - \Phi(k_{\alpha_0}) + \Phi(-k_{\alpha_0}) \\ &= 1 - \Phi(k_{\alpha_0}) + (1 - \Phi(k_{\alpha_0})) \\ &= 2 - 2\Phi(k_{\alpha_0}) \\ &= 2(1 - \Phi(k_{\alpha_0})) \\ &= 2 \left(1 - \Phi \left(\Phi^{-1}(1 - \alpha_0/2) \right) \right) \\ &= 2(1 - 1 + \alpha_0/2) \\ &= 2\alpha_0/2 \\ &= \alpha_0, \end{aligned} \quad (43)$$

wobei die zweite Gleichung mit der Symmetrie der Standardnormalverteilung folgt.

Es folgt also direkt, dass bei der Wahl von $k = k_{\alpha_0}$, $q_\phi(\mu_0) \leq \alpha_0$ ist und der betrachtete Test somit ein Level- α_0 -Test ist. Weiterhin folgt direkt, dass der Testumfang des betrachteten Tests bei der Wahl von $k = k_{\alpha_0}$ gleich α_0 ist.

Z-Test (3) Testumfangkontrolle

Wahl von $k_{\alpha_0} := \Phi^{-1}(1 - \alpha_0/2)$ mit $\alpha_0 := 0.05$ und resultierender Ablehnungsbereich



Praktisches Vorgehen

- Man nimmt an, dass ein vorliegender Datensatz $x_1, \dots, x_n$ eine Realisation von $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit unbekanntem μ und bekanntem Varianzparameter $\sigma^2 > 0$ ist.
- Man möchte entscheiden ob für ein $\mu_0 \in \mathbb{R}$ eher $H_0 : \mu = \mu_0$ oder $H_1 : \mu \neq \mu_0$ zutrifft.
- Man wählt ein Signifikanzniveau α_0 und bestimmt den zugehörigen kritischen Wert k_{α_0} . Zum Beispiel gilt bei Wahl von $\alpha_0 := 0.05$, dass $k_{0.05} = \Phi^{-1}(1 - 0.05/2) \approx 1.96$ ist.
- Anhand von n, μ_0, σ^2 und $\bar{x}_n$ berechnet man die Realisierung der Z -Teststatistik

$$z := \sqrt{n} \left(\frac{\bar{x}_n - \mu_0}{\sigma} \right) \quad (44)$$

- Wenn z größer-gleich k_{α_0} ist oder wenn z kleiner-gleich $-k_{\alpha_0}$ ist, lehnt man die Nullhypothese ab, andernfalls lehnt man sie nicht ab.
- Die oben entwickelte Theorie des Z -Tests garantiert dann, dass man in höchstens $\alpha_0 \cdot 100$ von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.

Praktisches Vorgehen

Beck Depressions-Inventar II (BDI II)

VP	BDI II Σ	Modellannahmen
1	11	Wir legen das Normalverteilungsmodell zugrunde, d.h. wir nehmen an, dass die BDI II Σ Werte Realisierungen von u.i.v. Zufallsvariablen $X_1, \dots, X_{12} \sim N(\mu, \sigma^2)$ sind. Dies entspricht der Annahme, dass sich der BDI II Σ Wert einer VP durch Addition einer normalverteilten Fehlervariable mit Erwartungswertparameter 0 und Varianzparameter σ^2 zu dem über VP identischen Wert μ ergibt.
2	14	
3	10	
4	9	
5	8	
6	17	
7	3	Fragestellungen (1) Was sind sinnvolle Tipps für die wahren, aber unbekannten, Parameterwerte μ und σ^2 ? (2) Wie hoch ist im frequentistischen Sinne die mit diesen Tipps assoziierte Unsicherheit? (3) Entscheiden wir uns sinnvollerweise für die Hypothese, dass $\mu = 8$ ist oder dass $\mu \neq 8$ ist, wenn wir annehmen, dass $\sigma^2 = 30$ ist?
8	20	
9	13	
10	3	
11	6	
12	2	

Z-Test (3) Testumfangkontrolle

Praktisches Vorgehen

```
X      = c(11, 14, 10, 9, 8, 17, 3, 20, 13, 3, 6, 2)  # BDI Score Datensatz
n      = length(X)                                   # Stichprobengröße
mu_0   = 8                                           # Nullhypothese  $H_0$ 
sigsqr = 30                                          # bekannter Varianzparameter
alpha_0 = 0.05                                       # Signifikanzniveau
k_alpha_0 = qnorm(1 - (alpha_0/2))                 # kritischer Wert
Z      = sqrt(n)*(mean(X) - mu_0)/sqrt(sigsqr)      # Teststatistik
if(abs(Z) >= k_alpha_0){                             # Test  $1_{\{|Z(X)| \geq k_{\alpha_0}\}}$ 
  phi = 1                                           # Ablehnen von  $H_0$ 
} else {
  phi = 0                                           # Nicht Ablehnen von  $H_0$ 
}
cat("x      = ", X,
    "\nalpha_0 = ", alpha_0,
    "\nk_alpha_0 = ", k_alpha_0,
    "\nz      = ", Z,
    "\nphi     = ", phi)
```

```
x      = 11 14 10 9 8 17 3 20 13 3 6 2
alpha_0 = 0.05
k_alpha_0 = 1.96
z      = 1.05
phi     = 0
```


Z-Test (3) Testumfangkontrolle

Praktisches Vorgehen

Beck Depressions-Inventar II (BDI II)

VP	BDI II Σ	Modellannahmen
1	11	Wir legen das Normalverteilungsmodell zugrunde, d.h. wir nehmen an, dass die BDI II Σ Werte Realisierungen von u.i.v. Zufallsvariablen $X_1, \dots, X_{12} \sim N(\mu, \sigma^2)$ sind. Dies entspricht der Annahme, dass sich der BDI II Σ Wert einer VP durch Addition einer normalverteilten Fehlervariable mit Erwartungswertparameter 0 und Varianzparameter σ^2 zu dem über VP identischen Wert μ ergibt.
2	14	
3	10	
4	9	
5	8	
6	17	Fragestellungen
7	3	(3) Antwort: Ein zweiseitiger Z-Test von $H_0 : \mu = 8$ bei einer als bekannt angenommenen Varianz von $\sigma^2 = 30$ ergibt den Z-Teststatistikwert $z = 1.05$ und damit bei einem Signifikanzniveau von $\alpha_0 = 0.05$, einem dementsprechenden kritischen Wert von $k_{\alpha_0} = 1.96$, und somit
8	20	
9	13	
10	3	
11	6	
12	2	$ z = 1.05 = 1.05 < 1.96 \quad (45)$ das Nicht-Ablehnen der Nullhypothese.

Z-Test (3) Testumfangkontrolle

Simulation

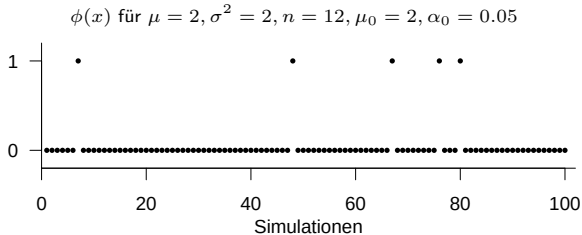
```
# Modellspezifikation
mu      = 2                                # w.a.u. Erwartungswertparameter
sigsqr  = 1                                # bekannter Varianzparameter
n       = 12                               # Stichprobengröße
mu_0    = mu                               # Nullhypothese  $H_0$ 
alpha_0 = 0.05                             # Signifikanzniveau
k_alpha_0 = qnorm(1 - (alpha_0/2))         # kritischer Wert

# Simulation
n_sim   = 1e4                              # Anzahl von Stichprobenrealisierung
phi     = rep(NA, n_sim)                   # Testergebnisarray
for(i in 1:n_sim){
  X      = rnorm(n, mu, sqrt(sigsqr))      # Stichprobenrealisation
  Z      = sqrt(n)*(mean(X) - mu_0)/sqrt(sigsqr) # Teststatistik
  if(abs(Z) >= k_alpha_0){                  # Test  $1_{\{|Z(X)| \geq k_{\alpha_0}\}}$ 
    phi[i] = 1                             # Ablehnen von  $H_0$ 
  } else {
    phi[i] = 0                             # Nicht Ablehnen von  $H_0$ 
  }
}
cat("Geschätzter Testumfang alpha =", mean(phi)) # Ausgabe
```

Geschätzter Testumfang alpha = 0.0517

Z-Test (3) Testumfangkontrolle

Simulation



Z-Test (4) Analyse der Powerfunktion

Wir betrachten die Testgütefunktion

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \Phi\left(k - \frac{\sqrt{n}}{\sigma}(\mu - \mu_0)\right) + \Phi\left(-k - \frac{\sqrt{n}}{\sigma}(\mu - \mu_0)\right) \quad (46)$$

bei bekanntem und festem Wert von $\sigma^2 > 0$, bei festgelegter Nullhypothese, also festem Wert von μ_0 und bei kontrolliertem Testumfang, also für $k := k_{\alpha_0} = \Phi^{-1}(1 - \alpha_0/2)$ mit festem α_0 , als Funktion der übrigen Testszenarioparameter.

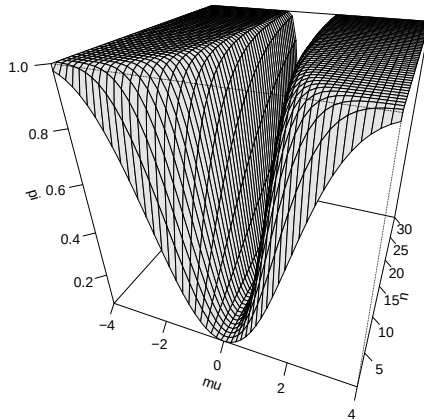
Es ergibt sich die bivariate reellwertige Funktion

$$\pi : \mathbb{R} \times \mathbb{N} \rightarrow [0, 1], (\mu, n) \mapsto \pi(\mu, n) := 1 - \Phi\left(k_{\alpha_0} - \frac{\sqrt{n}}{\sigma}(\mu - \mu_0)\right) + \Phi\left(-k_{\alpha_0} - \frac{\sqrt{n}}{\sigma}(\mu - \mu_0)\right) \quad (47)$$

Bei festgelegten $\sigma^2 > 0, \mu_0, k_{\alpha_0}$ hängt für $\mu \notin \{\mu_0\}$ die Powerfunktion des zweiseitigen Z-Tests mit einfacher Nullhypothese also vom unbekannten Wert μ und von der Stichprobengröße n ab. Wir visualisieren diese Abhängigkeiten untenstehend.

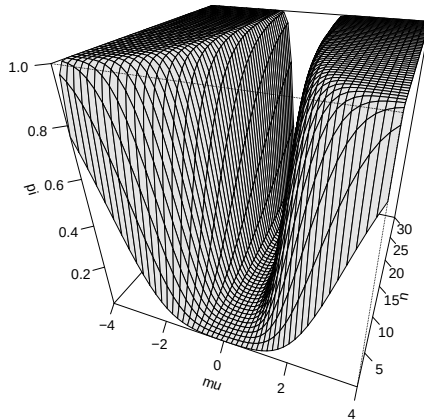
Z-Test (4) Analyse der Powerfunktion

Powerfunktion für $\mu_0 = 0$ und $\sigma^2 = 1$ bei $\alpha_0 = 0.05$



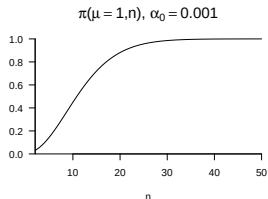
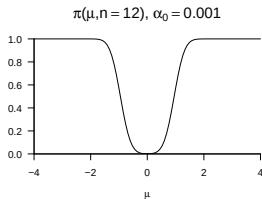
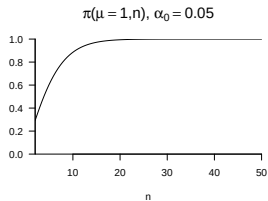
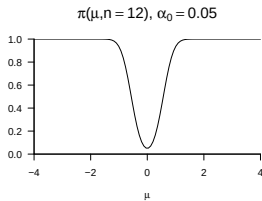
Z-Test (4) Analyse der Powerfunktion

Powerfunktion für $\mu_0 = 0$ und $\sigma^2 = 1$ bei $\alpha_0 = 0.001$



Z-Test (4) Analyse der Powerfunktion

Powerfunktionen für $\mu_0 = 0$ und $\sigma^2 = 1$



Z-Test (4) Analyse der Powerfunktion

Praktisches Vorgehen

Mit größerem n steigt die Powerfunktion des Tests an

- Ein großer Stichprobenumfang ist besser als ein kleiner Stichprobenumfang.
- Kosten für die Erhöhung des Stichprobenumfangs werden aber nicht berücksichtigt.

⇒ Die Theorie statistischer Hypothesentests ist nicht besonders lebensnah.

Die Powerfunktion hängt vom wahren, aber unbekannten, Parameterwert μ ab.

⇒ Wenn man μ schon kennen würde, würde man den Test nicht durchführen.

Generell wird folgendes Vorgehen favorisiert

- Man legt das Signifikanzniveau α_0 fest und evaluiert die Powerfunktion.
- Man wählt einen Mindestparameterwert μ^* , den man mit $\pi(n, \mu) = \beta$ detektieren möchte.
- Ein konventioneller Wert ist $\beta = 0.8$.
- Man liest die für $\pi(n, \mu = \mu^*) = \beta$ nötige Stichprobengröße n ab.

Z-Test (4) Analyse der Powerfunktion

Praktisches Vorgehen

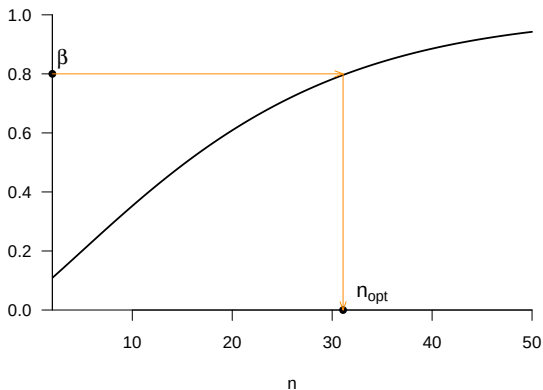
```
# Szenariospezifikation
sigma      = 1                # bekanntes \sigma
mu_0       = 0                # einfache Nullhypothese
n_min      = 2                # n Minimum
n_max      = 50               # n Maximum
n_res      = 1e2              # n Auflösung
n          = seq(n_min,n_max, len = n_res) # n Raum
alpha_0     = 0.05            # Signifikanzniveau
k_alpha_0   = qnorm(1 - alpha_0/2) # kritischer Wert

# Poweranalyse
mu_star     = .5              # Mindestparameterwert
pi_n        = (1-pnorm( k_alpha_0-sqrt(n)/sigma*(mu_star-mu_0)) # Powerfunktion
               +pnorm(-k_alpha_0-sqrt(n)/sigma*(mu_star-mu_0)))
beta        = 0.8             # gewünschter Powerfunktionswert
i           = 1               # Indexinitialisierung
n_min       = NaN             # minimales n Initialisierung
while(pi_n[i] < beta){        # Solange |pi(n,mu*)| < |beta|
  n_min     = n[i]            # Aufnahme des minimal nötigen ns
  i         = i + 1           # und Erhöhung des Indexes
}
cat("Minimal nötiges n =", ceiling(n_min)) # Ausgabe

> Minimal nötiges n = 32
```

Z-Test (4) Analyse der Powerfunktion

$\pi(\mu = \mu^*, n)$ für $\alpha_0 = 0.05$, $\sigma^2 = 1$, $\mu_0 = 0$, $\mu^* = 0.5$



Hypothesentests

- Vorbemerkungen
- Grundlegende Definitionen
- Z-Test
- **p-Werte**
- Konfidenzintervalle und Hypothesentests
- Selbstkontrollfragen

Motivation

- Es werde ein zweiseitiger Z-Test mit Signifikanzlevel $\alpha_0 = 0.05$ durchgeführt.
 - H_0 wird abgelehnt, wenn $|Z| \geq \Phi^{-1}(1 - 0.05/2) \approx 1.96$.
 - Nehmen wir an, es werde $z = 2.01$ beobachtet.
 - Das Testergebnis lautet " H_0 Ablehnen".
 - Nehmen wir an, es werde $z = 3.81$ beobachtet.
 - Das Testergebnis lautet " H_0 Ablehnen".
 - Der alleinige Bericht des Testergebnis supprimiert interessante Information.
- ⇒ Neben der Testumfangkontrolle durch z.B. $\alpha_0 = 0.05$ ist es daher üblich, alle Werte von α_0 anzugeben, für die ein Level- α_0 -Test zum Ablehnen von H_0 führen würde.
- Bei $z = 2.01$ würde H_0 für jedes α_0 mit $2.01 \geq \Phi^{-1}(1 - \alpha_0/2)$ abgelehnt werden.
 - Bei $z = 3.81$ würde H_0 für jedes α_0 mit $3.81 \geq \Phi^{-1}(1 - \alpha_0/2)$ abgelehnt werden.
 - Das kleinste Signifikanzlevel α_0 , bei dem man H_0 basierend auf einem Wert der Teststatistik ablehnen würde, wird *p-Wert* genannt.

Definition (p-Wert)

ϕ sei ein kritischer Wert-basierter Test. Der *p-Wert* ist das kleinste Signifikanzlevel α_0 , bei welchem man die Nullhypothese basierend auf einem vorliegendem Wert der Teststatistik ablehnen würde.

Beispiel (Zweiseitiger Z-Test mit einfacher Nullhypothese)

- Bei $Z = z$ würde H_0 für jedes α_0 mit $|z| \geq \Phi^{-1}(1 - \alpha_0/2)$ abgelehnt werden. Für diese α_0 gilt, wie unten gezeigt,

$$\alpha_0 \geq 2\mathbb{P}(Z \geq |z|). \quad (48)$$

- Das kleinste $\alpha_0 \in [0, 1]$ mit $\alpha_0 \geq 2\mathbb{P}(Z \geq |z|)$ ist dann $\alpha_0 = 2\mathbb{P}(Z \geq |z|)$, also folgt

$$\text{p-Wert} = 2\mathbb{P}(Z \geq |z|) = 2(1 - \Phi(|z|)). \quad (49)$$

- Zum Beispiel ist für $Z = 2.01$ der p-Wert 0.04, für $Z = -2.01$ ist der p-Wert auch 0.04, für $Z = 3.81$ ist der p-Wert 0.0001, und für $Z = -3.81$ ist der p-Wert auch 0.0001

Beispiel (Zweiseitiger Z-Test mit einfacher Nullhypothese, fortgeführt)

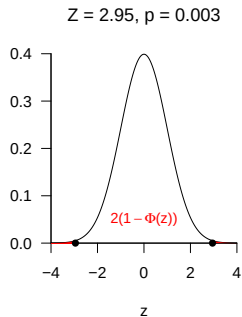
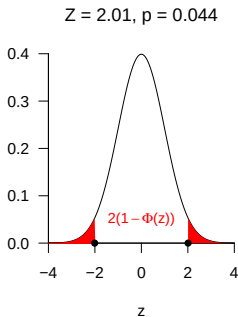
Es bleibt zu zeigen, dass gilt

$$|z| \geq \Phi^{-1}(1 - \alpha_0/2) \Leftrightarrow \alpha_0 \geq 2\mathbb{P}(Z \geq |z|) \quad (50)$$

Wir haben

$$\begin{aligned} |z| &\geq \Phi^{-1}\left(1 - \frac{\alpha_0}{2}\right) \\ \Leftrightarrow |z| &\geq \Phi^{-1}\left(1 - \frac{\alpha_0}{2}\right) \\ \Leftrightarrow \Phi(|z|) &\geq \Phi\left(\Phi^{-1}\left(1 - \frac{\alpha_0}{2}\right)\right) \\ \Leftrightarrow \Phi(|z|) &\geq 1 - \frac{\alpha_0}{2} \\ \Leftrightarrow \mathbb{P}(Z \leq |z|) &\geq 1 - \frac{\alpha_0}{2} \\ \Leftrightarrow \frac{\alpha_0}{2} &\geq 1 - \mathbb{P}(Z \leq |z|) \\ \Leftrightarrow \frac{\alpha_0}{2} &\geq \mathbb{P}(Z \geq |z|) \\ \Leftrightarrow \alpha_0 &\geq 2\mathbb{P}(Z \geq |z|). \end{aligned} \quad (51)$$

Beispiel (Zweiseitiger Z-Test mit einfacher Nullhypothese)



Bemerkungen

- p-Werte spiegeln die Antwort auf die intuitive Frage wie wahrscheinlich es im frequentistischen Sinne wäre, den beobachteten oder einen extremeren Wert der Teststatistik unter der Annahme eines Nullmodels zu observieren.
 - p-Werte sind extrem populär, ihre uninformierte Benutzung ist aber auch sehr umstritten.
- **The American Statistician (2019) Statistical Inference in the 21st Century: A World Beyond $p < 0.05$**
- p-Werte werden, wie Hypothesentestergebnisse generell, leider oft überinterpretiert.
 - Es gibt basierend auf dem Gesagten keinen Grund dies anzunehmen, trotzdem vorsorglich:
 - p-Werte quantifizieren nicht die Wahrscheinlichkeit, dass die Nullhypothese wahr ist.
 - Aufgrund von $p < 0.05$ sollte man nicht glauben, dass ein Effekt existiert.
 - Aufgrund von $p > 0.05$ sollte man nicht glauben, dass ein Effekt nicht existiert.
 - p-Werte sind eine Möglichkeit ein Signal-zu-Rauschen Verhältnis zu quantifizieren.
 - p-Werte sind eine Möglichkeit Unsicherheit zu quantifizieren.

Hypothesentests

- Vorbemerkungen
- Grundlegende Definitionen
- Z-Test
- p-Wert
- **Konfidenzintervalle und Hypothesentests**
- Selbstkontrollfragen

Theorem (Dualität von Konfidenzintervallen und Hypothesentests I)

$X = (X_1, \dots, X_n) \sim p_\theta$ sei eine Stichprobe mit Ergebnisraum $\mathcal{X}$ und Parameterraum Θ . Weiterhin sei $[b_l(X), b_u(X)]$ ein δ -Konfidenzintervall für θ .

Dann ist der Hypothesentest

$$\phi_\theta : \mathcal{X} \rightarrow \{0, 1\}, x \mapsto \phi(x) := \begin{cases} 0, & [b_l(x), b_u(x)] \ni \theta \\ 1, & [b_l(x), b_u(x)] \not\ni \theta \end{cases} \quad (52)$$

ein Test vom Signifikanzniveau $\alpha_0 = 1 - \delta$ für die Hypothesen

$$\Theta_0 := \{\theta\} \text{ und } \Theta_1 := \Theta \setminus \{\theta\}. \quad (53)$$

Beweis

Aufgrund der einfachen Nullhypothese und somit $\alpha_0 = \alpha$ folgt

$$\alpha_0 = \alpha = \mathbb{P}_\theta(\phi(X) = 1) = \mathbb{P}_\theta([b_l(x), b_u(x)] \not\ni \theta) = 1 - \mathbb{P}_\theta([b_l(x), b_u(x)] \ni \theta) = 1 - \delta. \quad (54)$$

Bemerkungen

- θ bezeichnet hier das Element der einfachen Nullhypothese
- Mit δ -Konfidenzintervallen kann man also Tests vom Signifikanzniveau $\alpha_0 = 1 - \delta$ konstruieren.

Konfidenzintervalle und Hypothesentests

Beispiel (Konstruktion eines Hypothesentests aus einem Konfidenzintervall)

Wir haben bereits gesehen, dass für eine Stichprobe $X_1, \dots, X_n \sim N(\mu_0, \sigma^2)$ mit unbekanntem Erwartungswertparameter μ_0 und bekanntem Varianzparameter $\sigma^2 > 0$, $\delta \in]0, 1[$ und $z_\delta := \Phi^{-1}\left(\frac{1+\delta}{2}\right)$,

$$C_n := \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_\delta, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_\delta \right]. \quad (55)$$

ein δ -Konfidenzintervall definiert.

Mit der Dualität von Konfidenzintervallen und Hypothesentests können wir also folgenden Test für die Hypothesen $\Theta_0 = \{\mu_0\}$ und $\Theta_1 = \mathbb{R} \setminus \mu_0$ definieren:

$$\phi : \mathcal{X} \rightarrow \{0, 1\}, x \mapsto \phi(x) := \begin{cases} 0, & \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_\delta, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_\delta \right] \ni \mu_0 \\ 1, & \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_\delta, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_\delta \right] \not\ni \mu_0 \end{cases} \quad (56)$$

Dann gilt

$$\begin{aligned} \mathbb{P}_{\mu_0}(\phi(X) = 1) &= 1 - \mathbb{P}_{\mu_0}(\phi(X) = 0) \\ &= 1 - \mathbb{P}_{\mu_0}\left(\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_\delta, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_\delta\right] \ni \mu_0\right) \\ &= 1 - \delta. \end{aligned} \quad (57)$$

und wir haben gezeigt, dass ϕ ein Test vom Signifikanzniveau $\alpha_0 = 1 - \delta$ ist.

Konfidenzintervalle und Hypothesentests

```
# Modellformulierung
mu_0      = 2
sigma     = 1
n         = 12
delta     = 0.95
phi_inv   = qnorm((1+delta)/2)

# w.a.u. Erwartungswertparameter/Nullhypothese
# Standardabweichungsparameter
# Stichprobengröße
# Konfidenzbedingung
#  $\Phi^{-1}((\delta + 1)/2)$ 

# Simulationssetup
ns        = 1e2
X_bar     = rep(NaN,ns)
C         = matrix(rep(NaN,2*ns), ncol = 2)
kfn       = rep(NaN,ns)
phi       = rep(NaN,ns)

# Anzahl Simulationen
# Stichprobenmittelarray
# Konfidenzintervallarray
# Überdeckungsindikatorarray
# Testarray

# Simulationsiterationen
for(i in 1:ns){

  # Datengeneration
  X      = rnorm(n,mu_0,sigma)

  # Stichprobenrealisierung

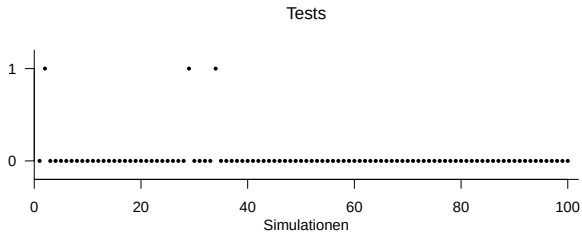
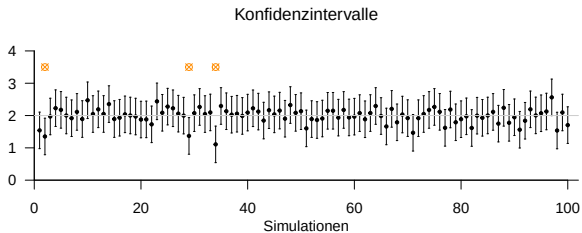
  # Konfidenzintervallevaluation
  X_bar[i] = mean(X)
  C[i,1]   = X_bar[i] - (sigma/sqrt(n))*phi_inv
  C[i,2]   = X_bar[i] + (sigma/sqrt(n))*phi_inv

  # Stichprobenmittel
  # untere KI Grenze
  # obere KI Grenze

  # Überdeckungsindikatorevaluation
  if(C[i,1] <= mu_0 & mu_0 <= C[i,2]){
    kfn[i] = 1} else{
    kfn[i] = 0}

  # Testevaluation
  if(C[i,1] <= mu_0 & mu_0 <= C[i,2]){
    phi[i] = 0} else{
    phi[i] = 1}
}
```

Konfidenzintervalle und Hypothesentests



Theorem (Dualität von Konfidenzintervallen und Hypothesentests II)

Unter den gleichen Annahmen wie im vorherigen Theorem sei

$$\Phi := \{\phi_\theta(X) | \theta \in \Theta\} \quad (58)$$

eine Familie von Tests, so dass $\phi_\theta(X)$ ein Test mit Signifikanzniveau α_0 für die Hypothesen

$$\Theta_0 := \{\theta\} \text{ and } \Theta_1 := \Theta \setminus \{\theta\}. \quad (59)$$

sei.

Man nehme weiter an, dass die Menge

$$C_n := \{\theta \in \Theta | \phi_\theta(X) = 0\} \quad (60)$$

als $C_n = [b_l(X), b_u(X)]$ für entsprechend bestimmte $b_l(X)$ und $b_u(X)$ geschrieben werden kann. Dann ist C_n ein $\delta := 1 - \alpha_0$ Konfidenzintervall für θ .

Beweis

Aufgrund der einfachen Nullhypothese und somit $\alpha_0 = \alpha$ folgt das Level $\delta = 1 - \alpha_0$ des Konfidenzintervalls aus

$$\delta = \mathbb{P}_\theta(C_n \ni \theta) = \mathbb{P}_\theta(\phi_\theta(X) = 0) = 1 - \mathbb{P}_\theta(\phi_\theta(X) = 1) = 1 - \alpha_0. \quad (61)$$

Bemerkungen

- Mit Tests von Signifikanzniveau α_0 kann man also $1 - \alpha_0$ -Konfidenzintervalle konstruieren.

Hypothesentests

- Vorbemerkungen
- Grundlegende Definitionen
- Z-Test
- p-Wert
- Konfidenzintervalle und Hypothesentests
- **Selbstkontrollfragen**

Selbstkontrollfragen

1. Erläutern Sie die grundlegende Logik statistischer Hypothesentests.
2. Geben Sie die Definition statistischer Hypothesen und eines Testszenarios wieder.
3. Definieren Sie die Begriffe der einfachen und zusammengesetzten Hypothesen.
4. Definieren Sie die Begriffe der einseitigen und zweiseitigen Hypothesen.
5. Definieren Sie den Begriff des Tests.
6. Definieren Sie den Begriff des Standardtests.
7. Definieren Sie den Begriff des kritischen Bereichs eines Tests.
8. Definieren Sie den Begriff des Ablehnungsbereichs eines Tests.
9. Definieren Sie den Begriff des kritischen Wert-basierten Tests.
10. Definieren Sie richtige Testentscheidungen, Typ I Fehler und Typ II Fehler.
11. Definieren Sie die Testgütefunktion.
12. Erläutern Sie die Bedeutung der Testgütefunktion im Rahmen der Konstruktion statistischer Tests.
13. Definieren Sie die Begriffe des Signifikanzniveaus und des Level- α_0 -Tests.
14. Definieren Sie den Begriff des Testumfangs.
15. Erläutern Sie die prinzipielle Strategie zur Wahl von Null- und Alternativhypothesen in der Wissenschaft.
16. Nennen Sie vier Schritte in der Konstruktion eines Tests.

Selbstkontrollfragen

17. Definieren Sie das statistische Modell eines Z-Tests.
18. Definieren Sie die Hypothesen eines Z-Tests mit einfacher Nullhypothese und zweiseitiger Alternativhypothese.
19. Definieren Sie den zweiseitigen Z-Test.
20. Skizzieren Sie qualitativ Testgütefunktionen eines zweiseitigen Z-Tests für verschiedene kritische Werte.
21. Wie muss der kritische Wert eines zweiseitigen Z-Tests definiert sein, damit der Test ein Level- α_0 -Test ist?
22. Skizzieren Sie qualitativ die Bestimmung des kritischen Wertes k_{α_0} bei einem zweiseitigen Z-Test.
23. Erläutern Sie das praktische Vorgehen zur Durchführung eines zweiseitigen Z-Tests.
24. Von welchen Werten hängt die Powerfunktion eines zweiseitigen Z-Tests ab?
25. Skizzieren Sie qualitativ die Powerfunktion des zweiseitigen Z-Tests bei fester Stichprobengröße.
26. Skizzieren Sie qualitativ die Powerfunktion des zweiseitigen Z-Tests bei festem Erwartungswertparameter.
27. Erläutern Sie das favorisierte praktische Vorgehen zur Durchführung einer Poweranalyse.
28. Erläutern Sie die Motivation zur Auswertung von p-Werten.
29. Definieren Sie den Begriff des p-Werts.
30. Geben Sie das erste Theorem zur Dualität von Konfidenzintervallen und Hypothesentests wieder.
31. Geben Sie das zweite Theorem zur Dualität von Konfidenzintervallen und Hypothesentests wieder.
32. Erläutern Sie die Dualität von Konfidenzintervallen und Hypothesentests.

References

- Czado, Claudia, and Thorsten Schmidt. 2011. *Mathematische Statistik*. Statistik und ihre Anwendungen. Berlin: Springer.
- DeGroot, Morris H., and Mark J. Schervish. 2012. *Probability and Statistics*. 4th ed. Boston: Addison-Wesley.
- Horvath, Lilla, Stanley Colcombe, Michael Milham, Shruti Ray, Philipp Schwartenbeck, and Dirk Ostwald. 2020. "Human Belief State-Based Exploration and Exploitation in an Information-Selective Symmetric Reversal Bandit Task." Preprint. Neuroscience. <https://doi.org/10.1101/2020.08.31.276139>.
- Ostwald, Dirk, Ludger Starke, and Ralph Hertwig. 2015. "A Normative Inference Approach for Optimal Sample Sizes in Decisions from Experience." *Frontiers in Psychology* 6 (September). <https://doi.org/10.3389/fpsyg.2015.01342>.
- Pratt, John, Howard Raiffa, and Robert Schlaifer. 1995. *Statistical Decision Theory*. MIT Press.
- Puterman, Martin. 2005. *Markov Decision Processes*. Wiley-Interscience.



Inferenzstatistik

BSc Psychologie SoSe 2021

Prof. Dr. Dirk Ostwald

(12) T-Tests

Die Darstellung der basiert im Wesentlichen auf Kapitel 9 in DeGroot and Schervish (2012).

T-Tests

- T-Teststatistik
- Einstichproben-T-Tests
 - $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$
 - $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$
- Zweistichproben-T-Tests
 - Unabhängige Stichproben
 - Abhängige Stichproben
- Selbstkontrollfragen

T-Tests

- **T-Teststatistik**
- Einstichproben-T-Tests
 - $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$
 - $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$
- Zweistichproben-T-Tests
 - Unabhängige Stichproben
 - Abhängige Stichproben
- Selbstkontrollfragen

Definition (T-Teststatistik)

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ sei die Stichprobe eines Normalverteilungmodells, $\bar{X}_n$ bezeichne das Stichprobenmittel, S_n bezeichne die Stichprobenstandardabweichung, und es sei $\mu_0 \in \mathbb{R}$. Die *T-Teststatistik* ist definiert als

$$T := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right). \quad (1)$$

Bemerkungen

- Im Gegensatz zur T-Statistik muss bei der T-Teststatistik nicht $\mu_0 = \mu$ gelten.
- Intuitiv kann die T-Teststatistik als mit der Stichprobengröße (Evidenz) gewichtetes Verhältnis von Signal (systematischer Variabilität) zu Rauschen (unsystematische Variabilität) verstanden werden:

$$\sqrt{\text{Stichprobengröße}} \left(\frac{\text{Signal}}{\text{Rauschen}} \right) = \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right) \quad (2)$$

- Die T-Teststatistik ist eine skalare Deskription des Effekt vs. Variabilität Verhältnisses eines Datensatzes.
- In der T-Teststatistik wird die Effektgröße in Einheiten der Stichprobenstandardabweichung gemessen
 - $T = 1 \Leftrightarrow \sqrt{n}(\bar{X}_n - \mu_0) = 1S_n$
 - $T = 2 \Leftrightarrow \sqrt{n}(\bar{X}_n - \mu_0) = 2S_n$

Definition (Nichtzentrale t -Zufallsvariable)

T sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}$ und WDF

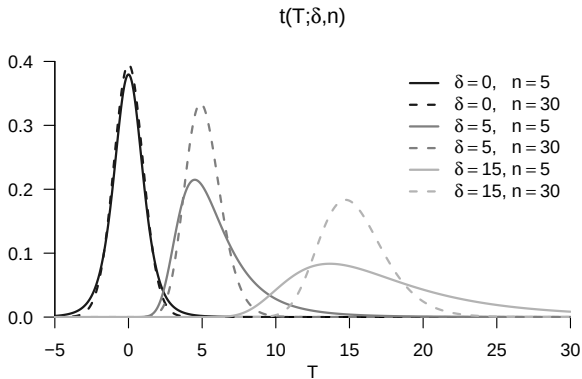
$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, t \mapsto p(t) := \frac{1}{2^{\frac{n-1}{2}} \Gamma\left(\frac{n}{2}\right) (n\pi)^{\frac{1}{2}}} \times \int_0^\infty \tau^{\frac{n-1}{2}} \exp\left(-\frac{\tau}{2}\right) \exp\left(-\frac{1}{2} \left(t \left(\frac{\tau}{n}\right)^{\frac{1}{2}} - \delta\right)^2\right) d\tau. \quad (3)$$

Dann sagen wir, dass T einer nichtzentralen t -Verteilung mit Nichtzentralitätsparameter δ und Freiheitsgradparameter n unterliegt und nennen T eine *nichtzentrale t -Zufallsvariable mit Nichtzentralitätsparameter δ und Freiheitsgradparameter n* . Wir kürzen dies mit $t(\delta, n)$ ab. Die WDF einer nichtzentralen t -Zufallsvariable bezeichnen wir mit $t(T; \delta, n)$. Die KVF und inverse KVF einer nichtzentralen t -Zufallsvariable bezeichnen wir mit $\psi(\cdot; \delta, n)$ und $\psi^{-1}(\cdot; \delta, n)$, respektive.

Bemerkungen

- Eine nichtzentrale t -Zufallsvariable mit $\delta = 0$ ist eine t -Zufallsvariable.
- Es gilt also $t(T; 0, n) = t(T; n)$.
- Die funktionale Form der WDF findet sich zum Beispiel in Lehmann (1986), Seite 254, Gl. (80)

Wahrscheinlichkeitsdichtefunktionen nichtzentraler t -Verteilungen



Theorem (Nichtzentrale T-Transformation)

$X \sim N(\mu, 1)$ sei eine normalverteilte Zufallsvariable, $U \sim \chi^2(n)$ sei eine χ^2 Zufallsvariable mit Freiheitsgradparameter n , und X und U seien unabhängige Zufallsvariablen. Dann ist die Zufallsvariable

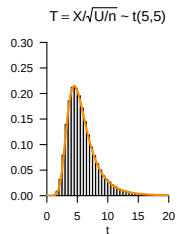
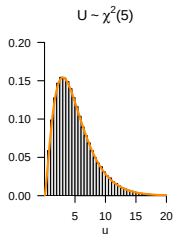
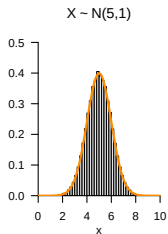
$$T := \frac{X}{\sqrt{U/n}} \quad (4)$$

eine nichtzentrale t -Zufallsvariable mit Nichtzentralitätsparameter μ und Freiheitsgradparameter n , es gilt also $T \sim t(\mu, n)$

Bemerkung

- Wir verzichten auf einen Beweis.

Nichtzentrale T-Transformation



Theorem (Verteilung der T-Teststatistik)

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ sei die Stichprobe eines Normalverteilungmodells, $\bar{X}_n$ sei das Stichprobenmittel, S_n sei die Stichprobenstandardabweichung, und es sei $\mu_0 \in \mathbb{R}$. Dann ist die T-Teststatistik

$$T := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right) \quad (5)$$

eine nichtzentrale t -Zufallsvariable mit Nichtzentralitätsparameter

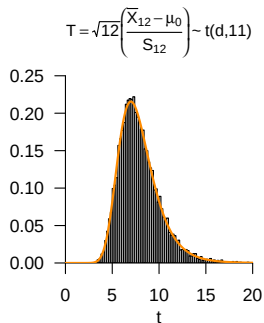
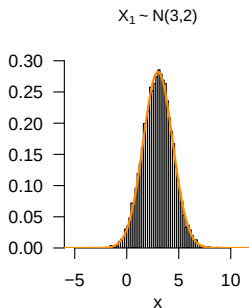
$$d = \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (6)$$

und Freiheitsgradparameter $n - 1$, es gilt also $T \sim t(d, n - 1)$

Bemerkung

- Wir verzichten auf einen Beweis.

T-Teststatistik



T-Tests

- T-Teststatistik
- **Einstichproben-T-Tests**
 - $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$
 - $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$
- Zweistichproben-T-Tests
 - Unabhängige Stichproben
 - Abhängige Stichproben
- Selbstkontrollfragen

Anwendungsszenario

- **Eine Stichprobe** experimenteller Einheiten.
- Annahme der unabhängigen und identischen Normalverteilung $N(\mu, \sigma^2)$ der Daten.
- μ und σ^2 unbekannt.
- Quantifizieren der Unsicherheit beim inferentiellen Vergleich von μ mit μ_0 beabsichtigt.

Anwendungsbeispiele

- Gruppenanalysen mit Wechsler Adult Intelligence Scale
 - $\mu \neq \mu_0 := 100 \Rightarrow$ Evidenz für über- oder unterdurchschnittliche WAIS Performanz
- BDI Score Datenanalyse einer Gruppe psychiatrischer Patient:innen
 - $\mu > \mu_0 := 18 \Rightarrow$ Klinisch auffällige Gruppe
- Gruppenanalysen in der funktionellen Kernspintomographie
 - $\mu > \mu_0 := 0 \Rightarrow$ Evidenz für regionale Gehirnaktivierung

Hypothesenszenarien

Einfache Nullhypothese, einfache Alternativhypothese $H_0 : \mu = \mu_0, H_1 : \mu = \mu_1$

- Theoretisch wichtiges Szenario (Neymann-Pearson Lemma)
- Praktische Relevanz eher gering

Einfache Nullhypothese, zusammengesetzte Alternativhypothese $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$

- Zweiseitiger Einstichproben-T-Test mit ungerichteter Hypothese
- Ungerichtete Fragestellung nach einem Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese $H_0 : \mu \leq \mu_0, H_1 : \mu > \mu_0$

- Einseitiger Einstichproben-T-Test mit gerichteter Hypothese
- Gerichtete Fragestellung nach einem positiven Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese $H_0 : \mu \geq \mu_0, H_1 : \mu < \mu_0$

- Gerichtete Fragestellung nach einem negativen Unterschied
- Qualitativ äquivalente Theorie zum umgekehrten Fall

Einstichproben-T-Tests

Im Folgenden näher betrachtete Hypothesenszenarien

Einfache Nullhypothese, einfache Alternativhypothese $H_0 : \mu = \mu_0, H_1 : \mu = \mu_1$

- Theoretisch wichtiges Szenario (Neymann-Pearson Lemma)
- Praktische Relevanz eher gering

Einfache Nullhypothese, zusammengesetzte Alternativhypothese $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$

- Zweiseitiger Einstichproben-T-Test mit ungerichteter Hypothese
- Ungerichtete Fragestellung nach einem Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese $H_0 : \mu \leq \mu_0, H_1 : \mu > \mu_0$

- Einseitiger Einstichproben-T-Test mit gerichteter Hypothese
- Gerichtete Fragestellung nach einem positiven Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese $H_0 : \mu \geq \mu_0, H_1 : \mu < \mu_0$

- Gerichtete Fragestellung nach einem negativen Unterschied
- Qualitativ äquivalente Theorie zum umgekehrten Fall

Gliederung

- (1) Statistisches Modell in klassischer Form
- (2) Statistisches Modell in generativer Form
- (3) Testhypothesen, Teststatistik, Test
- (4) Analyse der Testgütefunktion
- (5) Testumfangkontrolle
- (6) p-Werte
- (7) Analyse der Powerfunktion

T-Tests

- T-Teststatistik
- **Einstichproben-T-Tests**
 - $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$
 - $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$
- Zweistichproben-T-Tests
 - Unabhängige Stichproben
 - Abhängige Stichproben

Einstichproben-T-Test mit $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$

(1) Statistisches Modell in klassischer Form

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ sei die Stichprobe eines Normalverteilungsmodells mit unbekanntem Erwartungswertparameter μ und unbekanntem Varianzparameter $\sigma^2 > 0$. Als Parameter von Interesse betrachten wir $\theta = \mu$, so dass sich der Parameterraum von Interesse zu $\Theta = \mathbb{R}$ ergibt.

(2) Statistisches Modell in generativer Form

Es sei

$$X_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ für } i = 1, \dots, n \quad (7)$$

wobei

- $X_i, i = 1, \dots, n$ beobachtbare Zufallsvariablen,
- $\mu \in \mathbb{R}$ den festen identischen Erwartungswertparameter über Stichprobenvariablen, und
- $\varepsilon_i, i = 1, \dots, n$ unabhängige normalverteilte nicht-beobachtbare Zufallsvariablen

bezeichnen.

(2) Statistisches Modell in generativer Form (fortgeführt)

Die generative Form betont, dass im vorliegenden Modell beobachtete Daten durch einen systematischen deterministischen Prozess (hier μ) unter dem additiven Einfluss einer Vielzahl unabhängiger und deshalb in der Summe normalverteilter Störprozesse (hier $\varepsilon_i, i = 1, \dots, n$) erzeugt konzipiert werden.

Beweis

Die Äquivalenz beider Modellformen folgt direkt aus der Transformation normalverteilter Zufallsvariablen durch linear-affine Funktionen (cf. (6) Transformationen der Normalverteilung). Speziell gilt im vorliegenden Fall für $\varepsilon_i \sim N(0, \sigma^2)$, dass

$$X_i = f(\varepsilon_i) \text{ mit } f : \mathbb{R} \rightarrow \mathbb{R}, \varepsilon_i \mapsto f(\varepsilon_i) := \varepsilon_i + \mu. \quad (8)$$

Mit den WDF Transformationstheorem bei linear-affinen Abbildungen folgt dann

$$\begin{aligned} p_{X_i}(x_i) &= \frac{1}{|1|} p_{\varepsilon_i} \left(\frac{x_i - \mu}{1} \right) \\ &= N(x_i - \mu; 0, \sigma^2) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{1}{2\sigma^2} (x_i - \mu - 0)^2 \right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{1}{2\sigma^2} (x_i - \mu)^2 \right) \\ &= N(x_i; \mu, \sigma^2), \end{aligned} \quad (9)$$

also $X_i \sim N(\mu, \sigma^2)$.

□

(3) Testhypothesen, Teststatistik, Test

Für ein $\mu_0 \in \mathbb{R}$ betrachten wir die einfache Nullhypothese und die zusammengesetzte Alternativhypothese

$$H_0 : \mu = \mu_0 \Leftrightarrow \Theta_0 := \{\mu_0\} \text{ und } H_1 : \mu \neq \mu_0 \Leftrightarrow \Theta_1 := \mathbb{R} \setminus \{\mu_0\}, \quad (10)$$

respektive. Weiterhin betrachten wir die T-Teststatistik

$$T := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right) \quad (11)$$

und definieren den zweiseitigen kritischen Wert-basierten Test

$$\phi(X) := 1_{\{|T| \geq k\}} = \begin{cases} 1 & |T| \geq k \\ 0 & |T| < k \end{cases}. \quad (12)$$

(4) Analyse der Testgütefunktion

Theorem (Testgütefunktion)

ϕ sei der im obigen Testszenario definierte Test. Dann ist die Testgütefunktion von ϕ gegeben durch

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \psi(k; d_\mu, n - 1) + \psi(-k; d_\mu, n - 1) \quad (13)$$

wobei $\psi(\cdot; d_\mu, n - 1)$ die KVF der nichtzentralen t -Verteilung mit Nichtzentralitätsparameter

$$d_\mu := \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (14)$$

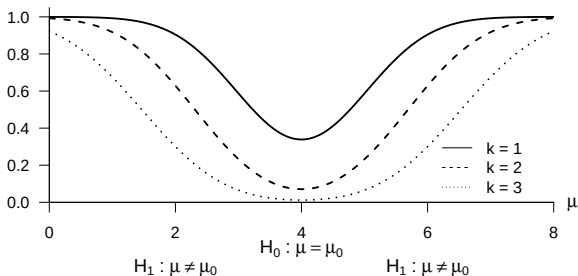
und Freiheitsgradparameter $n - 1$ bezeichnet.

Einstichproben-T-Test mit $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$

(4) Analyse der Testgütefunktion

Testgütefunktion q_ϕ für $\sigma^2 = 9$, $\mu_0 = 4$, $n = 12$ und $k = 1, 2, 3$.

$$q_\phi(\mu) = \mathbb{P}_\mu(\phi = 1)$$



(4) Analyse der Testgütefunktion

Beweis

Die Testgütefunktion des betrachteten Test im vorliegenden Testszenario ist definiert als

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := \mathbb{P}_\mu(\phi = 1). \quad (15)$$

Da die Wahrscheinlichkeiten für $\phi = 1$ und dafür, dass die zugehörige Teststatistik im Ablehnungsbereich des Tests liegt gleich sind, benötigen wir die also zunächst die Verteilung der Teststatistik. Wir haben oben bereits gesehen, dass die T-Teststatistik

$$T := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right) \quad (16)$$

unter der Annahme $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ nach einer nichtzentralen t -Verteilung $t(d, n - 1)$ mit Nichtzentralitätsparameter

$$d = \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (17)$$

verteilt ist. Der Ablehnungsbereich des zweiseitigen T-Tests ergibt sich, wie in ähnlicher Form bei der Betrachtung des zweiseitigen Z-Tests gesehen, zu

$$A =] - \infty, -k] \cup]k, \infty[. \quad (18)$$

(4) Analyse der Testgütefunktion

Beweis (fortgeführt)

Mit diesem Ablehnungsbereich ergibt sich dann

$$\begin{aligned} q_\phi(\mu) &= \mathbb{P}_\mu(\phi = 1) \\ &= \mathbb{P}_\mu(T \in]-\infty, -k] \cup]k, \infty[) \\ &= \mathbb{P}_\mu(T \in]-\infty, -k]) + \mathbb{P}_\mu(T \in [k, \infty[) \\ &= \mathbb{P}_\mu(T \leq -k) + \mathbb{P}_\mu(T \geq k) \\ &= \mathbb{P}_\mu(T \leq -k) + (1 - \mathbb{P}_\mu(T \leq k)) \\ &= 1 - \mathbb{P}_\mu(T \leq k) + \mathbb{P}_\mu(T \leq -k) \\ &= 1 - \psi(k; d_\mu, n-1) + \psi(-k; d_\mu, n-1), \end{aligned} \tag{19}$$

wobei $\psi(\cdot; d_\mu, n-1)$ die KVF der nichtzentralen T-Verteilung mit Nichtzentralitätsparameter d und Freiheitsgradparameter $n-1$ bezeichnet.

□

(5) Testumfangkontrolle

Theorem (Testumfangkontrolle)

ϕ sei der im obigen Testszenario definierte Test. Dann ist ϕ ein Level- α_0 -Test mit Testumfang α_0 , wenn der kritische Wert definiert ist durch

$$k_{\alpha_0} := \psi^{-1} \left(1 - \frac{\alpha_0}{2}; n - 1 \right), \quad (20)$$

wobei $\psi^{-1}(\cdot; n - 1)$ die inverse KVF der t -Verteilung mit $n - 1$ Freiheitsgraden ist.

Einstichproben-T-Test mit $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$

(5) Testumfangkontrolle

Beweis

Damit der betrachtete Test ein Level- α_0 -Test ist, muss bekanntlich $q_\phi(\mu) \leq \alpha_0$ für alle $\mu \in \{\mu_0\}$, also hier $q_\phi(\mu_0) \leq \alpha_0$, gelten. Weiterhin ist der Testumfang des betrachteten Tests durch $\alpha = \max_{\mu \in \{\mu_0\}} q_\phi(\mu)$, also hier durch $\alpha = q_\phi(\mu_0)$ gegeben. Wir müssen also zeigen, dass die Wahl von k_{α_0} garantiert, dass ϕ ein Level- α_0 -Test mit Testumfang α_0 ist. Dazu merken wir zunächst an, dass für $\mu = \mu_0$ gilt, dass

$$\begin{aligned} q_\phi(\mu_0) &= 1 - \psi(k; d_{\mu_0}, n-1) + \psi(-k; d_{\mu_0}, n-1) \\ &= 1 - \psi(k; 0, n-1) + \psi(-k; 0, n-1) \\ &= 1 - \psi(k; n-1) + \psi(-k; n-1), \end{aligned} \tag{21}$$

wobei $\psi(\cdot; d, n-1)$ und $\psi(\cdot; n-1)$ die KVF der nichtzentralen t -Verteilung mit Nichtzentralitätsparameter d und Freiheitsgradparameter $n-1$ sowie der t -Verteilung mit Freiheitsgradparameter $n-1$, respektive, bezeichnen. Sei nun also $k := k_{\alpha_0}$. Dann gilt

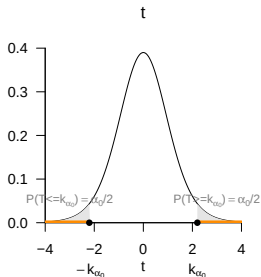
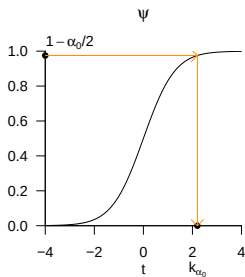
$$\begin{aligned} q_\phi(\mu_0) &= 1 - \psi(k_{\alpha_0}) + \psi(-k_{\alpha_0}) \\ &= 1 - \psi(k_{\alpha_0}) + (1 - \psi(k_{\alpha_0})) \\ &= 2(1 - \psi(k_{\alpha_0})) \\ &= 2 \left(1 - \psi \left(\psi^{-1}(1 - \alpha_0/2) \right) \right) \\ &= 2(1 - 1 + \alpha_0/2) \\ &= \alpha_0, \end{aligned} \tag{22}$$

wobei die zweite Gleichung mit der Symmetrie der t -Verteilung folgt. Es folgt also direkt, dass bei der Wahl von $k = k_{\alpha_0}$, $q_\phi(\mu_0) \leq \alpha_0$ ist und der betrachtete Test somit ein Level- α_0 -Test ist. Weiterhin folgt direkt, dass der Testumfang des betrachteten Tests bei der Wahl von $k = k_{\alpha_0}$ gleich α_0 ist.

Einstichproben-T-Test mit $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$

(5) Testumfangkontrolle

Wahl von $k_{\alpha_0} := \psi^{-1}(1 - \alpha_0/2; n - 1)$ mit $n = 12$, $\alpha_0 := 0.05$ und Ablehnungsbereich



(5) Testumfangkontrolle

Praktisches Vorgehen

- Man nimmt an, dass ein vorliegender Datensatz $x_1, \dots, x_n$ eine Realisation von $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit unbekannten Parametern μ und $\sigma^2 > 0$ ist.
- Man möchte entscheiden ob für ein $\mu_0 \in \mathbb{R}$ eher $H_0 : \mu = \mu_0$ oder $H_1 : \mu \neq \mu_0$ zutrifft.
- Man wählt ein Signifikanzniveau α_0 und bestimmt den zugehörigen Freiheitsgradparameter-abhängigen kritischen Wert k_{α_0} . Zum Beispiel gilt bei Wahl von $\alpha_0 := 0.05$ und $n = 12$, also Freiheitsgradparameter 11, dass $k_{0.05} = \psi^{-1}(1 - 0.05/2; 11) \approx 2.20$ ist.
- Anhand von $n, \mu_0, \bar{x}_n$ und s_n berechnet man die Realisierung der T-Teststatistik

$$t := \sqrt{n} \left(\frac{\bar{x}_n - \mu_0}{s_n} \right) \quad (23)$$

- Wenn t größer-gleich k_{α_0} ist oder wenn t kleiner- gleich $-k_{\alpha_0}$ ist, lehnt man die Nullhypothese ab, andernfalls lehnt man sie nicht ab.
- Die oben entwickelte Theorie garantiert dann, dass man in höchstens $\alpha_0 \cdot 100$ von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.

(6) p-Werte

Bestimmung des p-Wertes

- Per Definition ist der p-Wert das kleinste Signifikanzlevel α_0 , bei welchem man die Nullhypothese basierend auf einem vorliegendem Wert der Teststatistik ablehnen würde.
- Bei $T = t$ würde H_0 für jedes α_0 mit $|t| \geq \psi^{-1}(1 - \alpha_0/2; n - 1)$ abgelehnt werden. Für diese α_0 gilt, wie unten gezeigt,

$$\alpha_0 \geq 2\mathbb{P}(T \geq |t|). \quad (24)$$

- Das kleinste $\alpha_0 \in [0, 1]$ mit $\alpha_0 \geq 2\mathbb{P}(T \geq |t|)$ ist dann $\alpha_0 = 2\mathbb{P}(T \geq |t|)$, also folgt

$$\text{p-Wert} = 2\mathbb{P}(T \geq |t|) = 2(1 - \psi(|t|; n - 1)). \quad (25)$$

- Im Gegensatz zum Z-Test hängt bei T-Tests der p-Wert auch von der Stichprobengröße ab.
- Zum Beispiel ist für $T = 2.00$ und $n = 10$ der p-Wert 0.076, für $T = 2.00$ und $n = 100$ ist der p-Wert dagegen 0.048.

(6) p-Werte

Bestimmung des p-Wertes

- Es bleibt zu zeigen, dass gilt

$$|t| \geq \psi^{-1}(1 - \alpha_0/2; n - 1) \Leftrightarrow \alpha_0 \geq 2\mathbb{P}(T \geq |t|) \quad (26)$$

- Dies aber folgt aus

$$\begin{aligned} |t| &\geq \psi^{-1}\left(1 - \frac{\alpha_0}{2}; n - 1\right) \\ \Leftrightarrow \psi(|t|; n - 1) &\geq \psi\left(\psi^{-1}\left(1 - \frac{\alpha_0}{2}; n - 1\right); n - 1\right) \\ \Leftrightarrow \psi(|t|; n - 1) &\geq 1 - \frac{\alpha_0}{2} \\ \Leftrightarrow \mathbb{P}(T \leq |t|) &\geq 1 - \frac{\alpha_0}{2} \\ \Leftrightarrow \frac{\alpha_0}{2} &\geq 1 - \mathbb{P}(T \leq |t|) \\ \Leftrightarrow \frac{\alpha_0}{2} &\geq \mathbb{P}(T \geq |t|) \\ \Leftrightarrow \alpha_0 &\geq 2\mathbb{P}(T \geq |t|). \end{aligned} \quad (27)$$

(7) Analyse der Powerfunktion

Wir betrachten die Testgütefunktion

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \psi(k_{\alpha_0}; d_\mu, n - 1) + \psi(-k_{\alpha_0}; d_\mu, n - 1) \quad (28)$$

bei kontrolliertem Testumfang, also für $k_{\alpha_0}(n) := \psi^{-1}(1 - \alpha_0/2; n - 1)$ mit festem α_0 als Funktion des Nichtzentralitätsparameters und des Stichprobenumfangs. Namentlich hängt hier k_{α_0} auch von n ab.

Es ergibt sich die bivariate reellwertige Funktion

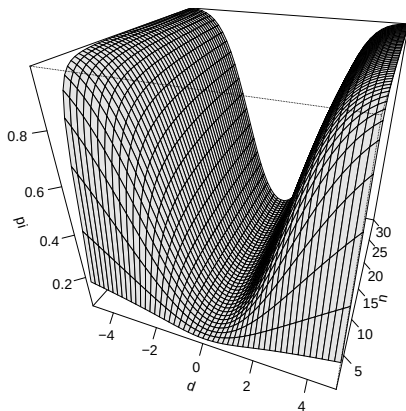
$$\pi : \mathbb{R} \times \mathbb{N} \rightarrow [0, 1], (d, n) \mapsto \pi(d, n) := 1 - \psi(k_{\alpha_0}; d, n - 1) + \psi(-k_{\alpha_0}; d, n - 1) \quad (29)$$

Bei festgelegten α_0 hängt die Powerfunktion des zweiseitigen T-Tests mit einfacher Nullhypothese also vom unbekannten Wert d und von der Stichprobengröße n ab. Wir visualisieren diese Abhängigkeiten untenstehend.

Einstichproben-T-Test mit $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$

(7) Analyse der Powerfunktion

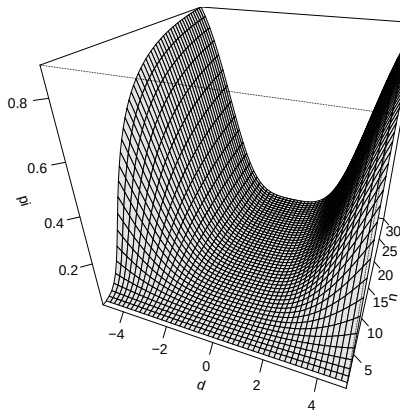
Powerfunktion für $\alpha_0 = 0.05$



Einstichproben-T-Test mit $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$

(7) Analyse der Powerfunktion

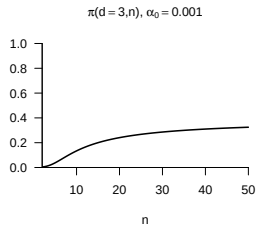
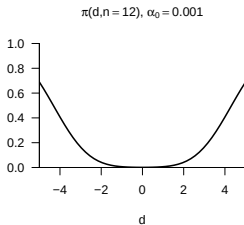
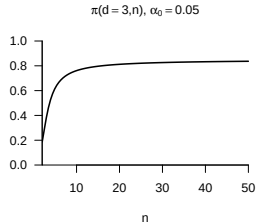
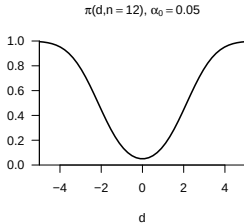
Powerfunktion für $\alpha_0 = 0.001$



Einstichproben-T-Test mit $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$

(7) Analyse der Powerfunktion

Powerfunktionen für $\mu_0 = 0$



(7) Analyse der Powerfunktion

Praktisches Vorgehen

Mit größerem n steigt die Powerfunktion des Tests an

- Ein großer Stichprobenumfang ist besser als ein kleiner Stichprobenumfang.
- Kosten für die Erhöhung des Stichprobenumfangs werden aber nicht berücksichtigt.

⇒ Die Theorie statistischer Hypothesentests ist nicht besonders lebensnah.

Die Powerfunktion hängt vom wahren, aber unbekannten, Parameterwert $d = \sqrt{n}(\mu - \mu_0)/\sigma$ ab.

⇒ Wenn man d schon kennen würde, würde man den Test nicht durchführen.

Generell wird folgendes Vorgehen favorisiert

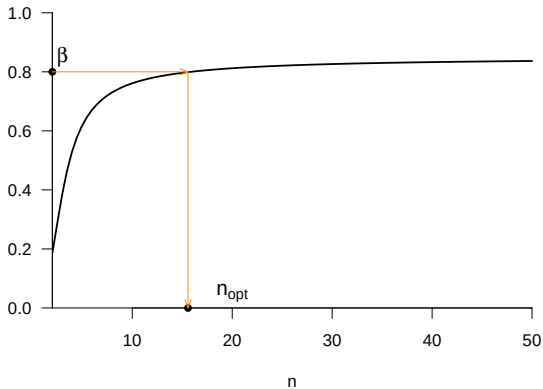
- Man legt das Signifikanzniveau α_0 fest und evaluiert die Powerfunktion.
- Man wählt einen Mindestparameterwert d^* , den man mit $\pi(d, n) = \beta$ detektieren möchte.
- Ein konventioneller Wert ist $\beta = 0.8$.
- Man liest die für $\pi(d = d^*, n) = \beta$ nötige Stichprobengröße n ab.

Einstichproben-T-Test mit $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$

(7) Analyse der Powerfunktion

Praktisches Vorgehen

$\pi(d=3, n)$ für $\alpha_0 = 0.05$, $\mu_0 = 0$



T-Tests

- T-Teststatistik
- **Einstichproben-T-Tests**
 - $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$
 - $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$
- Zweistichproben-T-Tests
 - Unabhängige Stichproben
 - Abhängige Stichproben

(1) Statistisches Modell in klassischer Form

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ sei die Stichprobe eines Normalverteilungsmodells mit unbekanntem Erwartungswertparameter μ und unbekanntem Varianzparameter $\sigma^2 > 0$. Als Parameter von Interesse betrachten wir $\theta = \mu$, so dass sich der Parameterraum von Interesse zu $\Theta = \mathbb{R}$ ergibt.

(2) Statistisches Modell in generativer Form

Es sei

$$X_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ für } i = 1, \dots, n \quad (30)$$

wobei

- $X_i, i = 1, \dots, n$ beobachtbare Zufallsvariablen,
- $\mu \in \mathbb{R}$ den festen identischen Erwartungswertparameter über Stichprobenvariablen, und
- $\varepsilon_i, i = 1, \dots, n$ unabhängige normalverteilte nicht-beobachtbare Zufallsvariablen

bezeichnen.

Einstichproben-T-Test mit $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$

(3) Testhypothesen, Teststatistik, Test

Für ein $\mu_0 \in \mathbb{R}$ betrachten wir die zusammengesetzte Nullhypothese

$$H_0 : \mu \leq \mu_0 \Leftrightarrow \Theta_0 :=] - \infty, \mu_0] \quad (31)$$

und die zusammengesetzte Alternativhypothese

$$H_1 : \mu > \mu_0 \Leftrightarrow \Theta_1 :=]\mu_0, \infty[. \quad (32)$$

Weiterhin betrachten wir die T-Teststatistik

$$T := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right) \quad (33)$$

und definieren den einseitigen kritischen Wert-basierten Test

$$\phi(X) := 1_{\{T \geq k\}} = \begin{cases} 1 & T \geq k \\ 0 & T < k \end{cases}. \quad (34)$$

(4) Analyse der Testgütefunktion

Theorem (Testgütefunktion)

ϕ sei der im obigen Testszenario definierte Test. Dann ist die Testgütefunktion von ϕ gegeben durch

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \psi(k; d_\mu, n - 1) \quad (35)$$

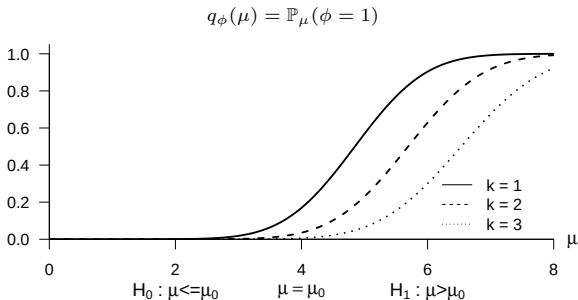
wobei $\psi(\cdot; d_\mu, n - 1)$ die KVF der nichtzentralen t -Verteilung mit Nichtzentralitätsparameter

$$d_\mu := \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (36)$$

und Freiheitsgradparameter $n - 1$ bezeichnet.

(4) Analyse der Testgütefunktion

Testgütefunktion q_ϕ für $\sigma^2 = 9$, $\mu_0 = 4$, $n = 12$ und $k = 1, 2, 3$.



(4) Analyse der Testgütefunktion

Beweis

Die Testgütefunktion des betrachteten Test im vorliegenden Testscenario ist definiert als

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := \mathbb{P}_\mu(\phi = 1). \quad (37)$$

Da die Wahrscheinlichkeiten für $\phi = 1$ und dafür, dass die zugehörige Teststatistik im Ablehnungsbereich des Tests liegt gleich sind, benötigen wir die also zunächst die Verteilung der Teststatistik. Wir haben oben bereits gesehen, dass die T-Teststatistik

$$T := \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_n} \right) \quad (38)$$

unter der Annahme $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ nach einer nichtzentralen t -Verteilung $t(d, n-1)$ mit Nichtzentralitätsparameter

$$d = \sqrt{n} \left(\frac{\mu - \mu_0}{\sigma} \right) \quad (39)$$

verteilt ist. Mit dem Ablehnungsbereich des einseitigen T-Tests $A :=]k, \infty[$ ergibt sich dann

$$q_\phi(\mu) = \mathbb{P}_\mu(\phi = 1) = \mathbb{P}_\mu(T \in]k, \infty[) = \mathbb{P}_\mu(T \geq k) = 1 - \mathbb{P}_\mu(T \leq k) = 1 - \psi(k; d, n-1), \quad (40)$$

wobei $\psi(\cdot; d, n-1)$ weiterhin die KVF der nichtzentralen T-Verteilung mit Nichtzentralitätsparameter d und Freiheitsgradparameter $n-1$ bezeichnet.

□

(5) Testumfangkontrolle

Theorem (Testumfangkontrolle)

ϕ sei der im obigen Szenario definierte Test. Dann ist ϕ ein Level- α_0 -Test mit Testumfang α_0 , wenn der kritische Wert k definiert ist durch

$$k_{\alpha_0} := \psi^{-1}(1 - \alpha_0; n - 1), \quad (41)$$

wobei $\psi^{-1}(\cdot; n - 1)$ die inverse KVF der t -Verteilung mit $n - 1$ Freiheitsgraden ist.

(5) Testumfangkontrolle

Beweis

Damit der betrachtete Test ein Level- α_0 -Test ist, muss bekanntlich $q_\phi(\mu) \leq \alpha_0$ für alle $\mu \in]-\infty, \mu_0]$ gelten. Weiterhin ist der Testumfang des betrachteten Tests durch $\alpha = \max_{-\infty, \mu_0]} q_\phi(\mu)$ gegeben. Mit der Monotonie von q_ϕ , die wir hier als selbstevident voraussetzen, gilt $\alpha = q_\phi(\mu_0)$. Wir müssen also zeigen, dass die Wahl von k_{α_0} garantiert, dass ϕ ein Level- α_0 -Test mit Testumfang α_0 ist. Dazu merken wir zunächst an, dass für $\mu = \mu_0$ gilt, dass

$$\begin{aligned} q_\phi(\mu_0) &= 1 - \psi(k; d_{\mu_0}, n - 1) \\ &= 1 - \psi(k; 0, n - 1) \\ &= 1 - \psi(k; n - 1) \end{aligned} \tag{42}$$

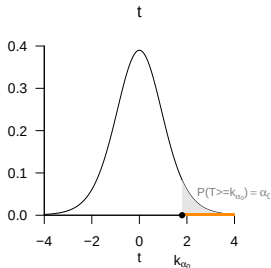
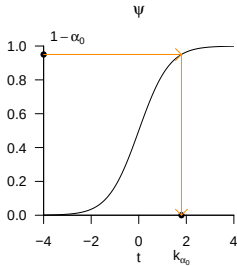
wobei $\psi(\cdot; d, n - 1)$ und $\psi(\cdot; n - 1)$ die KVF der nichtzentralen t -Verteilung mit Nichtzentralitätsparameter d und Freiheitsgradparameter $n - 1$ sowie der t -Verteilung mit Freiheitsgradparameter $n - 1$, respektive, bezeichnen. Sei nun also $k := k_{\alpha_0}$. Dann gilt

$$\begin{aligned} q_\phi(\mu_0) &= 1 - \psi(k_{\alpha_0}) \\ &= 1 - \psi\left(\psi^{-1}(1 - \alpha_0)\right) \\ &= 1 - 1 + \alpha_0 \\ &= \alpha_0, \end{aligned} \tag{43}$$

Es folgt also direkt, dass bei der Wahl von $k = k_{\alpha_0}$, $q_\phi(\mu_0) \leq \alpha_0$ ist und der betrachtete Test somit ein Level- α_0 -Test ist. Weiterhin folgt direkt, dass der Testumfang des betrachteten Tests bei der Wahl von $k = k_{\alpha_0}$ gleich α_0 ist.

(5) Testumfangkontrolle

Wahl von $k_{\alpha_0} := \psi^{-1}(1 - \alpha_0; n - 1)$ mit $n = 12$ $\alpha_0 := 0.05$ und Ablehnungsbereich



(5) Testumfangkontrolle

Praktisches Vorgehen

- Man nimmt an, dass ein vorliegender Datensatz $x_1, \dots, x_n$ eine Realisation von $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit unbekannten Parametern μ und $\sigma^2 > 0$ ist.
- Man möchte entscheiden ob für ein $\mu_0 \in \mathbb{R}$ eher $H_0 : \mu \leq \mu_0$ oder $H_1 : \mu > \mu_0$ zutrifft.
- Man wählt ein Signifikanzniveau α_0 und bestimmt den zugehörigen Freiheitsgradparameter-abhängigen kritischen Wert k_{α_0} . Zum Beispiel gilt bei Wahl von $\alpha_0 := 0.05$ und $n = 12$, also Freiheitsgradparameter 11, dass $k_{0.05} = \psi^{-1}(1 - 0.05; 11) \approx 1.80$ ist.
- Anhand von $n, \mu_0, \bar{x}_n$ und s_n berechnet man die Realisierung der T-Teststatistik

$$t := \sqrt{n} \left(\frac{\bar{x}_n - \mu_0}{s_n} \right) \quad (44)$$

- Wenn t größer-gleich k_{α_0} ist, lehnt man die Nullhypothese ab, andernfalls nicht.
- Die oben entwickelte Theorie garantiert dann, dass man in höchstens $\alpha_0 \cdot 100$ von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.

(6) p-Werte

Bestimmung des p-Wertes

- Per Definition ist der p-Wert das kleinste Signifikanzlevel α_0 , bei welchem man die Nullhypothese basierend auf einem vorliegendem Wert der Teststatistik ablehnen würde.
- Bei $T = t$ würde H_0 für jedes α_0 mit $t \geq \psi^{-1}(1 - \alpha_0; n - 1)$ abgelehnt werden. Für diese α_0 gilt, wie unten gezeigt,

$$\alpha_0 \geq \mathbb{P}(T \geq t). \quad (45)$$

- Das kleinste $\alpha_0 \in [0, 1]$ mit $\alpha_0 \geq \mathbb{P}(T \geq t)$ ist dann $\alpha_0 = \mathbb{P}(T \geq t)$, also folgt

$$\text{p-Wert} = \mathbb{P}(T \geq t) = 1 - \psi(t; n - 1). \quad (46)$$

- Im Gegensatz zum zweiseitigen T-Tests ergeben sich für negative Werte von t sehr große p-Werte.
- Zum Beispiel ist für $T = 2.00$ und $n = 10$ der p-Wert 0.038, für $T = -2.00$ und $n = 10$ ist der p-Wert dagegen 0.961.

(6) p-Werte

Bestimmung des p-Wertes

- Es bleibt zu zeigen, dass gilt

$$t \geq \psi^{-1}(1 - \alpha_0; n - 1) \Leftrightarrow \alpha_0 \geq 2\mathbb{P}(T \geq t) \quad (47)$$

- Dies aber folgt aus

$$\begin{aligned} t &\geq \psi^{-1}(1 - \alpha_0; n - 1) \\ \Leftrightarrow \psi(t; n - 1) &\geq \psi\left(\psi^{-1}(1 - \alpha_0; n - 1); n - 1\right) \\ \Leftrightarrow \psi(t; n - 1) &\geq 1 - \alpha_0 \\ \Leftrightarrow \mathbb{P}(T \leq t) &\geq 1 - \alpha_0 \\ \Leftrightarrow \alpha_0 &\geq 1 - \mathbb{P}(T \leq t) \\ \Leftrightarrow \alpha_0 &\geq \mathbb{P}(T \geq t) \end{aligned} \quad (48)$$

(7) Analyse der Powerfunktion

Wir betrachten die Testgütefunktion

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \psi(k_{\alpha_0}; d_\mu, n - 1) \quad (49)$$

bei kontrolliertem Testumfang, also für $k_{\alpha_0} := \psi^{-1}(1 - \alpha_0; n - 1)$ mit festem α_0 als Funktion des Nichtzentralitätsparameters und des Stichprobenumfangs. Namentlich hängt hier k_{α_0} auch von n ab.

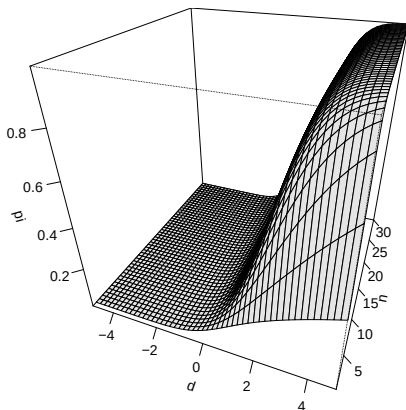
Es ergibt sich die bivariate reellwertige Funktion

$$\pi : \mathbb{R} \times \mathbb{N} \rightarrow [0, 1], (d, n) \mapsto \pi(\mu, n) := 1 - \psi(k_{\alpha_0}; d, n - 1) \quad (50)$$

Bei festgelegten α_0 hängt die Powerfunktion des einseitigen T-Tests mit zusammengesetzter Nullhypothese also vom unbekannten Wert d und von der Stichprobengröße n ab. Wir visualisieren diese Abhängigkeiten untenstehend.

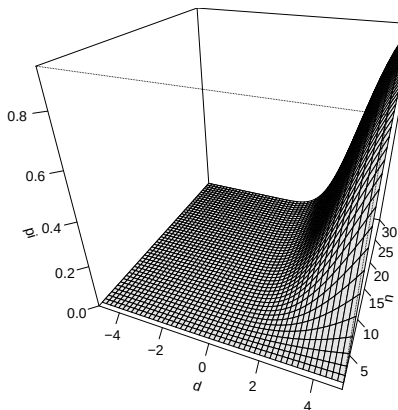
Einstichproben-T-Test mit $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$

Powerfunktion für $\alpha_0 = 0.05$



Einstichproben-T-Test mit $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$

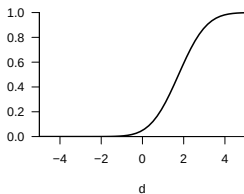
Powerfunktion für $\alpha_0 = 0.001$



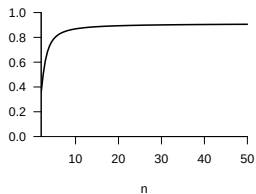
Einstichproben-T-Test mit $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$

Powerfunktionen für $\mu_0 = 0$

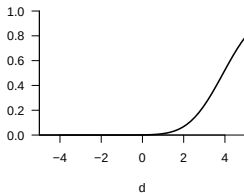
$\pi(d, n = 12), \alpha_0 = 0.05$



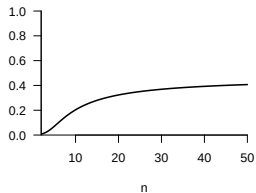
$\pi(d = 3, n), \alpha_0 = 0.05$



$\pi(d, n = 12), \alpha_0 = 0.001$



$\pi(d = 3, n), \alpha_0 = 0.001$



(7) Analyse der Powerfunktion

Praktisches Vorgehen

Mit größerem n steigt die Powerfunktion des Tests an

- Ein großer Stichprobenumfang ist besser als ein kleiner Stichprobenumfang.
- Kosten für die Erhöhung des Stichprobenumfangs werden aber nicht berücksichtigt.

⇒ Die Theorie statistischer Hypothesentests ist nicht besonders lebensnah.

Die Powerfunktion hängt vom wahren, aber unbekannten, Parameterwert $d = \sqrt{n}(\mu - \mu_0)/\sigma$ ab.

⇒ Wenn man d schon kennen würde, würde man den Test nicht durchführen.

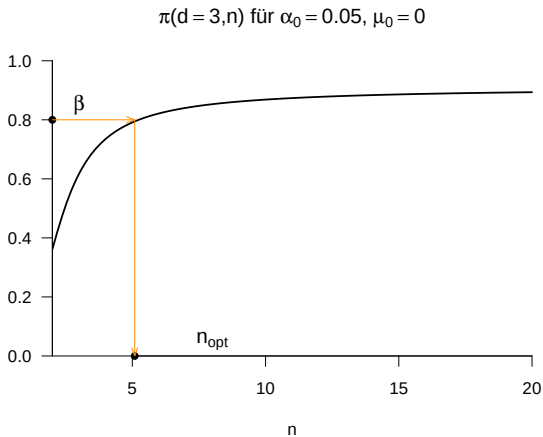
Generell wird folgendes Vorgehen favorisiert

- Man legt das Signifikanzniveau α_0 fest und evaluiert die Powerfunktion.
- Man wählt einen Mindestparameterwert d^* , den man mit $\pi(d, n) = \beta$ detektieren möchte.
- Ein konventioneller Wert ist $\beta = 0.8$.
- Man liest die für $\pi(d = d^*, n) = \beta$ nötige Stichprobengröße n ab.

Einstichproben-T-Test mit $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$

(7) Analyse der Powerfunktion

Praktisches Vorgehen



T-Tests

- T-Teststatistik
- Einstichproben-T-Tests
 - $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$
 - $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$
- Zweistichproben-T-Tests
 - **Unabhängige Stichproben**
 - Abhängige Stichproben
- Selbstkontrollfragen

Zweistichproben-T-Test bei unabhängigen Stichproben

Anwendungsszenario

- **Zwei Stichproben** experimenteller Einheiten
- Annahme unabhängiger identischer Normalverteilungen $N(\mu_1, \sigma_1^2)$ und $N(\mu_2, \sigma_2^2)$
- $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ unbekannt.
- Quantifizieren der Unsicherheit beim inferentiellen Vergleich von μ_1 mit μ_2 beabsichtigt.

Anwendungsbeispiele

- BDI Score Datenanalyse bei zwei Gruppen psychiatrischer Patient:innen
 - Gruppe 1 Psychoanalyse, Gruppe 2 Kognitive Verhaltenstherapie
 - $\mu_1 \neq \mu_2$? $\Leftrightarrow$ Welche Therapie ist besser?
- **Forcierte Schwimmtestdatenanalyse bei zwei Gruppen genmanipulierter Mäuse**
 - Gruppe 1 Wildtyp, Gruppe 2 Serotoninrezeptormutation
 - $\mu_1 \neq \mu_2$? $\Leftrightarrow$ Trägt Serotoninrezeptor zum Schwimmtestverhalten bei?

Modellszenarien

- Annahme identischer Varianzen
- Annahme eines bekannten Varianzverhältnisses
- Keine Annahmen zu Varianzen

Hypothesenszenarien

- $H_0 : \mu_1 = \mu_2$ und $H_1 : \mu_1 \neq \mu_2$
- $H_0 : \mu_1 \leq \mu_2$ und $H_1 : \mu_1 > \mu_2$
- $H_0 : \mu_1 \geq \mu_2$ und $H_1 : \mu_1 < \mu_2$

Wir betrachten hier exemplarisch $H_0 : \mu_1 = \mu_2$ und $H_1 : \mu_1 \neq \mu_2$

Modellszenarien

- **Annahme identischer Varianzen**
- Annahme eines bekannten Varianzverhältnisses
- Keine Annahmen zu Varianzen

Hypothesenszenarien

- $H_0 : \mu_1 = \mu_2$ und $H_1 : \mu_1 \neq \mu_2$
- $H_0 : \mu_1 \leq \mu_2$ und $H_1 : \mu_1 > \mu_2$
- $H_0 : \mu_1 \geq \mu_2$ und $H_1 : \mu_1 < \mu_2$

(1) Statistisches Modell in klassischer Form

$X_{11}, \dots, X_{1n_1} \sim N(\mu_1, \sigma^2)$ sei eine Stichprobe eines Normalverteilungsmodells mit unbekanntem Erwartungswertparameter μ_1 und unbekanntem Varianzparameter $\sigma^2 > 0$. $X_{21}, \dots, X_{2n_2} \sim N(\mu_2, \sigma^2)$ sei eine weitere Stichprobe eines Normalverteilungsmodells mit unbekanntem Erwartungswertparameter μ_2 und unbekanntem Varianzparameter $\sigma^2 > 0$. Die Varianzparameter beider Stichproben werden also als identisch vorausgesetzt. Der Parameter von Interesse ist (μ_1, μ_2) , der Parameterraum des Modells ist $\Theta := \mathbb{R}^2$.

(2) Statistisches Modell in generativer Form

Es sei

$$X_{ij} = \mu_i + \varepsilon_{ij} \text{ mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ für } i = 1, 2 \text{ und } j = 1, \dots, n_i \quad (51)$$

wobei

- i die Stichproben indiziert,
- j die experimentellen Einheiten indiziert,
- n_i die Stichprobengrößen sind,
- X_{ij} beobachtbare Zufallsvariablen sind,
- μ_i feste Erwartungswertparameter der Stichprobenvariablen sind, und
- ε_{ij} unabhängige normalverteilte nicht-beobachtbare Zufallsvariablen sind.

(3) Testhypothesen, Teststatistik, Test

Wir betrachten die einfache Nullhypothese

$$H_0 : \mu_1 = \mu_2 \Leftrightarrow \Theta_0 := \{(\mu_1, \mu_2) \in \mathbb{R}^2 \mid \mu_1 = \mu_2\} \quad (52)$$

und die zusammengesetzte Alternativhypothese

$$H_1 : \mu_1 \neq \mu_2 \Leftrightarrow \Theta_1 := \{(\mu_1, \mu_2) \in \mathbb{R}^2 \mid \mu_1 \neq \mu_2\}. \quad (53)$$

Weiterhin betrachten wir die *Zweistichproben-T-Teststatistik*

$$T := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{X}_1 - \bar{X}_2}{S_{12}} \right) \quad (54)$$

wobei für $i = 1, 2$

$$\bar{X}_i := \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \text{ und } S_{12} := \sqrt{\frac{\sum_{j=1}^{n_1} (X_{1j} - \bar{X}_1)^2 + \sum_{j=1}^{n_2} (X_{2j} - \bar{X}_2)^2}{(n_1 - 1) + (n_2 - 1)}} \quad (55)$$

die Stichprobenmittelwerte und die *gepoolte Stichprobenstandardabweichung*, respektive, bezeichnen. Schließlich definieren wir den zweiseitigen kritischen Wert-basierten Test

$$\phi(X) := 1_{\{|T| \geq k\}}. \quad (56)$$

(4) Analyse der Testgütefunktion

Theorem (Testgütefunktion)

Es sei ϕ der im obigen Modell formulierte Zweistichproben-T-Test. Dann ist die Testgütefunktion von ϕ gegeben durch

$$\begin{aligned} q_\phi : \mathbb{R}^2 &\rightarrow [0, 1], (\mu_1, \mu_2) \mapsto q_\phi(\mu_1, \mu_2) \\ &:= 1 - \psi(k; d_{\mu_1, \mu_2}, n_1 + n_2 - 2) + \psi(-k; d_{\mu_1, \mu_2}, n_1 + n_2 - 2) \end{aligned} \quad (57)$$

wobei $\psi(\cdot; d_{\mu_1, \mu_2}, n_1 + n_2 - 2)$ die KVF der nichtzentralen t -Verteilung mit Nichtzentralitätsparameter

$$d_{\mu_1, \mu_2} := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\mu_1 - \mu_2}{\sigma} \quad (58)$$

und Freiheitsgradparameter $n_1 + n_2 - 2$ bezeichnet.

Bemerkungen

- q_ϕ ist eine bivariate reellwertige Funktion.
- q_ϕ kann alternativ als univariate reellwertige Funktion von $\Delta := \mu_1 - \mu_2$ konzipiert werden.
- Im Vergleich zum Einstichprobenszenario gelten

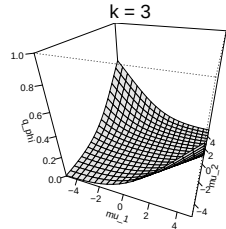
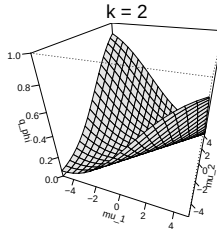
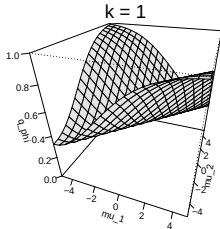
$$n \hookrightarrow n_1 + n_2 - 2, \quad \sqrt{n} \hookrightarrow \sqrt{\frac{n_1 n_2}{n_1 + n_2}}, \quad \mu - \mu_0 \hookrightarrow \mu_1 - \mu_2 \quad (59)$$

- Für einen Beweisansatz, siehe DeGroot and Schervish (2012) Seite 591.

(4) Analyse der Testgütefunktion

Testgütefunktion q_ϕ für $\sigma^2 = 9, n_1 = 12, n_2 = 12$.

$$q_\phi(\mu) = \mathbb{P}_\mu(\phi = 1)$$



(5) Testumfangkontrolle

Theorem (Testumfangkontrolle)

ϕ sei der im obigen Testszenario definierte Test. Dann ist ϕ ein Level- α_0 -Test mit Testumfang α_0 , wenn der kritische Wert definiert ist durch

$$k_{\alpha_0} := \psi^{-1} \left(1 - \frac{\alpha_0}{2}; n_1 + n_2 - 2 \right), \quad (60)$$

wobei $\psi^{-1}(\cdot; n_1 + n_2 - 2)$ die inverse KVF der t -Verteilung mit $n_1 + n_2 - 2$ Freiheitsgraden ist.

Bemerkungen

- Das Resultat folgt in Analogie zum Einstichproben-T-Test.
- Im Vergleich zum Einstichproben-T-Testfall gilt lediglich

$$n - 1 \hookrightarrow n_1 + n_2 - 2. \quad (61)$$

(5) Testumfangkontrolle

Praktisches Vorgehen

- Man nimmt an, dass zwei vorliegende Datensätze $x_{11}, \dots, x_{1n_1}$ und $x_{21}, \dots, x_{2n_2}$ Realisationen von $X_{11}, \dots, X_{1n_1} \sim N(\mu_1, \sigma^2)$ und $X_{21}, \dots, X_{2n_2} \sim N(\mu_2, \sigma^2)$ mit unbekannten Parametern μ_1, μ_2, σ^2 sind.
- Man möchte entscheiden, ob eher $H_0 : \mu_1 = \mu_2$ oder $H_1 : \mu_1 \neq \mu_2$ zutrifft.
- Man wählt ein Signifikanzniveau α_0 und bestimmt den zugehörigen Freiheitsgradparameter-abhängigen kritischen Wert k_{α_0} . Zum Beispiel gilt bei Wahl von $\alpha_0 := 0.05$ und $n_1 = 12, n_2 = 12$, also Freiheitsgradparameter $12+12-2 = 22$, dass $k_{0.05} = \psi^{-1}(1 - 0.05/2; 22) \approx 2.07$ ist.
- Anhand von $n_1, n_2, \bar{x}_1, \bar{x}_2$ und der gepoolten Stichprobenstandardabweichung s_{12} berechnet man die Realisierung der Zweistichproben-T-Teststatistik

$$t := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left(\frac{\bar{x}_1 - \bar{x}_2}{s_{12}} \right) \quad (62)$$

- Wenn t größer-gleich k_{α_0} ist oder wenn t kleiner-gleich $-k_{\alpha_0}$ ist, lehnt man die Nullhypothese ab, andernfalls lehnt man sie nicht ab.
- Die oben entwickelte Theorie des Zweistichproben-T-Tests garantiert dann, dass man in höchstens $\alpha_0 \cdot 100$ von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.

(6) p-Werte

Bestimmung des p-Wertes

- Per Definition ist der p-Wert das kleinste Signifikanzlevel α_0 , bei welchem man die Nullhypothese basierend auf einem vorliegendem Wert der Teststatistik ablehnen würde.
- Bei $T = t$ würde H_0 für jedes α_0 mit $|t| \geq \psi^{-1}(1 - \alpha_0/2; n_1 + n_2 - 2)$ abgelehnt werden. Für diese α_0 gilt, wie bereits mehrfach gezeigt,

$$\alpha_0 \geq 2\mathbb{P}(T \geq |t|). \quad (63)$$

- Das kleinste $\alpha_0 \in [0, 1]$ mit $\alpha_0 \geq 2\mathbb{P}(T \geq |t|)$ ist dann $\alpha_0 = 2\mathbb{P}(T \geq |t|)$, also folgt

$$\text{p-Wert} = 2\mathbb{P}(T \geq |t|) = 2(1 - \psi(|t|; n_1 + n_2 - 2)). \quad (64)$$

- Im Vergleich zum Einstichprobenfall gilt lediglich $n \mapsto n_1 + n_2 - 2$.

(7) Analyse der Powerfunktion

Wir betrachten die Testgütefunktion

$$\begin{aligned} q_\phi : \mathbb{R}^2 &\rightarrow [0, 1], (\mu_1, \mu_2) \mapsto q_\phi(\mu_1, \mu_2) \\ &:= 1 - \psi(k; d_{\mu_1, \mu_2}, n_1 + n_2 - 2) + \psi(-k; d_{\mu_1, \mu_2}, n_1 + n_2 - 2) \end{aligned} \quad (65)$$

als Funktion des Nichtzentralitätsparameters und der Summe der Stichprobenumfänge $n := n_1 + n_2$ bei kontrolliertem Testumfang, also für $k_{\alpha_0} := \psi^{-1}(1 - \alpha_0/2; n - 2)$ mit festem α_0 .

Es ergibt sich die multivariate reellwertige Funktion

$$\begin{aligned} \pi : \mathbb{R} \times \mathbb{N}^2 &\rightarrow [0, 1], (d, n) \mapsto \\ \pi(d, n) &:= 1 - \psi(k_{\alpha_0}; d, n - 2) + \psi(-k_{\alpha_0}; d, n - 2) \end{aligned} \quad (66)$$

Bei festgelegten α_0 hängt die Powerfunktion des zweiseitigen T-Tests mit einfacher Nullhypothese also vom unbekannten Wert d und von der Summe der Stichprobengrößen n ab. De-facto handelt es sich also um die gleiche Powerfunktion wie beim zweiseitigen Einstichproben-T-Test mit dem einzigen Unterschied, dass für den Freiheitsgradparameter $n - 2$ anstelle von $n - 2$ gilt. Wir verzichten auf eine erneute Visualisierung.

(7) Analyse der Powerfunktion

Praktisches Vorgehen

Mit größerem $n = n_1 + n_2$ steigt die Powerfunktion des Tests an

- Ein großer Stichprobenumfang ist besser als ein kleiner Stichprobenumfang.
- Kosten für die Erhöhung des Stichprobenumfangs werden aber nicht berücksichtigt.
- Ungleichgewichte zwischen n_1 und n_2 werden durch die Tatsache ausglich, dass Datenpunkte einer Stichproben auch zur Varianzschätzung in der anderen Stichprobe beitragen, da eine identische Varianz vorausgesetzt wurde.

Die Powerfunktion hängt vom wahren, aber unbekannten, Parameterwert $d = \sqrt{n}(\mu_1 - \mu_0)/\sigma$ ab.

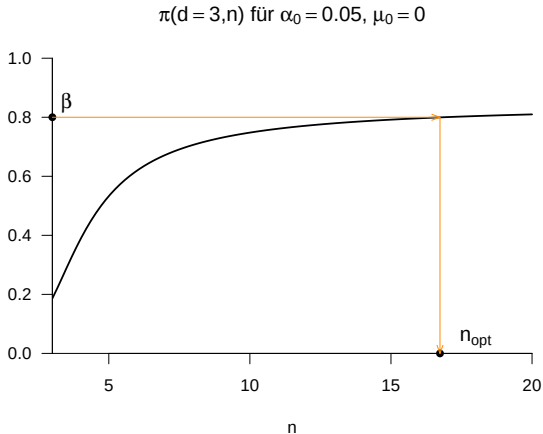
⇒ Wenn man d schon kennen würde, würde man den Test nicht durchführen.

Generell wird folgendes Vorgehen favorisiert

- Man legt das Signifikanzniveau α_0 fest und evaluiert die Powerfunktion.
- Man wählt einen Mindestparameterwert d^* , den man mit $\pi(d, n) = \beta$ detektieren möchte.
- Ein konventioneller Wert ist $\beta = 0.8$.
- Man liest die für $\pi(d = d^*, n) = \beta$ nötige Stichprobengröße n ab.

(7) Analyse der Powerfunktion

Praktisches Vorgehen



Modellszenarien

- Annahme identischer Varianzen
- **Annahme eines bekannten Varianzverhältnisses**
- Keine Annahmen zu Varianzen

Hypothesenszenarien

- $H_0 : \mu_1 = \mu_2$ und $H_1 : \mu_1 \neq \mu_2$
- $H_0 : \mu_1 \leq \mu_2$ und $H_1 : \mu_1 > \mu_2$
- $H_0 : \mu_1 \geq \mu_2$ und $H_1 : \mu_1 < \mu_2$

Überblick

- Die obige Theorie kann für eine bekannte Konstante $k > 0$ auf den Fall des statistischen Modells

$$X_{11}, \dots, X_{1n_1} \sim N(\mu_1, \sigma_1^2) \text{ und } X_{21}, \dots, X_{2n_2} \sim N(\mu_2, k\sigma_1^2) \quad (67)$$

erweitert werden.

- In diesem Fall hat die Zweistichproben-T-Teststatistik definiert durch (cf. DeGroot and Schervish (2012), S. 593)

$$T := \frac{(n_1 + n_2 - 2)^{\frac{1}{2}} (\bar{X}_1 - \bar{X}_2)}{\left(\frac{1}{n_1} + \frac{k}{n_2}\right)^{\frac{1}{2}} \left(S_1^2 + \frac{S_2^2}{k}\right)^{\frac{1}{2}}} \quad (68)$$

mit den jeweiligen Stichprobenmittelwerten $\bar{X}_1$ und $\bar{X}_2$, sowie den Stichprobenvarianzen S_1^2 und S_2^2 eine nichtzentrale t -Verteilung mit Freiheitsgradparameter $n_1 + n_2 - 2$.

- Wir wollen diesen Fall hier nicht weiter vertiefen.

Modellszenarien

- Annahme identischer Varianzen
- Annahme eines bekannten Varianzverhältnisses
- **Keine Annahmen zu Varianzen**

Hypothesenszenarien

- $H_0 : \mu_1 = \mu_2$ und $H_1 : \mu_1 \neq \mu_2$
- $H_0 : \mu_1 \leq \mu_2$ und $H_1 : \mu_1 > \mu_2$
- $H_0 : \mu_1 \geq \mu_2$ und $H_1 : \mu_1 < \mu_2$

Überblick

- Wenn im Rahmen des statistischen Modells

$$X_{11}, \dots, X_{1n_1} \sim N(\mu_1, \sigma_1^2) \text{ und } X_{21}, \dots, X_{2n_2} \sim N(\mu_2, \sigma_2^2) \quad (69)$$

alle vier Parameter $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ sowie das Varianzverhältnis $k = \sigma_2^2 / \sigma_1^2$ als unbekannt vorausgesetzt werden, **kann obige Theorie nicht auf diesen Fall angewendet werden**. Das Problem der statistischen Inferenz wird in diesem Fall als **Behrens-Fisher Problem** bezeichnet.

- Eine populäre Methode zur statistischen Inferenz im Behrens-Fisher Problem geht auf Welch (1938) zurück. Diese Methode bedient sich der Teststatistik

$$\tilde{T} := (\bar{X}_1 - \bar{X}_2) / \left(\left(\frac{S_1^2}{n_1(n_1 - 1)} + \frac{S_2^2}{n_2(n_2 - 1)} \right)^{1/2} \right), \quad (70)$$

wobei S_1^2 und S_2^2 die Stichprobenvarianzen bezeichnen. Die Verteilung von $\tilde{T}$ ist unbekannt, kann aber mit einer (nichtzentralen) t -Verteilung approximiert werden.

- Der Freiheitsgradparameter dieser approximierenden (nichtzentralen) t -Verteilung ergibt sich dabei mithilfe der beobachteten Werte s_1^2 und s_2^2 von S_1^2 und S_2^2 , respektive, zu

$$\nu = \left(\frac{s_1^2}{n_1(n_1 - 1)} + \frac{s_2^2}{n_2(n_2 - 1)} \right) / \left(\frac{1}{(n_1 - 1)^3} \left(\frac{s_1^2}{n_1} \right)^2 + \frac{1}{(n_2 - 1)^3} \left(\frac{s_2^2}{n_2} \right)^2 \right). \quad (71)$$

- Wir wollen die theoretische Begründung dieses Vorgehens hier nicht weiter vertiefen.

T-Tests

- T-Teststatistik
- Einstichproben-T-Tests
 - $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$
 - $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$
- Zweistichproben-T-Tests
 - Unabhängige Stichproben
 - **Abhängige Stichproben**
- Selbstkontrollfragen

Zweistichproben-T-Test bei abhängigen Stichproben

Anwendungsszenario

- **Zwei Messungen** an einer Gruppe experimenteller Einheiten.
- Annahme “abhängiger” Normalverteilungen $N(\mu_1, \sigma_1^2)$ und $N(\mu_2, \sigma_2^2)$
- $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ unbekannt.
- Quantifizieren der Unsicherheit beim inferentiellen Vergleich von μ_1 mit μ_2 beabsichtigt.

Anwendungsbeispiele

- Tense Arousal Datenanalyse bei einer Filmclipbedingung
 - Erste Messung vor Filmclipexposition, zweite Messung nach Filmclipexposition
 - $\mu_1 \neq \mu_2?$ $\Leftrightarrow$ Hat Filmclipexposition einen Einfluss auf TA?
- Reaktionszeitdatenanalyse zur visuellen Wahrnehmung
 - Bedingung 1: Hoher Kontrast, Bedingung 2: Niedriger Kontrast
 - $\mu_1 \neq \mu_2?$ $\Leftrightarrow$ Hat Stimuluskontrast Einfluss auf perzeptuelle Entscheidungen?

Zweistichproben-T-Test bei abhängigen Stichproben

(1) Statistisches Modell in generativer Form

Es sei

$$X_{ij} = \mu_i + \beta_j + \varepsilon_{ij} \text{ mit } \beta_j \sim N(0, \sigma_\beta^2) \text{ und } \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \text{ für } i = 1, 2, j = 1, \dots, n, \quad (72)$$

wobei

- i die Messungen (abhängige Stichproben) und j die experimentellen Einheiten indizieren,
- X_{ij} beobachtbare Zufallsvariablen sind,
- μ_i feste Erwartungswertparameter der Messungen sind,
- β_j unabhängige normalverteilte nicht-beobachtbare Zufallsvariablen sind, und
- ε_{ij} unabhängige normalverteilte nicht-beobachtbare Zufallsvariablen sind.

Alle β_i und ε_{ij} werden als voneinander unabhängig vorausgesetzt.

- Die X_{ij} modellieren Datenpunkte,
- die μ_i modellieren feste Messungs-abhängige Effekte,
- die β_j modellieren zufällige experimentelle Einheits-abhängige Effekte,
- die ε_{ij} modellieren (wie immer) zufällige Störeffekte.

Lineare Modelle mit festen Effekten (fixed effects) und zufälligen Effekten (random effects) heißen gemischte Modelle (linear mixed models).

Zweistichproben-T-Test bei abhängigen Stichproben

(1) Statistisches Modell in generativer Form

Datengeneration

```
# Modellspezifikation
n          = 12
mu         = c(10,12)
sigsqr_beta = 5
sigsqr_veps = 1

# Stichprobengröße
# feste Effektparameter      \mu_1, \mu_2
# zufälliger Effektparameter \sigsgqr_\beta
# Störvariablenparameter    \sigqr_\varepsilon

# Datengeneration
X          = data.frame(
  mu_1      = rep(NaN,n),
  mu_2      = rep(NaN,n),
  beta_j    = rep(NaN,n),
  veps_1_j  = rep(NaN,n),
  veps_2_j  = rep(NaN,n),
  X_1j      = rep(NaN,n),
  X_2j      = rep(NaN,n)

  # Datensatzinitialisierung
  # \mu_1
  # \mu_2
  # \beta_j,
  # \varepsilon_{1j},
  # \varepsilon_{2j},
  # X_{1j}, j = 1,...,n
  # X_{2j}, j = 1,...,n

# Iterationen über experimentelle Einheiten
for(j in 1:n){
  beta_j      = rnorm(1,0,sqrt(sigsqr_beta)) # zufälliger experimenteller Einheits-abhängiger Effekt
  X[[j,3]]    = beta_j                      # \beta_j

  # Iterationen über Messungen
  for(i in 1:2){
    eps_ij    = rnorm(1,0,sqrt(sigsqr_veps)) # zufälliger Störvariablenwert
    X[[j,i]]  = mu[i]                       # \mu_i
    X[[j,3+i]] = eps_ij                     # \varepsilon_{ij}
    X[[j,5+i]] = mu[i] + beta_j + eps_ij    # X_{ij} = \mu_i + \beta_j + \varepsilon_{ij}
  }
}
```

Zweistichproben-T-Test bei abhängigen Stichproben

(1) Statistisches Modell in generativer Form

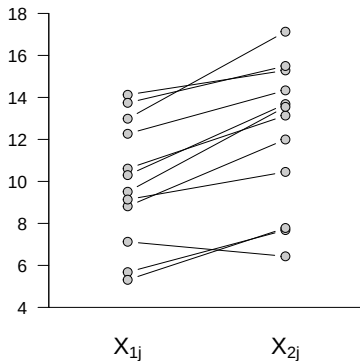
Datenbeispiel

mu_1	mu_2	beta_j	veps_1_j	veps_2_j	X_1j	X_2j
10	12	-0.792	0.796	1.136	10.00	12.34
10	12	-0.538	-0.951	0.166	8.51	11.63
10	12	-3.158	1.011	-0.571	7.85	8.27
10	12	-0.066	0.183	-1.227	10.12	10.71
10	12	0.537	-1.381	0.475	9.16	13.01
10	12	3.592	0.245	1.137	13.84	16.73
10	12	4.749	0.342	-0.610	15.09	16.14
10	12	-1.080	0.311	0.599	9.23	11.52
10	12	0.675	0.503	0.609	11.18	13.28
10	12	-1.062	-1.293	-0.362	7.64	10.58
10	12	2.490	0.432	-0.169	12.92	14.32
10	12	-0.458	-0.697	-0.488	8.85	11.05

Zweistichproben-T-Test bei abhängigen Stichproben

(1) Statistisches Modell in generativer Form

Datenbeispiel



Zweistichproben-T-Test bei abhängigen Stichproben

(1) Statistisches Modell in generativer Form

Theorem (Verteilung der Differenzvariablen)

Das oben formulierte statistische Modell eines Zweistichproben-T-Test bei abhängigen Stichproben sei gegeben und für $j = 1, \dots, n$ sei $X_j := X_{j2} - X_{j1}$ die Differenzvariable der Zufallsvariablen X_{j1} und X_{j2} . Dann gilt

$$X_j \sim N(\mu, \sigma^2) \text{ mit } \mu := \mu_2 - \mu_1 \text{ und } \sigma^2 := 2\sigma_\varepsilon^2. \quad (73)$$

Bemerkungen

- Hinsichtlich der Differenzvariablen X_j handelt es sich um das Einstichproben-T-Testscenario
- Die Hypothesen des Zweistichproben-T-Test bei abhängigen Stichproben hinsichtlich der Erwartungswertparameter μ_1 und μ_2 übertragen sich wie folgt auf die Einstichproben-T-Test Hypothesen in bezug auf den Erwartungswertparameter μ :

Zweistichproben-T-Test			Einstichproben-T-Test	
$H_0 : \mu_1 = \mu_2$	$H_1 : \mu_1 \neq \mu_2$	$\Rightarrow$	$H_0 : \mu = 0$	$H_1 : \mu \neq 0$
$H_0 : \mu_1 \leq \mu_2$	$H_1 : \mu_1 > \mu_2$	$\Rightarrow$	$H_0 : \mu \geq 0$	$H_1 : \mu < 0$
$H_0 : \mu_1 \geq \mu_2$	$H_1 : \mu_1 < \mu_2$	$\Rightarrow$	$H_0 : \mu \leq 0$	$H_1 : \mu > 0$

Zweistichproben-T-Test bei abhängigen Stichproben

(1) Statistisches Modell in generativer Form

Beweis

Wir halten zunächst fest, dass

$$\begin{aligned}X_j &= X_{2j} - X_{1j} \\&= \mu_2 + \beta_j + \varepsilon_{2j} - \mu_1 - \beta_j - \varepsilon_{1j} \\&= \mu_2 - \mu_1 + \varepsilon_{2j} - \varepsilon_{1j} \\&= \mu + \varepsilon_j,\end{aligned}\tag{74}$$

wobei wir $\mu := \mu_2 - \mu_1$ und $\varepsilon_j := \varepsilon_{2j} - \varepsilon_{1j}$ definiert haben. In Generalisierung des Summentransformationstheorems für normalverteilte Zufallsvariablen (cf. (6) Transformationen der Normalverteilung) halten wir weiterhin fest, dass für unabhängige Zufallsvariablen $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \dots, n$ und Koeffizienten a_i , $i = 1, \dots, n$ gilt, dass

$$\sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right)\tag{75}$$

Mit $a_1 := 1$ und $a_2 := -1$ folgt dann aus $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, $i = 1, 2$ direkt, dass

$$\varepsilon_j \sim N\left(0, 2\sigma_\varepsilon^2\right)\tag{76}$$

Mit der Definition $\sigma^2 := 2\sigma_\varepsilon^2$ und dem Transformationstheorem für normalverteilte Zufallsvariablen bei linear-affiner Transformation folgt das Resultat dann direkt.

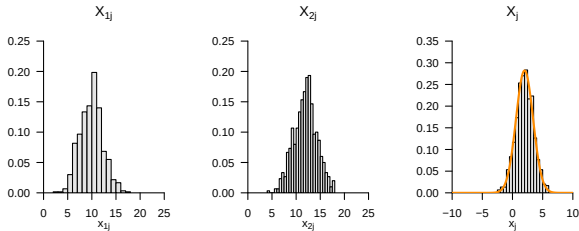
□

Zweistichproben-T-Test bei abhängigen Stichproben

(1) Statistisches Modell in generativer Form

Visualisierung des Zweistichproben-T-Test Theorems

$$n = 12, \mu_1 = 10, \mu_2 = 12, \sigma_\beta^2 = 5, \sigma_\varepsilon^2 = 1$$



(3) Praktisches Vorgehen

- Man nimmt an, dass ein vorliegender Datensatz von Messpaaren x_{11} und x_{21} , x_{12} und x_{22} , ..., x_{1n} und x_{2n} eine Realisation von X_{ij} , $i = 1, 2$, $j = 1, \dots, n$ wie in der Beschreibung des generativen Modells definiert ist.
 - Man bildet die Differenzwerte $x_1 = x_{21} - x_{11}$, $x_2 = x_{22} - x_{12}$, ..., $x_n = x_{2n} - x_{1n}$.
- ⇒ Man nutzt die Theorie des Einstichproben-T-Tests zum Testen der Hypothesen des Zweistichproben-T-Tests mit abhängigen Stichproben.

Zweistichproben-T-Test bei abhängigen Stichproben

(3) Praktisches Vorgehen

Datenbeispiel

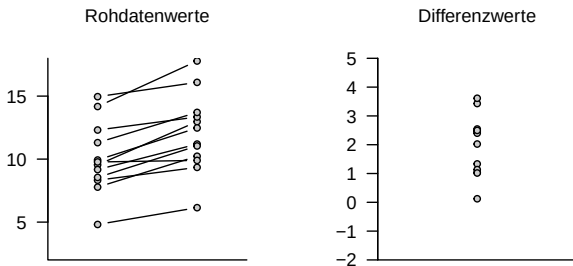
μ_1	μ_2	β_j	veps_{1_j}	veps_{2_j}	X_{1j}	X_{2j}	X_j
10	12	-0.600	0.144	1.574	9.54	12.97	3.430
10	12	3.962	0.211	1.824	14.17	17.79	3.613
10	12	-2.874	0.651	1.102	7.78	10.23	2.451
10	12	-5.039	-0.148	-0.815	4.81	6.15	1.333
10	12	0.907	1.402	0.437	12.31	13.35	1.036
10	12	-0.207	-0.624	-0.600	9.17	11.19	2.024
10	12	2.197	-0.890	-0.489	11.31	13.71	2.400
10	12	-0.482	0.409	0.955	9.93	12.47	2.546
10	12	3.993	0.963	0.098	14.96	16.09	1.135
10	12	-1.217	-0.458	-1.439	8.32	9.34	1.018
10	12	-0.194	-1.259	-0.762	8.55	11.04	2.497
10	12	-0.589	0.363	-1.514	9.77	9.90	0.123

$\Rightarrow$ Einstichproben-T-Test basierend auf $X_j, j = 1, \dots, n$

Zweistichproben-T-Test bei abhängigen Stichproben

(3) Praktisches Vorgehen

Datenbeispiel



⇒ Einstichproben-T-Test basierend auf $X_j, j = 1, \dots, n$

T-Tests

- T-Teststatistik
- Einstichproben-T-Tests
 - $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$
 - $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$
- Zweistichproben-T-Tests
 - Unabhängige Stichproben
 - Abhängige Stichproben
- **Selbstkontrollfragen**

Selbstkontrollfragen

1. Definieren Sie die T-Teststatistik.
2. Geben Sie das Theorem zur nichtzentralen T-Transformation wieder.
3. Geben Sie das Theorem zur Verteilung der T-Teststatistik wieder.
4. Erläutern Sie das Anwendungsszenario eines Einstichproben-T-Tests.
5. Erläutern Sie mögliche Hypothesenszenarien eines Einstichproben-T-Tests.
6. Geben Sie das statistische Modell eines Einstichproben-T-Tests in klassischer Form an.
7. Geben Sie das statistische Modell eines Einstichproben-T-Tests in generativer Form an.
8. Erläutern Sie die Bedeutung des eines statistischen Modells in generativer Form.
9. Definieren Sie den zweiseitigen Einstichproben-T-Test.
10. Skizzieren Sie qualitativ die Testgütefunktion eines zweiseitigen Einstichproben-T-Test
11. Geben Sie die Definition des kritischen Werts für einen zweiseitigen Level- α_0 -Einstichproben-T-Test wieder.
12. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level- α_0 -Einstichproben-T-Tests.
13. Geben Sie die Definition des p-Wertes Werts für einen zweiseitigen Einstichproben-T-Test wieder.
14. Von welchen Werten hängt die Powerfunktion eines Einstichproben-zweiseitigen T-Tests ab?
15. Skizzieren Sie qualitativ die Powerfunktion des zweiseitigen Einstichproben-T-Tests bei fester Stichprobengröße.

Selbstkontrollfragen

16. Skizzieren Sie qualitativ die Powerfunktion des zweiseitigen Einstichproben-T-Tests bei festem Nichtzentralitätsparameter.
17. Erläutern Sie das Anwendungsszenario eines Zweistichproben-T-Tests bei unabhängigen Stichproben.
18. Nennen Sie drei Modellszenarien für Zweistichproben-T-Tests bei unabhängigen Stichproben.
19. Geben Sie das statistische Modell eines Zweistichproben-T-Tests bei unabhängigen Stichproben in klassischer Form an.
20. Geben Sie das statistische Modell eines Zweistichproben-T-Tests bei unabhängigen Stichproben in generativer Form an.
21. Definieren Sie die gepoolte Stichprobenstandardabweichung.
22. Definieren Sie die Zweistichproben-T-Teststatistik.
23. Definieren Sie den kritischen Wert für einen zweiseitigen Level- α_0 -Zweistichproben-T-Tests bei unabhängigen Stichproben.
24. Erläutern Sie das praktische Vorgehen bei Durchführung eines zweiseitigen Level- α_0 -Zweistichproben-T-Tests bei unabhängigen Stichproben.
25. Erläutern Sie das Szenario eines Zweistichproben-T-Tests bei unabhängigen Stichproben mit bekanntem Varianzverhältnis.
26. Definieren Sie das Behrens-Fisher Problem.
27. Erläutern Sie das Anwendungsszenario eines Zweistichproben-T-Tests bei abhängigen Stichproben.
28. Geben Sie das statistische Modell eines Zweistichproben-T-Tests bei abhängigen Stichproben in generativer Form an.
29. Geben Sie das Theorem zur Verteilung der Differenzvariablen im Zweistichproben-T-Tests bei abhängigen Stichproben wieder.
30. Erläutern Sie das praktische Vorgehen bei Durchführung eines Zweistichproben-T-Tests bei abhängigen Stichproben.

References

DeGroot, Morris H., and Mark J. Schervish. 2012. *Probability and Statistics*. 4th ed. Boston: Addison-Wesley.

Lehmann, E. L. 1986. *Testing Statistical Hypotheses*. Wiley Series in Probability and Statistics.

Welch, B. L. 1938. "The Significance of the Difference Between Two Means When the Population Variances Are Unequal." *Biometrika* 29 (3-4): 350–62. <https://doi.org/10.1093/biomet/29.3-4.350>.

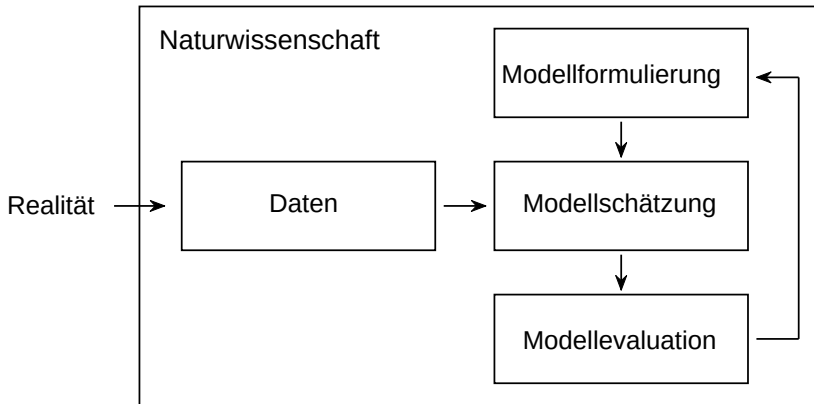


Inferenzstatistik

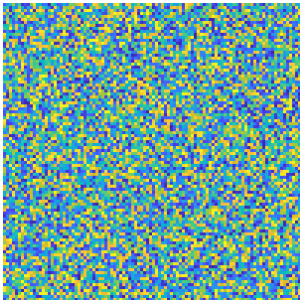
BSc Psychologie SoSe 2021

Prof. Dr. Dirk Ostwald

(12) Varianzanalysen

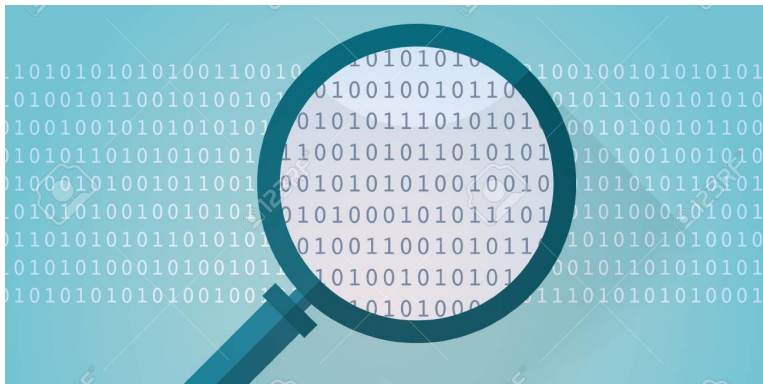


Rohdaten



Reduzierte Daten





Einfaktorielle Varianzanalyse

- Anwendungsszenario und Datendeskription
- Modellformulierung
- Modellschätzung
- Modellevaluation

Zweifaktorielle Varianzanalyse

- Anwendungsszenario und Datendeskription
- Modellformulierung
- Modellschätzung
- Modellevaluation

Selbstkontrollfragen

Die Darstellung basiert im Wesentlichen auf Kapitel 11 in DeGroot and Schervish (2012) und Kapitel 14 in Fahrmeir et al. (2016).

Einfaktorielle Varianzanalyse

- **Anwendungsszenario und Datendeskription**
- Modellformulierung
- Modellschätzung
- Modellevaluation

Zweifaktorielle Varianzanalyse

- Anwendungsszenario und Datendeskription
- Modellformulierung
- Modellschätzung
- Modellevaluation

Selbstkontrollfragen

Einfaktorielle Varianzanalyse = One-way Analysis of Variance (ANOVA)

Anwendungsszenario

- **Zwei oder mehr Stichproben** (im Folgenden oft **Gruppen** genannt) experimenteller Einheiten.
- Annahme der unabhängigen und identischen Normalverteilung $N(\mu_i, \sigma^2)$ der Daten.
- μ_i und σ^2 unbekannt.
- Absicht des inferentiellen Testens der Nullhypothese identischer Gruppenerwartungswerte.

Generalisierung des Zweistichproben-T-Tests bei unabhängigen Stichproben mit einfacher Nullhypothese für mehr als Zweistichproben

Anwendungsbeispiele

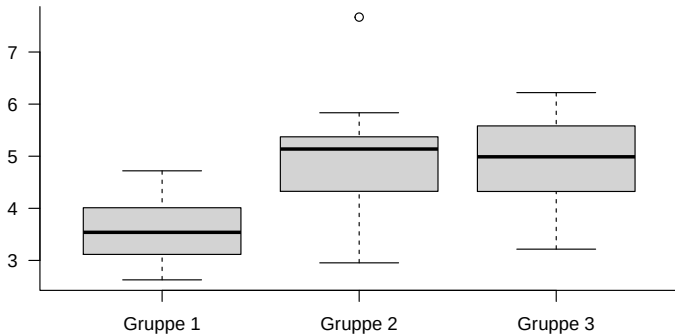
- BDI Score Analyse von drei Gruppen psychiatrischer Patient:innen nach PA, CBT, ST
 - Inferentielle Evidenz für Gruppenerwartungswertunterschiede?
 - Evidenz für unterschiedliche Therapiewirksamkeit?
- Tense Arousal Differenzwertanalyse der vier Filmclipbedingungen im Affect Datensatz
 - Inferentielle Evidenz für Gruppenerwartungswertunterschiede?
 - Evidenz für unterschiedliche Tense Arousal Induktion?

x_{ij} : Datenpunkt der j ten experimentellen in Einheit in der i ten Gruppe

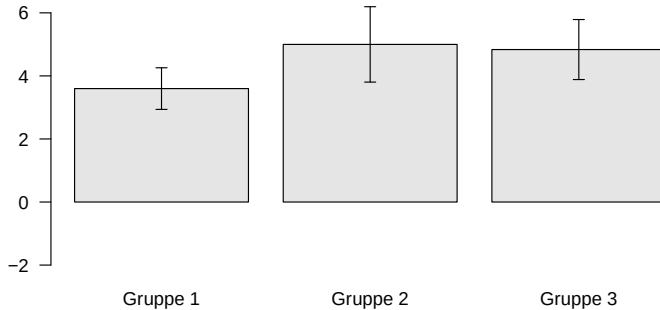
	x_{1j}	x_{2j}	x_{3j}
1	3.55	5.12	4.35
2	2.79	4.88	5.60
3	4.04	3.66	3.22
4	4.64	5.17	5.33
5	3.21	5.16	5.56
6	3.62	5.03	6.22
7	3.52	2.95	4.79
8	4.72	5.21	5.70
9	3.98	7.67	4.29
10	2.63	3.77	4.54
11	3.02	5.83	3.23
12	3.45	5.53	5.19

$i = 1, 2, 3$ mit $j = 1, \dots, 12$

Boxplot



Gruppenmittelwerte ($\pm$ Gruppenstandardabweichungen)



Einfaktorielle Varianzanalyse

- Anwendungsszenario und Datendeskription
- **Modellformulierung**
- Modellschätzung
- Modellevaluation

Zweifaktorielle Varianzanalyse

- Anwendungsszenario und Datendeskription
- Modellformulierung
- Modellschätzung
- Modellevaluation

Selbstkontrollfragen

Definition (Modell der einfaktoriellen Varianzanalyse)

Für $i = 1, \dots, p$ sei

$$X_{i1}, \dots, X_{im} \sim N(\mu_i, \sigma^2) \quad (1)$$

eine Stichprobe eines Normalverteilungsmodells von Größe m mit unbekanntem Erwartungswertparameter μ_i und unbekanntem Varianzparameter $\sigma^2 > 0$. Dann heißt (1) *Klassisches Modell der einfaktoriellen Varianzanalyse in Erwartungswertparameterdarstellung*. Die Gesamtstichprobengröße bezeichnen wir mit $n := pm$. Äquivalent sei

$$X_{ij} = \mu_i + \varepsilon_{ij} \text{ mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ für } i = 1, \dots, p \text{ und } j = 1, \dots, m \quad (2)$$

wobei

- i die Stichproben und j die experimentellen Einheiten indizieren,
- m die Stichprobengrößen und $n = pm$ die Gesamtstichprobengrößen sind,
- X_{ij} beobachtbare Zufallsvariablen sind,
- μ_i feste Erwartungswertparameter der Stichprobenvariablen sind, und
- ε_{ij} unabhängige normalverteilte nicht-beobachtbare Zufallsvariablen sind.

Dann heißt (2) *Generatives Modell der einfaktoriellen Varianzanalyse in Erwartungswertparameterdarstellung*.

Bemerkungen

- Die Varianzparameter aller Stichproben werden als identisch vorausgesetzt.
- Wir verzichten auf einen Beweis der Äquivalenz von (1) und (2).
- Es handelt sich um die Generalisierung des Zweistichproben-T-Test Modells für unabhängige Stichproben unter Annahme identischer Varianzparameter von $p = 2$ auf ein beliebiges $p \in \mathbb{N}$.
- Die Gesamtheit aller Stichprobenvariablen bezeichnen wir mit $X := \{X_{ij}\}_{i=1, \dots, p, j=1, \dots, m}$.

Theorem (Effektdarstellung des einfaktoriellen Varianzanalysemodells)

Es sei das generative Modell der einfaktoriellen Varianzanalyse in Erwartungswertparameterdarstellung wie oben definiert gegeben. Dann können die beobachtbaren Zufallsvariablen äquivalent in der Form

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij} \text{ mit } \varepsilon_{ij} \sim N(0, \sigma^2) \text{ für } i = 1, \dots, p, j = 1, \dots, m, n = pm \quad (3)$$

mit

$$\alpha_i := \mu_i - \mu, \quad \mu := \frac{1}{p} \sum_{i=1}^p \mu_i \text{ und } \sum_{i=1}^p \alpha_i = 0 \quad (4)$$

geschrieben werden. (3) und (4) heißen *Generatives Modell der einfaktoriellen Varianzanalyse in Effektdarstellung*.

Beweis

Für ein beliebig gewähltes μ gilt

$$\mu_i = \mu + (\mu_i - \mu) =: \mu + \alpha_i \quad (5)$$

Aus der Definition von μ in obigem Theorem und mit $\mu_i = \alpha_i + \mu$ folgt dann außerdem

$$\mu = \frac{1}{p} \sum_{i=1}^p \mu_i = \frac{1}{p} \sum_{i=1}^p (\mu + \alpha_i) = \mu + \frac{1}{p} \sum_{i=1}^p \alpha_i \Leftrightarrow \frac{1}{p} \sum_{i=1}^p \alpha_i = 0. \quad (6)$$

□

Bemerkungen

- α_i heißt *i*ter Gruppeneffektparameter oder Effektparameter des *i*ten Faktorlevels.
- Die Effektdarstellung betont die Analogie zum zweifaktoriellen Varianzanalysemodell.

Einfaktorielle Varianzanalyse

- Anwendungsszenario und Datendeskription
- Modellformulierung
- **Modellschätzung**
- Modellevaluation

Zweifaktorielle Varianzanalyse

- Anwendungsszenario und Datendeskription
- Modellformulierung
- Modellschätzung
- Modellevaluation

Selbstkontrollfragen

Theorem (Parameterschätzer des einfaktoriellen Varianzanalysemodells)

$\mu_1, \dots, \mu_p$ und $\sigma^2 > 0$ seien die Parameter des einfaktoriellen Varianzanalysemodells wie oben definiert. Dann sind die Maximum Likelihood Schätzer von $\mu_1, \dots, \mu_p$ und $\sigma^2 > 0$ gegeben durch

$$\hat{\mu}_i := \frac{1}{n} \sum_{j=1}^n X_{ij} \text{ und } \hat{\sigma}_{\text{ML}}^2 := \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2, \quad (7)$$

respektive, wobei die Erwartungswertparameterschätzer unverzerrte Schätzer sind und der Varianzschätzer ein verzerrter Schätzer ist. Ein unverzerrter Schätzer von $\sigma^2 > 0$ ist gegeben durch

$$\hat{\sigma}^2 := \frac{1}{n-p} \sum_{i=1}^p \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2. \quad (8)$$

Bemerkungen

- Wir verzichten auf einen Beweis.
- Die Stichprobenmittelwerte sind unverzerrte Schätzer der wahren, aber unbekannten, Stichprobenerwartungswertparameter, die Stichprobenvarianz der Gesamtstichprobe ist ein unverzerrter Schätzer des Varianzparameters.

Einfaktorielle Varianzanalyse

- Anwendungsszenario und Datendeskription
- Modellformulierung
- Modellschätzung
- **Modellevaluation**

Zweifaktorielle Varianzanalyse

- Anwendungsszenario und Datendeskription
- Modellformulierung
- Modellschätzung
- Modellevaluation

Selbstkontrollfragen

Hypothesen und Überblick

- Primäres Ziel der einfaktoriellen Varianzanalyse ist das Testen der Nullhypothese

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_p \quad (9)$$

bzw.

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_p = 0. \quad (10)$$

- Die Nullhypothese besagt also, dass zwischen den Gruppen keine Erwartungswertparameterunterschiede vorliegen bzw. dass die Gruppeneffektparameter identisch Null sind.
- Die Alternativhypothese lautet somit

$$H_1 : \mu_i \neq \mu_j \text{ für mindestens ein Paar } (i, j) \text{ mit } i \neq j \quad (11)$$

bzw.

$$H_1 : \alpha_i \neq 0 \text{ für mindestens ein } i \in \mathbb{N}_p. \quad (12)$$

- Die Alternativhypothese besagt also, dass sich mindestens zwei Erwartungswertparameter unterscheiden bzw. mindestens ein Gruppeneffektparameter ungleich Null ist.
- Ein kritischer Wert-basierter Test dieser Hypothesen kann mithilfe der *F-Teststatistik* und ihrer Verteilung konstruiert werden. Die F-Teststatistik wiederum beruht auf einer *Quadratsummenzerlegung*. Wir führen beide Konzepte im Folgenden ein.

Theorem (Quadratsummenzerlegung bei einfaktorieller Varianzanalyse)

Für $i = 1, \dots, p$ und $j = 1, \dots, m$ bezeichne X_{ij} die j te Stichprobenvariable der i ten Stichproben-
gruppe eines einfaktoriellen Varianzanalysemodells. Weiterhin seien

$$\bar{X} := \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^m X_{ij} \text{ und } \bar{X}_i := \frac{1}{m} \sum_{j=1}^m X_{ij} \quad (13)$$

den *Gesamtmittelwert (grand mean)* und das i te Stichprobenmittel, respektive. Schließlich seien

$$\text{SQT} := \sum_{i=1}^p \sum_{j=1}^m (X_{ij} - \bar{X})^2 \quad \text{die Total Sum of Squares}$$

$$\text{SQB} := \sum_{i=1}^p m(\bar{X}_i - \bar{X})^2 \quad \text{die Between-Group Sum of Squares}$$

$$\text{SQW} := \sum_{i=1}^p \sum_{j=1}^m (X_{ij} - \bar{X}_i)^2 \quad \text{die Within-Group Sum of Squares}$$

Dann gilt

$$\text{SQT} = \text{SQW} + \text{SQB}. \quad (14)$$

Bemerkungen

- SQT repräsentiert die totale Variabilität der Daten um den Gesamtmittelwert.
- SQB repräsentiert die Variabilität der Gruppenmittelwerte um den Gesamtmittelwert.
- SQW repräsentiert die Variabilität der Daten um ihre jeweiligen Gruppenmittelwerte.
- Die totale Variabilität wird hier also in zwei unabhängige Beiträge von Variabilität zerlegt.
- SQW wird auch *Residualvariabilität* genannt, weil sie die verbleibende Variabilität nach Schätzung der Gruppenerwartungswertparameter quantifiziert und gilt, dass

$$\hat{\sigma}^2 = \text{SQW}. \quad (15)$$

- Die Quadratsummenzerlegung ergibt sich aus algebraischen Identitäten wie unten gezeigt.

Beweis

$$\begin{aligned}\text{SQT} &= \sum_{i=1}^p \sum_{j=1}^m (X_{ij} - \bar{X})^2 \\&= \sum_{i=1}^p \sum_{j=1}^m (X_{ij} - \bar{X}_i + \bar{X}_i - \bar{X})^2 \\&= \sum_{i=1}^p \sum_{j=1}^m \left((X_{ij} - \bar{X}_i) + (\bar{X}_i - \bar{X}) \right)^2 \\&= \sum_{i=1}^p \sum_{j=1}^m \left((X_{ij} - \bar{X}_i)^2 + 2(X_{ij} - \bar{X}_i)(\bar{X}_i - \bar{X}) + (\bar{X}_i - \bar{X})^2 \right) \\&= \sum_{i=1}^p \left(\sum_{j=1}^m (X_{ij} - \bar{X}_i)^2 + \sum_{j=1}^m 2(X_{ij} - \bar{X}_i)(\bar{X}_i - \bar{X}) + \sum_{j=1}^m (\bar{X}_i - \bar{X})^2 \right) \\&= \sum_{i=1}^p \left(\sum_{j=1}^m (X_{ij} - \bar{X}_i)^2 + 2(\bar{X}_i - \bar{X}) \sum_{j=1}^m (X_{ij} - \bar{X}_i) + m(\bar{X}_i - \bar{X})^2 \right)\end{aligned}$$

Beweis (fortgeführt)

$$\begin{aligned}
 &= \sum_{i=1}^p \left(\sum_{j=1}^m (X_{ij} - \bar{X}_i)^2 + 2(\bar{X}_i - \bar{X}) \sum_{j=1}^m \left(X_{ij} - \frac{1}{m} \sum_{j=1}^m X_{ij} \right) + m(\bar{X}_i - \bar{X})^2 \right) \\
 &= \sum_{i=1}^p \left(\sum_{j=1}^m (X_{ij} - \bar{X}_i)^2 + 2(\bar{X}_i - \bar{X}) \left(\sum_{j=1}^m X_{ij} - \sum_{j=1}^m \left(\frac{1}{m} \sum_{j=1}^m X_{ij} \right) \right) + m(\bar{X}_i - \bar{X})^2 \right) \\
 &= \sum_{i=1}^p \left(\sum_{j=1}^m (X_{ij} - \bar{X}_i)^2 + 2(\bar{X}_i - \bar{X}) \left(\sum_{j=1}^m X_{ij} - \frac{m}{m} \sum_{j=1}^m X_{ij} \right) + m(\bar{X}_i - \bar{X})^2 \right) \\
 &= \sum_{i=1}^p \left(\sum_{j=1}^m (X_{ij} - \bar{X}_i)^2 + 2(\bar{X}_i - \bar{X}) \left(\sum_{j=1}^m X_{ij} - \sum_{j=1}^m X_{ij} \right) + m(\bar{X}_i - \bar{X})^2 \right) \\
 &= \sum_{i=1}^p \left(\sum_{j=1}^m (X_{ij} - \bar{X}_i)^2 + m(\bar{X}_i - \bar{X})^2 \right) \\
 &= \sum_{i=1}^p \sum_{j=1}^m (X_{ij} - \bar{X}_i)^2 + \sum_{i=1}^p m(\bar{X}_i - \bar{X})^2 \\
 &= \text{SQW} + \text{SQB}.
 \end{aligned}$$

Definition (F-Teststatistik)

Es seien das Modell der einfaktoriellen Varianzanalyse sowie die Between-Group Sum of Squares SQB und die Within-Group Sum of Squares SQW definiert wie oben. Dann ist die *F-Teststatistik* definiert als

$$F := \frac{SQB/(p-1)}{SQW/(n-p)} \quad (16)$$

Der Zähler und der Nenner dieser Definition werden auch als *Mean SQB* und *Mean SQW*, respektive, bezeichnet.

Bemerkungen

- Die F-Teststatistik vergleicht die Variabilität zwischen Gruppen mit der Variabilität innerhalb von Gruppen.
- Große Variabilität der Gruppenmittelwerte um den Gesamtmittelwert bei gleichzeitiger kleiner Residualvariabilität resultiert in hohen F -Werten $\Rightarrow$ Bei Nichtzutreffen der Nullhypothese erwartet man also hohe F -Werte bei Zutreffen der Nullhypothese eher geringe F -Werte
- Die F-Teststatistik misst die Größe der SQB pro zusätzlicher Gruppe in Einheiten der geschätzten Varianz $\hat{\sigma}^2$.
- Ein F -Werte ≤ 1 besagt, dass die SQB pro zusätzlicher Gruppe kleiner bzw. gleich der Residualvarianz ist.
- Ohne Beweis halten wir fest, dass für $p = 2$ gilt, dass $F = T^2$, wobei T die T-Teststatistik bezeichnet.

Theorem (Verteilung der F-Teststatistik bei H_0)

Es seien das Modell der einfaktoriellen Varianzanalyse sowie die F-Teststatistik F wie oben definiert und es gelte zudem

$$H_0 : \mu_1 = \dots = \mu_p \text{ bzw. } H_0 : \alpha_1 = \dots = \alpha_p = 0. \quad (17)$$

Dann ist die F-Teststatistik nach einer f -Verteilung mit Freiheitsgradparametern $p - 1$ und $n - p$ verteilt,

$$F \sim f(p - 1, n - p). \quad (18)$$

Bemerkung

- Wir verzichten auf einen Beweis.
- Trifft H_0 nicht zu, so ist die F-Teststatistik analog zur T-Teststatistik nach einer *nichtzentralen f-Verteilung* verteilt. Wir gehen auf diesen Fall nicht genauer ein, d.h. wir verzichten hier auf eine Analyse der Testgütefunktion.

Zur Erinnerung an (6) Transformationen der Normalverteilung

Definition (f -Zufallsvariable)

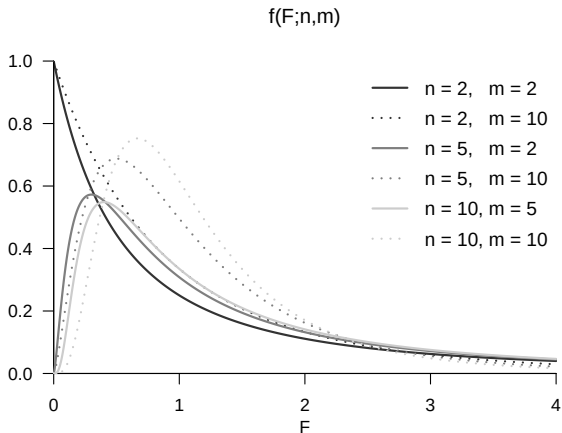
F sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}_{>0}$ und WDF

$$p_F : \mathbb{R} \rightarrow \mathbb{R}_{>0}, f \mapsto p_F(f) := m^{\frac{m}{2}} n^{\frac{n}{2}} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} \frac{f^{\frac{m}{2}-1}}{\left(1 + \frac{m}{n} f\right)^{\frac{m+n}{2}}}, \quad (19)$$

wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass F einer f -Verteilung mit Freiheitsgradparametern n und m unterliegt und nennen F eine f -Zufallsvariable mit Freiheitsgradparametern n, m . Wir kürzen dies mit $F \sim f(n, m)$ ab. Die WDF einer f -Zufallsvariable mit Freiheitsgradparametern n, m bezeichnen wir mit $f(F; n, m)$, die KVF einer f -Zufallsvariable mit Freiheitsgradparametern n, m bezeichnen wir mit $\varphi(F; n, m)$, und die inverse KVF einer f -Zufallsvariable mit Freiheitsgradparametern n, m bezeichnen wir mit $\varphi^{-1}(F; n, m)$.

Zur Erinnerung an (6) Transformationen der Normalverteilung

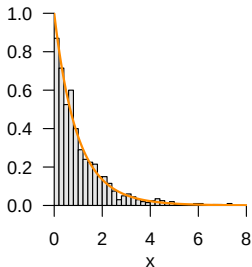
Wahrscheinlichkeitsdichtefunktionen der f -Verteilung



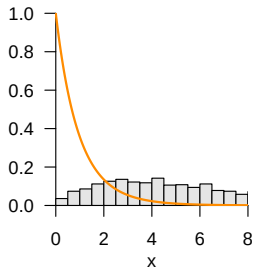
Simulation der Verteilung der F-Teststatistik

$H_0 : \mu_i := 3 \text{ für } i = 1, 2, 3, 4 \text{ und } H_1 := \mu_1 = 4, \mu_i = 3 \text{ für } i = 2, 3, 4, \text{ und } \sigma^2 := 1$

$\mu_1 = \mu_2 = \mu_3$



$\mu_1 \neq \mu_2 = \mu_3$



Theorem (F-Test und F-Testumfangkontrolle)

Es sei das Modell der einfaktoriellen Varianzanalyse gegeben und die F-Teststatistik wie oben definiert. Weiterhin sei der kritische Wert-basierte Test

$$\phi(X) : 1_{\{F \geq k\}} := \begin{cases} 1 & F \geq k \\ 0 & F < k \end{cases} \quad (20)$$

definiert. Dann ist ϕ genau dann ein Level- α_0 -Test mit Testumfang α , wenn

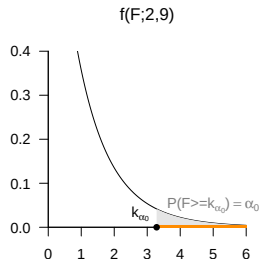
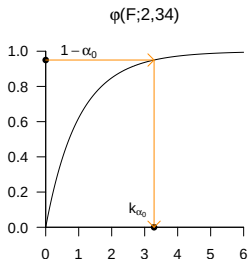
$$k = k_{\alpha_0} := \varphi^{-1}(1 - \alpha_0; p - 1, n - m) \quad (21)$$

gilt.

Bemerkungen

- Ein Beweis kann in Analogie zur Testumfangkontrolle bei T-Tests formuliert werden.

Wahl von $k_{\alpha_0} := \varphi^{-1}(1 - \alpha_0; p - 1, n - p)$ mit $p = 3, n = 12$ $\alpha_0 := 0.05$ und Ablehnungsbereich



Praktisches Vorgehen

- Man nimmt an, dass ein vorliegender Datensatz von p Gruppendatensätzen $x_{11}, \dots, x_{1n_1}, x_{21}, \dots, x_{2n_2}, \dots, x_{p1}, \dots, x_{pn_p}$ Realisationen von $X_{ij} \sim N(\mu_i, \sigma^2)$ mit unbekannten Parametern $\mu_i, i = 1, \dots, p$ und $\sigma^2 > 0$ sind.
- Man möchte entscheiden ob $H_0 : \mu_1 = \dots = \mu_p$ eher zutrifft oder eher nicht.
- Man wählt ein Signifikanzniveau α_0 und bestimmt den zugehörigen Freiheitsgradparameter-abhängigen kritischen Wert k_{α_0} . Zum Beispiel gilt bei Wahl von $\alpha_0 := 0.05, p = 3, m = 12, i = 1, 2, 3$ und somit $n = 36$, dass $k_{\alpha_0} = \varphi^{-1}(1 - 0.05; 2, 34) \approx 3.28$ ist.
- Anhand der SQB und SQW sowie von n und p berechnet man den realisierten Wert der F-Teststatistik, den wir hier mit $\tilde{f}$ bezeichnen.
- Wenn $\tilde{f}$ größer gleich k_{α_0} ist, lehnt man die Nullhypothese ab, andernfalls nicht.
- Die oben entwickelte Theorie garantiert dann, dass man im Mittel in höchstens $\alpha_0 \cdot 100$ von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.
- Schließlich ergibt sich der assoziierte p-Wert der realisierter F-Teststatistik $\tilde{f}$ in Analogie zum Einstichproben-T-Test mit zusammengesetzter Nullhypothese zu

$$\text{p-Wert} = \mathbb{P}(F \geq \tilde{f}) = 1 - \varphi(\tilde{f}; p - 1, n - p) \quad (22)$$

Anwendungsbeispiel

x_{ij} : Datenpunkt der j ten experimentellen in Einheit in der i ten Gruppe

	x_{1j}	x_{2j}	x_{3j}
1	3.55	5.12	4.35
2	2.79	4.88	5.60
3	4.04	3.66	3.22
4	4.64	5.17	5.33
5	3.21	5.16	5.56
6	3.62	5.03	6.22
7	3.52	2.95	4.79
8	4.72	5.21	5.70
9	3.98	7.67	4.29
10	2.63	3.77	4.54
11	3.02	5.83	3.23
12	3.45	5.53	5.19

Anwendungsbeispiel - Manuelle Berechnung

```
p      = ncol(X)           # Gruppenanzahl
m      = nrow(X)           # Gruppengröße bei gleichen Gruppengrößen
n      = p*m               # Gesamtstichprobengröße
x_bar  = mean(as.matrix(X)) # Gesamtmittelwert
x_bar_i = colMeans(as.matrix(X)) # Gruppenmittelwerte
alpha_0 = 0.05             # Signifikanzlevel
k_alpha_0 = qf(1 - alpha_0, p-1, n-p) # kritischer Wert
SQT     = sum((as.vector(as.matrix(X)) - x_bar)^2) # Total Sum of Squares
SQB     = sum(m*(x_bar_i - x_bar)^2) # Between-Group Sum of Squares
SQW     = sum((t(as.matrix(X)) - as.vector(x_bar_i))^2) # Within-Group Sum of Squares
MSB     = SQB/(p-1)         # Between-Group Mean Squares
MSW     = SQW/(n-p)         # Within-Group Mean Squares
f_tilde = MSB/MSW           # Realisierung der F-Teststatistik
if(f_tilde >= k_alpha_0){   # \tilde{f} >= k_alpha_0
  phi = 1                   # Testwert
} else {
  phi = 0                   # Testwert
}
p_value = 1 - pf(f_tilde, p-1, n-p) # p-Wert
```

```
Gruppenanzahl      = 3
Gruppengrößen      = 12
Gesamtstichprobengröße = 36
Gesamtmittelwert   = 4.48
Gruppenmittelwerte  = 3.6 5 4.84
alpha_0             = 0.05
k_alpha_0           = 3.28
SQT                 = 44.6
SQB                 = 14.1
SQW                 = 30.5
MSB                 = 7.05
MSW                 = 0.923
f_tilde             = 7.63
phi                 = 1
p-Wert              = 0.00189
```

Anwendungsbeispiel - Automatische Berechnung

```
XAOV = data.frame(measurements = c(X$X_1j, X$X_2j, X$X_3j),  
                  group        = as.factor(c(rep(1,m),  
                                              rep(2,m),  
                                              rep(3,m))))  
# Reformatierung des Datensatzes:  
# Datenpunktgruppenlabel als Faktorlevelspalte  
  
res.aov = aov(measurements ~ group, data = XAOV)  
# Benutzung von R's aov Funktion  
summary(res.aov)  
# Ausgabe
```

```
      Df Sum Sq Mean Sq F value Pr(>F)  
group    2  14.1    7.05    7.63 0.0019 **  
Residuals 33   30.5     0.92  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Einfaktorielle Varianzanalyse

- Anwendungsszenario und Datendeskription
- Modellformulierung
- Modellschätzung
- Modellevaluation

Zweifaktorielle Varianzanalyse

- **Anwendungsszenario und Datendeskription**
- Modellformulierung
- Modellschätzung
- Modellevaluation

Selbstkontrollfragen

Zweifaktorielle Varianzanalyse = Two-way Analysis of Variance (ANOVA)

Anwendungsszenario

- **Zweifaktorielle Versuchspläne mit crossed design (cf. Modul A.2)**
 - Eine univariate Messvariable bestimmt an individuellen experimentellen Einheiten.
 - Zwei diskrete experimentelle Variablen, die mindestens zweistufig sind.
 - Die experimentellen Variablen werden *Faktoren* genannt.
 - Die Stufen der Faktoren werden auch *Faktorlevel* genannt.
 - Jede Stufe eines Faktors wird mit allen Leveln des anderen Faktors kombiniert
- Zweifaktorielle Versuchspläne werden üblicherweise anhand ihrer Faktorlevel bezeichnet
 - 2×2 ANOVA : Faktor A mit Level 1,2 Faktor B mit Level 1,2
 - 2×3 ANOVA : Faktor A mit Level 1,2 Faktor B mit Level 1,2,3
 - 4×2 ANOVA : Faktor A mit Level 1,2,3,4 Faktor B mit Level 1,2
 - 3×1 ANOVA : Faktor A mit Level 1,2,3 Faktor B mit Level 1
- In der Psychologie sind 2×2 Designs sehr populär, wir fokussieren auf diesen Fall.
- Die Zellen eines 2×2 Designs werden im folgenden auch als Gruppen bezeichnet.

Anwendungsszenario 2 × 2 ANOVA

Konzeptuelles Design

		Faktor B	
		Level 1	Level 2
Faktor A	Level 1	A1B1	A1B2
	Level 2	A2B1	A2B2

Anwendungsszenario 2 × 2 ANOVA

Datennotation

		Faktor B	
		Level 1	Level 2
Faktor A	Level 1	$x_{111} = 1.2$ $x_{112} = 0.4$ $x_{113} = 1.7$ $\vdots$ $x_{11n_{11}} = 2.1$	$x_{121} = 4.1$ $x_{122} = 1.8$ $x_{123} = 3.3$ $\vdots$ $x_{12n_{12}} = 5.9$
	Level 2	$x_{211} = 0.1$ $x_{212} = 0.8$ $x_{213} = 2.7$ $\vdots$ $x_{21n_{21}} = 1.4$	$x_{221} = 7.4$ $x_{222} = 6.2$ $x_{223} = 9.5$ $\vdots$ $x_{22n_{22}} = 6.1$

x_{ijk} bezeichnet den Datenpunkt der k ten experimentellen Einheit ($k = 1, \dots, m$)
im i ten Level von Faktor A und j ten Level von Faktor B ($i = 1, 2, j = 1, 2$)

Anwendungsszenario 2 × 2 ANOVA

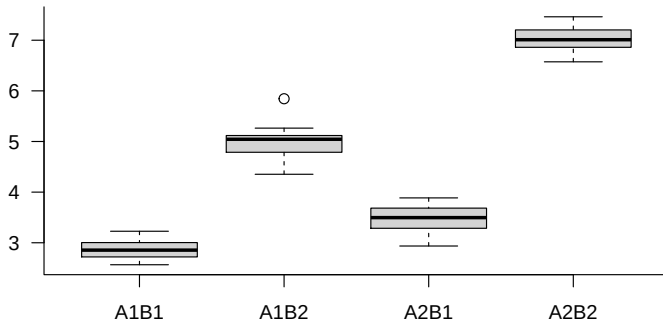
Datentabellenform mit $m := 12$

	x_{11k}	x_{12k}	x_{21k}	x_{22k}
1	2.86	5.04	3.29	6.88
2	2.62	4.96	3.69	7.33
3	3.01	4.58	2.94	6.76
4	3.20	5.05	3.61	6.57
5	2.75	5.05	3.68	6.84
6	2.88	5.01	3.89	7.45
7	2.85	4.35	3.43	7.06
8	3.23	5.07	3.72	6.99
9	2.99	5.84	3.28	6.93
10	2.57	4.61	3.35	7.46
11	2.69	5.26	2.94	7.07
12	2.83	5.17	3.56	7.03

x_{ijk} bezeichnet den Datenpunkt der k ten experimentellen Einheit ($k = 1, \dots, m$)
im i ten Level von Faktor A und j ten Level von Faktor B ($i = 1, 2, j = 1, 2$)

Anwendungsszenario 2 × 2 ANOVA

Gruppendatendeskription durch Boxplot



Anwendungsszenario 2 × 2 ANOVA

Man unterscheidet hinsichtlich der Gruppenmittelwerte **Haupteffekte** und **Interaktionen**

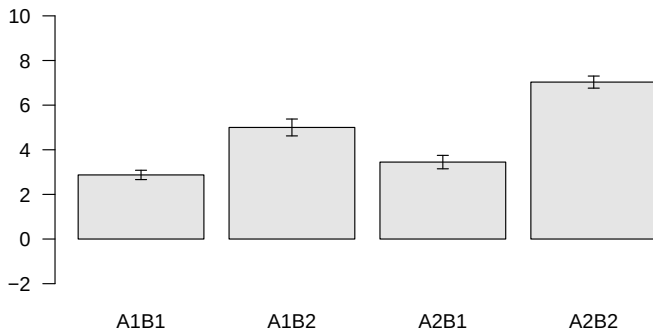
- Intuitiv spricht man vom Vorliegen eines *Haupteffekts von Faktor A*, wenn sich die Gruppenmittelwerte zwischen Level 1 und Level 2 von Faktor A, jeweils gemittelt über die zwei Level von Faktor B, unterscheiden.
- Intuitiv spricht man vom Vorliegen eines *Haupteffekts von Faktor B*, wenn sich die Gruppenmittelwerte zwischen Level 1 und Level 2 von Faktor B, jeweils gemittelt über die zwei Level von Faktor A, unterscheiden.
- Intuitiv spricht man vom Vorliegen einer *Interaktion der Faktoren A und B*, wenn der Unterschied der Gruppenmittelwerte von Faktor A zwischen Level 1 und 2 unterschiedlich für Level 1 und Level 2 von Faktor B ausgeprägt ist.

Intuitiv beziehen sich Haupteffekte also auf (marginale) Unterschiede (Differenzen), während sich Interaktionen auf Unterschiede von Unterschieden (Differenzen von Differenzen) beziehen.

Das Vorhandensein einer Interaktion besagt lediglich, dass sich die Unterschiede der Gruppenmittelwerte zwischen den Leveln eines experimentellen Faktors in Abhängigkeit von den Leveln des anderen experimentellen Faktors ändern, es macht aber keine Aussage darüber, warum dies so ist. Haupteffekte und Interaktionen sind lediglich Datenmuster, keine mechanistischen wissenschaftlichen Theorien.

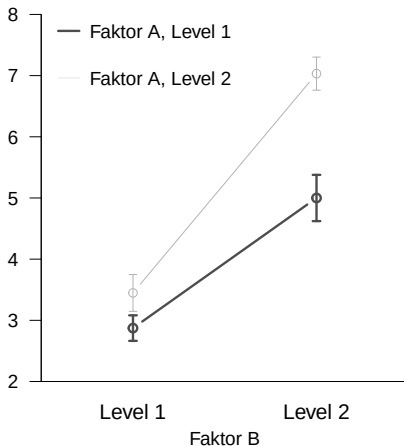
Anwendungsszenario 2 × 2 ANOVA

Darstellung der Gruppenmittelwerte und Gruppenstandardabweichungen durch Balkendiagramm



Anwendungsszenario 2 × 2 ANOVA

Darstellung der Gruppenmittelwerte und Gruppenstandardabweichungen durch Liniendiagramm



Anwendungsbeispiele

BDI Score Pre-Post-Differenzwertanalyse bei Klassischer und Online PT für zwei Altersklassen

- Faktor (A) Therapie mit Level (1) Klassische PT und Level (2) Online PT
- Faktor (B) Altersklasse mit Level (1) Junge Erwachsene und (2) Ältere Erwachsene
- 2×2 faktorieller Versuchsplan mit crossed design $\Rightarrow 2 \times 2$ ANOVA
 - Inferentielle Evidenz für einen Haupteffekt der Therapieart?
 - Inferentielle Evidenz für einen Haupteffekt der Altersklasse?
 - Inferentielle Evidenz für Interaktion (Verschiedene Therapieeffekte zwischen Altersklassen)?

TA Score Pre-Post-Differenzwertanalyse der vier Filmclips und zwei Studien im Affect Datensatz

- Faktor (A) Filmclips mit Level (1) Frontline, (2) Halloween, (3) Geographic, (4) Parenthood
- Faktor (B) Studien mit Level (1) Flat und (2) Maps
- 4×2 faktorieller Versuchsplan mit crossed design $\Rightarrow 4 \times 2$ ANOVA
 - Inferentielle Evidenz für einen Haupteffekt der Filmclipbedingung?
 - Inferentielle Evidenz für einen Haupteffekt der Studie?
 - Inferentielle Evidenz für eine Interaktion (Änderung der Filmclipeffekte zwischen den Studien)?

Einfaktorielle Varianzanalyse

- Anwendungsszenario und Datendeskription
- Modellformulierung
- Modellschätzung
- Modellevaluation

Zweifaktorielle Varianzanalyse

- Anwendungsszenario und Datendeskription
- **Modellformulierung**
- Modellschätzung
- Modellevaluation

Definition (Modell der additiven zweifaktoriellen Varianzanalyse)

Es sei

$$X_{ij1}, \dots, X_{ijm} \sim N(\mu_{ij}, \sigma^2) \text{ für } i = 1, \dots, p \text{ und } j = 1, \dots, q \quad (23)$$

eine Stichprobe eines Normalverteilungsmodells der Größe m mit unbekanntem Erwartungswertparameter μ_{ij} und unbekanntem Varianzparameter $\sigma^2 > 0$. Dann heißt (23) das *Klassische Modell der additiven zweifaktoriellen Varianzanalyse in Erwartungswertparameterdarstellung*. Gilt für die Erwartungswertparameter in der Erwartungswertparameterdarstellung der zweifaktoriellen Varianzanalyse, dass

$$\mu_{ij} = \mu + \alpha_i + \beta_j \text{ für } i = 1, \dots, p, j = 1, \dots, q \text{ mit } \sum_{i=1}^p \alpha_i = 0 \text{ und } \sum_{j=1}^q \beta_j = 0, \quad (24)$$

dann heißen (23) und (24) *Klassisches Modell der additiven zweifaktoriellen Varianzanalyse in Effektdarstellung*. Die Parameter $\alpha_1, \dots, \alpha_p$ heißen die *Haupteffekte von Faktor A*, die Parameter $\beta_1, \dots, \beta_q$ heißen die *Haupteffekte von Faktor B*, und wir bezeichnen die Gesamtstichprobengröße der zweifaktoriellen Varianzanalyse mit $n := pqm$. Schließlich heißt

$$X_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \text{ mit } \sum_{i=1}^p \alpha_i = 0, \sum_{j=1}^q \beta_j = 0 \text{ und } \varepsilon_{ij} \sim N(0, \sigma^2) \quad (25)$$

Generatives Modell der additiven zweifaktoriellen Varianzanalyse in Effektdarstellung.

Bemerkungen

- Der Einfachheit halber setzen wir hier die identische Stichprobengröße m voraus.
- Es seien

$$\bar{\mu} := \frac{1}{pq} \sum_{i=1}^p \sum_{j=1}^q \mu_{ij}, \quad \bar{\mu}_i := \frac{1}{q} \sum_{j=1}^q \mu_{ij}, \quad \text{und} \quad \bar{\mu}_j := \frac{1}{p} \sum_{i=1}^p \mu_{ij}. \quad (26)$$

Dann gelten folgende Beziehungen zwischen den Parametern der Erwartungswertparameterdarstellung und der Effektdarstellung für $i = 1, \dots, p$ und $j = 1, \dots, q$:

$$\mu = \bar{\mu}, \quad \alpha_i = \bar{\mu}_i - \mu, \quad \beta_j = \bar{\mu}_j - \mu \quad (27)$$

In der Effektdarstellung der additiven zweifaktoriellen Varianzanalyse gilt also

- Der Parameter μ encodiert den Mittelwert der Gruppenerwartungswerte μ_{ij}
- Die Parameter α_i encodieren die Differenz der über j gemittelten Gruppenerwartungswerte μ_{ij} vom Mittelwert der Gruppenerwartungswerte. Die Haupteffektparameter α_i erklären also den Unterschied zwischen den Mittelwerten in der jeweiligen Faktorstufe des Faktors zum Gesamtmittel ohne Berücksichtigung der Faktorstufe des anderen Faktors.
- Die Parameter β_j encodieren die Differenz der über i gemittelten Gruppenerwartungswerte μ_{ij} vom Mittelwert der Gruppenerwartungswerte. Die Haupteffektparameter β_j erklären also den Unterschied zwischen den Mittelwerten in der jeweiligen Faktorstufe des Faktors zum Gesamtmittel ohne Berücksichtigung der Faktorstufe des anderen Faktors.

Wir verzichten auf einen Beweis dieser Tatsache.

Bemerkungen (fortgeführt)

- Die Bedingungen $\sum_{i=1}^p \alpha_i = 0$ und $\sum_{j=1}^q \beta_j = 0$ stellen die Eindeutigkeit der Darstellung der pq Parameter μ_{ij} durch die $1 + p + q$ Parameter μ, α_i, β_j sicher. Dass diese ohne weitere Bedingungen an die α_i, β_j nicht gegeben ist, macht man sich an folgendem Beispiel klar

- Für $p = 2$ und $q = 2$ gelte

$$\mu_{11} = 1, \mu_{12} = 2, \mu_{21} = 3, \mu_{22} = 4. \quad (28)$$

Dann können die μ_{ij} sowohl durch

$$\mu = 0, \alpha_1 = 1, \alpha_2 = 3, \beta_1 = 0, \beta_2 = 1, \quad (29)$$

als auch durch

$$\mu = 1, \alpha_1 = 0, \alpha_2 = 2, \beta_1 = 0, \beta_2 = 1 \quad (30)$$

dargestellt werden

- Wir verzichten hier auf einen Beweis, dass die Bedingungen $\sum_{i=1}^p \alpha_i = 0$ und $\sum_{j=1}^q \beta_j = 0$ eine eindeutige Darstellung der μ_{ij} erlauben.

Parameterbeispiele bei 2×2 Varianzanalysen mit Interaktionseffekten

(1) Es sei $\mu := 1, \alpha_1 := 1, \beta_1 := 0$

- Aus der Bedingung $\alpha_1 + \alpha_2 = 0$ folgt $\alpha_2 = -1$; aus der Bedingung $\beta_1 + \beta_2 = 0$ folgt $\beta_2 = 0$.
- Es ergibt sich also die Erwartungswertparameterdarstellung

$$\begin{aligned}\mu_{11} &= \mu + \alpha_1 + \beta_1 = 1 + 1 + 0 = \mathbf{2} & \mu_{12} &= \mu + \alpha_1 + \beta_2 = 1 + 1 - 0 = \mathbf{2} \\ \mu_{21} &= \mu + \alpha_2 + \beta_1 = 1 - 1 + 0 = \mathbf{0} & \mu_{22} &= \mu + \alpha_2 + \beta_2 = 1 - 1 + 0 = \mathbf{0}\end{aligned}$$

(2) Es sei $\mu := 1, \alpha_1 := 0, \beta_1 := 1$

- Aus der Bedingung $\alpha_1 + \alpha_2 = 0$ folgt $\alpha_2 = 0$; aus der Bedingung $\beta_1 + \beta_2 = 0$ folgt $\beta_2 = -1$.
- Es ergibt sich also die Erwartungswertparameterdarstellung

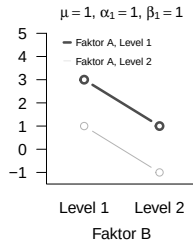
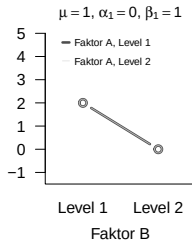
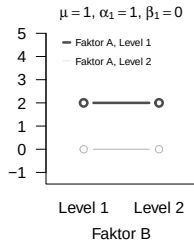
$$\begin{aligned}\mu_{11} &= \mu + \alpha_1 + \beta_1 = 1 + 0 + 1 = \mathbf{2} & \mu_{12} &= \mu + \alpha_1 + \beta_2 = 1 + 0 - 1 = \mathbf{0} \\ \mu_{21} &= \mu + \alpha_2 + \beta_1 = 1 + 0 + 1 = \mathbf{2} & \mu_{22} &= \mu + \alpha_2 + \beta_2 = 1 + 0 - 1 = \mathbf{0}\end{aligned}$$

(3) Es sei $\mu := 1, \alpha_1 := 1, \beta_1 := 1$

- Aus der Bedingung $\alpha_1 + \alpha_2 = 0$ folgt $\alpha_2 = -1$; aus der Bedingung $\beta_1 + \beta_2 = 0$ folgt $\beta_2 = -1$.
- Es ergibt sich also die Erwartungswertparameterdarstellung

$$\begin{aligned}\mu_{11} &= \mu + \alpha_1 + \beta_1 = 1 + 1 + 1 = \mathbf{3} & \mu_{12} &= \mu + \alpha_1 + \beta_2 = 1 + 1 - 1 = \mathbf{1} \\ \mu_{21} &= \mu + \alpha_2 + \beta_1 = 1 - 1 + 1 = \mathbf{1} & \mu_{22} &= \mu + \alpha_2 + \beta_2 = 1 - 1 - 1 = \mathbf{-1}\end{aligned}$$

Parameterbeispiele bei additiven 2×2 Varianzanalysen



Definition (Modell der zweifaktoriellen Varianzanalyse)

Es sei

$$X_{ij1}, \dots, X_{ijm} \sim N(\mu_{ij}, \sigma^2) \text{ für } i = 1, \dots, p \text{ und } j = 1, \dots, q \quad (31)$$

eine Stichprobe eines Normalverteilungsmodells der Größe m mit unbekanntem Erwartungswertparameter μ_{ij} und unbekanntem Varianzparameter $\sigma^2 > 0$. Dann heißt (31) das *Klassische Modell der zweifaktoriellen Varianzanalyse in Erwartungswertparameterdarstellung*. Gilt für die Erwartungswertparameter in der Erwartungswertparameterdarstellung der zweifaktoriellen Varianzanalyse, dass

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} \text{ für } i = 1, \dots, p, j = 1, \dots, q, \quad (32)$$

mit

$$\sum_{i=1}^p \alpha_i = 0, \sum_{j=1}^q \beta_j = 0, \sum_{i=1}^p \gamma_{ij} = 0, \sum_{j=1}^q \gamma_{ij} = 0 \text{ und } n := pqm \quad (33)$$

dann heißen (31), (32) und (33) *Klassisches Modell der zweifaktoriellen Varianzanalyse in Effektdarstellung*. $\alpha_1, \dots, \alpha_p$ heißen weiterhin die *Haupteffekte von Faktor A*, $\beta_1, \dots, \beta_q$ heißen weiterhin die *Haupteffekte von Faktor B*, und $\gamma_{11}, \dots, \gamma_{pq}$ heißen die *Interaktionseffekte der Faktoren A und B*. Schließlich heißt

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \text{ mit } \varepsilon_{ijk} \sim N(0, \sigma^2) \quad (34)$$

mit den Parameterbedingungen (33) *Generatives Modell der zweifaktoriellen Varianzanalyse in Effektdarstellung*.

Zweifaktorielle Varianzanalyse | Modellformulierung

Bemerkungen

- Es seien

$$\bar{\mu} := \frac{1}{pq} \sum_{i=1}^p \sum_{j=1}^q \mu_{ij}, \quad \bar{\mu}_i := \frac{1}{q} \sum_{j=1}^q \mu_{ij}, \quad \text{und} \quad \bar{\mu}_j := \frac{1}{p} \sum_{i=1}^p \mu_{ij}. \quad (35)$$

Dann gelten folgende Beziehungen zwischen den Parametern der Erwartungswertparameterdarstellung und der Effektdarstellung für $i = 1, \dots, p$ und $j = 1, \dots, q$:

$$\mu = \bar{\mu}, \quad \alpha_i = \bar{\mu}_i - \mu, \quad \beta_j = \bar{\mu}_j - \mu, \quad \gamma_{ij} = \mu_{ij} - (\mu + \alpha_i + \beta_j) \quad (36)$$

- In der Effektdarstellung der zweifaktoriellen Varianzanalyse gilt also
 - Die Parameter μ und die Haupteffektparameter α_i und β_j enkodieren die gleichen Aspekte wie in der additiven zweifaktoriellen Varianzanalyse.
 - Die Interaktionsparameter γ_{ij} enkodieren den Unterschied zwischem dem Mittelwert in der jeweiligen Faktorlevelkombination zu dem Teil des Gesamtmittelwertes, der nicht durch Addition der Haupteffektparameter beschrieben werden kann.
- Die Bedingungen

$$\sum_{i=1}^p \alpha_i = 0, \quad \sum_{j=1}^q \beta_j = 0, \quad \sum_{i=1}^p \gamma_{ij} = 0 \quad \text{und} \quad \sum_{j=1}^q \gamma_{ij} = 0 \quad (37)$$

stellen die Eindeutigkeit der Darstellung der pq Parameter μ_{ij} durch die $1 + p + q + qp$ Parameter $\mu, \alpha_i, \beta_j, \gamma_{ij}$ sicher. Wir verzichten wiederum auf einen Beweis dieser Tatsache

Parameterbeispiele bei 2 x 2 Varianzanalysen

(1) Es sei $\mu := 0, \alpha_1 := 0, \beta_1 := 0, \gamma_{11} = 1$

- Aus der Bedingung $\alpha_1 + \alpha_2 = 0$ folgt $\alpha_2 = 0$
- Aus der Bedingung $\beta_1 + \beta_2 = 0$ folgt $\beta_2 = 0$.
- Aus $\sum_{i=1}^p \gamma_{ij} = 0$ für $j = 1, 2$, also $\gamma_{11} + \gamma_{21} = 0$ und $\gamma_{12} + \gamma_{22} = 0$, folgt $\gamma_{21} = -1$
- Aus $\sum_{j=1}^q \gamma_{ij} = 0$ für $i = 1, 2$, also $\gamma_{11} + \gamma_{12} = 0$ und $\gamma_{21} + \gamma_{22} = 0$, folgt $\gamma_{12} = -1$
- Mit $\gamma_{12} + \gamma_{22} = 0$ folgt dann auch $\gamma_{22} = 1$

Die Wahl von $\mu := 1, \alpha_1 := 1, \beta_1 := 0, \gamma_{11} := 1$ legt also $\alpha_2, \beta_2, \gamma_{21}, \gamma_{12}, \gamma_{22}$ fest.

Es ergibt sich also die Erwartungswertparameterdarstellung

$$\mu_{11} = \mu + \alpha_1 + \beta_1 + \gamma_{11} = 0 + 0 + 0 + 1 = 1 \qquad \mu_{12} = \mu + \alpha_1 + \beta_2 + \gamma_{12} = 0 + 0 + 0 - 1 = -1$$

$$\mu_{21} = \mu + \alpha_2 + \beta_1 + \gamma_{21} = 0 + 0 + 0 - 1 = -1 \qquad \mu_{22} = \mu + \alpha_2 + \beta_2 + \gamma_{22} = 0 + 0 + 0 + 1 = 1$$

Parameterbeispiele bei 2 x 2 Varianzanalysen

(2) Es sei $\mu := 0, \alpha_1 := 0, \beta_1 := 0, \gamma_{11} = -1$

- Aus der Bedingung $\alpha_1 + \alpha_2 = 0$ folgt $\alpha_2 = 0$
- Aus der Bedingung $\beta_1 + \beta_2 = 0$ folgt $\beta_2 = 0$.
- Aus $\sum_{i=1}^p \gamma_{ij} = 0$ für $j = 1, 2$, also $\gamma_{11} + \gamma_{21} = 0$ und $\gamma_{12} + \gamma_{22} = 0$, folgt $\gamma_{21} = 1$
- Aus $\sum_{j=1}^q \gamma_{ij} = 0$ für $i = 1, 2$, also $\gamma_{11} + \gamma_{12} = 0$ und $\gamma_{21} + \gamma_{22} = 0$, folgt $\gamma_{12} = 1$
- Mit $\gamma_{12} + \gamma_{22} = 0$ folgt dann auch $\gamma_{22} = -1$

Die Wahl von $\mu := 1, \alpha_1 := 1, \beta_1 := 0, \gamma_{11} := 1$ legt also $\alpha_2, \beta_2, \gamma_{21}, \gamma_{12}, \gamma_{22}$ fest.

Es ergibt sich also die Erwartungswertparameterdarstellung

$$\mu_{11} = \mu + \alpha_1 + \beta_1 + \gamma_{11} = 0 + 0 + 0 - 1 = -1 \qquad \mu_{12} = \mu + \alpha_1 + \beta_2 + \gamma_{12} = 0 + 0 + 0 + 1 = 1$$

$$\mu_{21} = \mu + \alpha_2 + \beta_1 + \gamma_{21} = 0 + 0 + 0 + 1 = 1 \qquad \mu_{22} = \mu + \alpha_2 + \beta_2 + \gamma_{22} = 0 + 0 + 0 - 1 = -1$$

Parameterbeispiele bei 2 x 2 Varianzanalysen

(3) Es sei $\mu := 0, \alpha_1 := 1, \beta_1 := 0, \gamma_{11} = 1$

- Aus der Bedingung $\alpha_1 + \alpha_2 = 0$ folgt $\alpha_2 = -1$
- Aus der Bedingung $\beta_1 + \beta_2 = 0$ folgt $\beta_2 = 0$.
- Aus $\sum_{i=1}^p \gamma_{ij} = 0$ für $j = 1, 2$, also $\gamma_{11} + \gamma_{21} = 0$ und $\gamma_{12} + \gamma_{22} = 0$, folgt $\gamma_{21} = -1$
- Aus $\sum_{j=1}^q \gamma_{ij} = 0$ für $i = 1, 2$, also $\gamma_{11} + \gamma_{12} = 0$ und $\gamma_{21} + \gamma_{22} = 0$, folgt $\gamma_{12} = -1$
- Mit $\gamma_{12} + \gamma_{22} = 0$ folgt dann auch $\gamma_{22} = 1$

Die Wahl von $\mu := 1, \alpha_1 := 1, \beta_1 := 0, \gamma_{11} := 1$ legt also $\alpha_2, \beta_2, \gamma_{21}, \gamma_{12}, \gamma_{22}$ fest.

Es ergibt sich also die Erwartungswertparameterdarstellung

$$\mu_{11} = \mu + \alpha_1 + \beta_1 + \gamma_{11} = 0 + 1 + 0 + 1 = \mathbf{2} \qquad \mu_{12} = \mu + \alpha_1 + \beta_2 + \gamma_{12} = 0 + 1 + 0 - 1 = \mathbf{0}$$

$$\mu_{21} = \mu + \alpha_2 + \beta_1 + \gamma_{21} = 0 - 1 + 0 - 1 = \mathbf{-2} \qquad \mu_{22} = \mu + \alpha_2 + \beta_2 + \gamma_{22} = 0 - 1 + 0 + 1 = \mathbf{0}$$

Parameterbeispiele bei 2 x 2 Varianzanalysen

(4) Es sei $\mu := 1, \alpha_1 := 1, \beta_1 := -1, \gamma_{11} = 1$

- Aus der Bedingung $\alpha_1 + \alpha_2 = 0$ folgt $\alpha_2 = -1$
- Aus der Bedingung $\beta_1 + \beta_2 = 0$ folgt $\beta_2 = 1$.
- Aus $\sum_{i=1}^p \gamma_{ij} = 0$ für $j = 1, 2$, also $\gamma_{11} + \gamma_{21} = 0$ und $\gamma_{12} + \gamma_{22} = 0$, folgt $\gamma_{21} = -1$
- Aus $\sum_{j=1}^q \gamma_{ij} = 0$ für $i = 1, 2$, also $\gamma_{11} + \gamma_{12} = 0$ und $\gamma_{21} + \gamma_{22} = 0$, folgt $\gamma_{12} = -1$
- Mit $\gamma_{12} + \gamma_{22} = 0$ folgt dann auch $\gamma_{22} = 1$

Die Wahl von $\mu := 1, \alpha_1 := 1, \beta_1 := 0, \gamma_{11} := 1$ legt also $\alpha_2, \beta_2, \gamma_{21}, \gamma_{12}, \gamma_{22}$ fest.

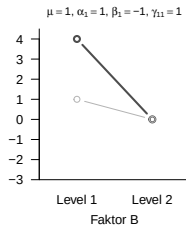
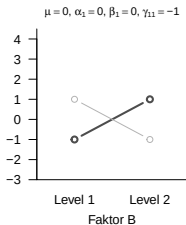
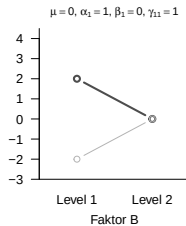
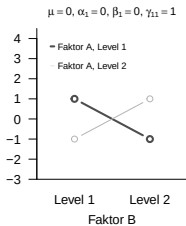
Es ergibt sich also die Erwartungswertparameterdarstellung

$$\mu_{11} = \mu + \alpha_1 + \beta_1 + \gamma_{11} = 1 + 1 + 1 + 1 = 4 \quad \mu_{12} = \mu + \alpha_1 + \beta_2 + \gamma_{12} = 1 + 1 - 1 - 1 = 0$$

$$\mu_{21} = \mu + \alpha_2 + \beta_1 + \gamma_{21} = 1 - 1 + 1 - 1 = 0 \quad \mu_{22} = \mu + \alpha_2 + \beta_2 + \gamma_{22} = 1 - 1 - 1 + 1 = 0$$

Zweifaktorielle Varianzanalyse | Modellformulierung

Parameterbeispiele bei 2 x 2 Varianzanalysen



Einfaktorielle Varianzanalyse

- Anwendungsszenario und Datendeskription
- Modellformulierung
- Modellschätzung
- Modellevaluation

Zweifaktorielle Varianzanalyse

- Anwendungsszenario und Datendeskription
- Modellformulierung
- **Modellschätzung**
- Modellevaluation

Selbstkontrollfragen

Theorem (Parameterschätzer der zweifaktoriellen Varianzanalyse)

Für $i = 1, \dots, p$ und $j = 1, \dots, q$ seien μ , α_i , β_j , γ_{ij} und $\sigma^2 > 0$ die Parameter des Modells der zweifaktoriellen Varianzanalyse in (klassischer oder generativer) Effektdarstellung wie oben definiert. Weiterhin seien

$$\bar{X} := \frac{1}{pqn} \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^m X_{ijk} \quad (38)$$

und für $i = 1, \dots, p$, $j = 1, \dots, q$

$$\bar{X}_i := \frac{1}{qm} \sum_{j=1}^q \sum_{k=1}^m X_{ijk}, \bar{X}_j := \frac{1}{pm} \sum_{i=1}^p \sum_{k=1}^m X_{ijk} \text{ und } \bar{X}_{ij} := \frac{1}{m} \sum_{k=1}^m X_{ijk} \quad (39)$$

definiert. Dann sind die Maximum Likelihood Schätzer von μ , α_i , β_j und γ_{ij} gegeben durch

$$\hat{\mu} = \bar{X}, \quad \hat{\alpha}_i = \bar{X}_i - \bar{X}, \quad \hat{\beta}_j = \bar{X}_j - \bar{X} \text{ und } \hat{\gamma}_{ij} = \bar{X}_{ij} - (\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j) \quad (40)$$

Ein unverzerrter Schätzer von $\sigma^2 > 0$ ist gegeben durch

$$\hat{\sigma}^2 = \frac{1}{pq(m-1)} \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^m (X_{ijk} - \bar{X}_{ij})^2. \quad (41)$$

- Wir verzichten auf einen Beweis.

Einfaktorielle Varianzanalyse

- Anwendungsszenario und Datendeskription
- Modellformulierung
- Modellschätzung
- Modellevaluation

Zweifaktorielle Varianzanalyse

- Anwendungsszenario und Datendeskription
- Modellformulierung
- Modellschätzung
- **Modellevaluation**

Selbstkontrollfragen

Hypothesen und Überblick

- Primäre Ziele der zweifaktoriellen Varianzanalyse sind das Testen von Haupteffekt- und Interaktionshypothesen
- Die Null- und Alternativhypothesen bezüglich der Haupteffektparameter von Faktor A lauten

$$H_0^A : \alpha_i = 0 \text{ für alle } i = 1, \dots, p \text{ und } H_1^A : \alpha_i \neq 0 \text{ für mindestens zwei } i \in \mathbb{N}_p.$$

- Die Null- und Alternativhypothesen bezüglich der Haupteffektparameter von Faktor B lauten

$$H_0^B : \beta_j = 0 \text{ für alle } j = 1, \dots, q \text{ und } H_1^B : \beta_j \neq 0 \text{ für mindestens zwei } j \in \mathbb{N}_q.$$

- Die Null- und Alternativhypothesen bezüglich der Interaktionsparameter lauten

$$H_0^{A \times B} : \gamma_{ij} = 0 \text{ für alle } i = 1, \dots, p, j = 1, \dots, q$$

$$H_1^{A \times B} : \gamma_{ij} \neq 0 \text{ für mindestens zwei } (i, j) \in \mathbb{N}_p \times \mathbb{N}_q.$$

Die Interaktionsalternativhypothese macht keine Aussage über die Form der Interaktion.

- Die Alternativhypothesen sind jeweils zweimal ausgeprägt, da die Effektsummen nach Voraussetzung 0 sind.
- In Analogie zur einfaktoriellen Varianzanalyse können mithilfe entsprechend definierter F-Teststatistiken und ihrer Verteilungen kritische Wert-basierte Tests konstruiert werden. Die F-Teststatistiken beruhen auch im zweifaktoriellen Fall auf einer Quadratsummenzerlegung.

Theorem (Quadratsummenzerlegung bei zweifaktorieller Varianzanalyse)

Für $i = 1, \dots, p$, $j = 1, \dots, q$ und $k = 1, \dots, m$ bezeichne X_{ijk} die k te Stichprobenvariable bei Faktorlevelkombination ij eines zweifaktoriellen Varianzanalysemodells und es seien die Parameterschätzer $\hat{\mu}$, $\hat{\alpha}_i$, $\hat{\beta}_j$, $\hat{\gamma}_{ij}$, $\hat{\sigma}^2$ wie in der Parameterschätzung des Modells definiert gegeben. Schließlich seien

$$\text{SQT} := \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^m (X_{ijk} - \hat{\mu})^2 \quad \begin{array}{ll} \text{die Total} & \text{Sum of Squares} \end{array}$$

$$\text{SQA} := qm \sum_{i=1}^p \hat{\alpha}_i^2 \quad \begin{array}{ll} \text{die Faktor A} & \text{Sum of Squares} \end{array}$$

$$\text{SQB} := pm \sum_{j=1}^q \hat{\beta}_j^2 \quad \begin{array}{ll} \text{die Faktor B} & \text{Sum of Squares} \end{array}$$

$$\text{SQG} := m \sum_{i=1}^p \sum_{j=1}^q \hat{\gamma}_{ij}^2 \quad \begin{array}{ll} \text{die Interaktion} & \text{Sum of Squares} \end{array}$$

$$\text{SQR} := pq(m-1)\hat{\sigma}^2 \quad \begin{array}{ll} \text{die Residual} & \text{Sum of Squares} \end{array}$$

Dann gilt

$$\text{SQT} = \text{SQA} + \text{SQB} + \text{SQG} + \text{SQR}. \quad (42)$$

- Wir verzichten auf einen Beweis.

Definition (F-Teststatistiken für Haupteffekte)

Es seien das Modell der zweifaktoriellen Varianzanalyse sowie die Faktor A Sum of Squares SQA, die Faktor B Sum of Squares SQB und die Residual Sum of Squares SQR wie oben definiert. Dann sind die F-Teststatistiken für die Haupteffekte von Faktor A und Faktor B definiert als

$$F_A := \frac{SQA/(p-1)}{SQR/(pq(m-1))} \text{ und } F_B := \frac{SQB/(q-1)}{SQR/(pq(m-1))} \quad (43)$$

respektive. Die Zähler und Nenner dieser Definitionen werden auch als *Mean SQA*, *Mean SQB*, *Mean SQR* bezeichnet.

Theorem (Verteilungen der F-Teststatistiken für Haupteffekte)

Es seien das Modell der zweifaktoriellen Varianzanalyse sowie die F-Teststatistiken F_A und F_B für die Haupteffekte von Faktor A und Faktor B wie oben definiert. Es gelte zudem

$$H_0^A : \alpha_i = 0 \text{ für alle } i = 1, \dots, p \text{ bzw. } H_0^B : \beta_j = 0 \text{ für alle } j = 1, \dots, q. \quad (44)$$

Dann sind die F-Teststatistiken nach f -Verteilungen mit Freiheitsgradparametern $p-1, pq(m-1)$ bzw. $q-1, pq(m-1)$ verteilt,

$$F_A \sim f(p-1, pq(m-1)) \text{ bzw. } F_B \sim f(q-1, pq(m-1)) \quad (45)$$

- Wir verzichten auf einen Beweis.

Definition (F-Teststatistik für Interaktionseffekte)

Es seien das Modell der zweifaktoriellen Varianzanalyse sowie die Interaktions Sum of Squares SQG und die Residual Sum of Squares SQR wie oben definiert. Dann ist die F-Teststatistik für Interaktionseffekte definiert als

$$F_{A \times B} := \frac{SQG/(p-1)(q-1)}{SQR/(pq(m-1))}. \quad (46)$$

Die Zähler und Nenner dieser Definitionen werden auch als *Mean SQG* und *textitMean SQR* bezeichnet.

Theorem (Verteilung der F-Teststatistik für Interaktionseffekte)

Es seien das Modell der zweifaktoriellen Varianzanalyse sowie die F-Teststatistik $F_{A \times B}$ für Interaktionseffekte wie oben definiert. Es gelte zudem

$$H_0^{A \times B} : \gamma_{ij} = 0 \text{ für alle } i = 1, \dots, p, j = 1, \dots, q \quad (47)$$

Dann ist die F-Teststatistik nach einer f -Verteilung mit Freiheitsgradparametern $(p-1)(q-1), pq(m-1)$ verteilt,

$$F_{A \times B} \sim f((p-1)(q-1), pq(m-1)). \quad (48)$$

- Wir verzichten auf einen Beweis.

Theorem (F-Tests und F-Testumfangkontrolle)

Es sei das Modell der zweifaktoriellen Varianzanalyse gegeben und die F-Teststatistiken F_A , F_B und $F_{A \times B}$ wie oben definiert. Weiterhin seien die kritischen Wert-basierten Tests

$$\phi_A := 1_{\{F_A \geq k_A\}}, \phi_B := 1_{\{F_B \geq k_B\}} \text{ und } \phi_{A \times B} := 1_{\{F_{A \times B} \geq k_{A \times B}\}} \quad (49)$$

definiert. Dann sind ϕ_A , ϕ_B und $\phi_{A \times B}$ genau dann Level- α_0 -Tests mit Testumfang α_0 , wenn

$$\begin{aligned} k_A &:= k_{\alpha_0}^A := \varphi^{-1}(1 - \alpha_0; p - 1, pq(m - 1)) \\ k_B &:= k_{\alpha_0}^B := \varphi^{-1}(1 - \alpha_0; q - 1, pq(m - 1)) \end{aligned} \quad (50)$$

und

$$k_{A \times B} := k_{\alpha_0}^{A \times B} := \varphi^{-1}(1 - \alpha_0; (p - 1)(q - 1), pq(m - 1)), \quad (51)$$

respektive.

- Wir verzichten auf einen Beweis.

Praktisches Vorgehen

- Man hat einen Datensatz x_{ijk} , $i = 1, \dots, p$, $j = 1, \dots, q$, $k = 1, \dots, m$ eines zweifaktoriellen Versuchsplans mit crossed design vorliegen und nimmt an, dass die x_{ijk} Realisationen von $X_{ijk} \sim N(\mu_{ij}, \sigma^2)$ mit unbekannten Parametern μ_{ij} und σ^2 und entsprechenden Haupteffekt und Interaktionsparametern sind.
- Man möchte entscheiden ob die Haupteffekts- und Interaktionsnullhypothesen H_0^A , H_0^B und $H_0^{A \times B}$ eher zutreffen oder eher nicht.
- Man wählt ein Signifikanzniveau α_0 und bestimmt die zugehörigen Freiheitsgradparameter-abhängigen kritischen Werte k_A , k_B und $k_{A \times B}$.
- Anhand der SQA, SQB, SQG und SQR sowie p , q und n berechnet man die realisierten Werte der F-Teststatistiken, die wir hier mit $\tilde{f}_A$, $\tilde{f}_B$ und $\tilde{f}_{A \times B}$ bezeichnen.
- Wenn für $i \in \{A, B, A \times B\}$ $\tilde{f}_i$ größer gleich $k_{\alpha_0}^i$ ist, lehnt man die jeweilige Nullhypothese ab, sonst nicht.
- Die oben entwickelte Theorie garantiert dann, dass man im Mittel in höchstens $\alpha_0 \cdot 100$ von 100 Fällen die jeweilige Nullhypothese fälschlicherweise ablehnt.
- Schließlich ergeben sich die assoziierten p-Werte der realisierten F-Teststatistiken in Analogie zum einfaktoriellen Varianzanalysefall zu

$$\begin{array}{lll} p_A\text{-Wert} & = \mathbb{P} \left(F_A \geq \tilde{f}_A \right) & = 1 - \varphi \left(\tilde{f}_A; p - 1, pq(m - 1) \right) \\ p_B\text{-Wert} & = \mathbb{P} \left(F_B \geq \tilde{f}_B \right) & = 1 - \varphi \left(\tilde{f}_B; q - 1, pq(m - 1) \right) \\ p_{A \times B}\text{-Wert} & = \mathbb{P} \left(F_{A \times B} \geq \tilde{f}_{A \times B} \right) & = 1 - \varphi \left(\tilde{f}_{A \times B}; (p - 1)(q - 1), pq(m - 1) \right) \end{array}$$

Anwendungsbeispiel

	x_{11k}	x_{12k}	x_{21k}	x_{22k}
1	2.86	5.04	3.29	6.88
2	2.62	4.96	3.69	7.33
3	3.01	4.58	2.94	6.76
4	3.20	5.05	3.61	6.57
5	2.75	5.05	3.68	6.84
6	2.88	5.01	3.89	7.45
7	2.85	4.35	3.43	7.06
8	3.23	5.07	3.72	6.99
9	2.99	5.84	3.28	6.93
10	2.57	4.61	3.35	7.46
11	2.69	5.26	2.94	7.07
12	2.83	5.17	3.56	7.03

x_{ijk} bezeichnet den Datenpunkt der k ten experimentellen Einheit ($k = 1, \dots, m$)
im i ten Level von Faktor A und j ten Level von Faktor B ($i = 1, 2, j = 1, 2$)

Anwendungsbeispiel - Manuelle Berechnung

```
# Analyseparameter
p           = 2
q           = 2
m           = 12
n           = 2*2*12
alpha_0     = 0.05

# Mittelwerte
x_bar       = mean(as.matrix(X))
x_bar_i     = colMeans(cbind(as.matrix(c(X$A1B1, X$A1B2)),
                             as.matrix(c(X$A2B1, X$A2B2))))
x_bar_j     = colMeans(cbind(as.matrix(c(X$A1B1, X$A2B1)),
                             as.matrix(c(X$A1B2, X$A2B2))))
x_bar_ij    = colMeans(X)

# Parameterschätzer
mu_hat      = x_bar
alpha_hat_i = x_bar_i - x_bar
beta_hat_j  = x_bar_j - x_bar
gamma_hat_ij = c(x_bar_ij[1] - (mu_hat + alpha_hat_i[1] + beta_hat_j[1]),
                 x_bar_ij[2] - (mu_hat + alpha_hat_i[1] + beta_hat_j[2]),
                 x_bar_ij[3] - (mu_hat + alpha_hat_i[2] + beta_hat_j[1]),
                 x_bar_ij[4] - (mu_hat + alpha_hat_i[2] + beta_hat_j[2]))
sigsqr_hat  = sum((t(as.matrix(X)) - as.vector(x_bar_ij))^2)/(p*q*(m-1))

# Sum of squares
SQT         = mean((as.matrix(X) - x_bar)^2)
SQA         = q*m*sum(alpha_hat_i^2)
SQB         = p*m*sum(beta_hat_j^2)
SQG         = m*sum(gamma_hat_ij^2)
SQR         = p*q*(m-1)*sigsqr_hat

# Anzahl Level Faktor A
# Anzahl Level Faktor B
# Anzahl Datenpunkte pro Zelle
# Gesamtstichprobengröße
# Signifikanzlevel

# Gesamtmittelwert \bar{X}
# Faktor A Mittelwerte \bar{X}_i
# Faktor B Mittelwerte \bar{X}_j
# Gruppenmittelwerte \bar{X}_{ij}

# \mu Schätzer
# \alpha_i Schätzer
# \beta_j Schätzer
# \gamma_{ij} Schätzer

# \sigma^2 Schätzers

# Total      SOS
# Faktor A   SOS
# Faktor B   SOS
# Interaktions SOS
# Residual   SOS
```

Anwendungsbeispiel - Manuelle Berechnung

```
# Mean Sum of Squares
MSA      = SQA/(p-1)                # Mean Square Faktor A
MSB      = SQB/(q-1)                # Mean Square Faktor B
MSG      = SQG/((p-1)*(q-1))        # Mean Square Interaktion
MSR      = SQR/(p*q*(m-1))          # Mean Square Residuals

# Teststatistiken
F_A      = MSA/MSR                  # Faktor A F-Teststatistik
F_B      = MSB/MSR                  # Faktor B F-Teststatistik
F_AB     = MSG/MSR                  # Interaktion F-Teststatistik

# kritische Werte
k_A_alpha_0 = qf(1-alpha_0, p-1, p*q*(m-1)) # k^A_\alpha_0
k_B_alpha_0 = qf(1-alpha_0, q-1, p*q*(m-1)) # k^B_\alpha_0
k_AB_alpha_0 = qf(1-alpha_0, (p-1)*(q-1), p*q*(m-1)) # k^AB_\alpha_0

# Tests
phi       = c(F_A, F_B, F_AB) >= c(k_A_alpha_0, k_B_alpha_0, k_AB_alpha_0) # simultaner Test

# p-Werte
p_values  = 1-c(pf(F_A, p-1, p*q*(m-1)),
                pf(F_B, q-1, p*q*(m-1)),
                pf(F_AB, (p-1)*(q-1), p*q*(m-1)))
```


Anwendungsbeispiel - Manuelle Berechnung

Zweifaktorielle Varianzanalyse

Anzahl Level Faktor A = 2
Anzahl Level Faktor B = 2
Gruppengrößen = 12
Gesamtstichprobengröße = 48
 α_0 = 0.05

Haupteffekt Faktor A
SQA = 20.4
MSA = 20.4
F_A = 233
 k_{A,α_0} = 4.06
 ϕ_A = 1
p_A-Wert = 0

Haupteffekt Faktor B
SQB = 97.9
MSB = 97.9
F_B = 1118
 k_{B,α_0} = 4.06
 ϕ_B = 1
p_B-Wert = 0

Interaktion A x B
SQG = 6.37
MSG = 6.37
F_AxB = 72.8
 k_{AxB,α_0} = 4.06
 ϕ_{AxB} = 1
p_AxB-Wert = 7.07e-11

Anwendungsbeispiel - Automatische Berechnung

```
XAOV = data.frame(X = c(X$A1B1, X$A1B2, X$A2B1, X$A2B2), # Reformatierung des Datensatzes
                  A = as.factor(c(rep(1,2*m), rep(2,2*m))), # Faktor A Level Labels
                  B = as.factor(c(rep(1,m), rep(2,m), rep(1,m), rep(2,m)))) # Faktor B Level Labels

res.aov = aov(X ~ A + B + A:B, data = XAOV) # Benutzung von R's aov Funktion
summary(res.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	20.4	20.4	233.1	< 2e-16 ***
B	1	97.9	97.9	1117.8	< 2e-16 ***
A:B	1	6.4	6.4	72.8	7.1e-11 ***
Residuals	44	3.9	0.1		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Einfaktorielle Varianzanalyse

- Anwendungsszenario und Datendeskription
- Modellformulierung
- Modellschätzung
- Modellevaluation

Zweifaktorielle Varianzanalyse

- Anwendungsszenario und Datendeskription
- Modellformulierung
- Modellschätzung
- Modellevaluation

Selbstkontrollfragen

Einfaktorielle Varianzanalysen

1. Erläutern Sie das Anwendungsszenario für einfaktorielle Varianzanalysen.
2. Definieren Sie das klassische Modell der einfaktoriellen Varianzanalyse in Erwartungswertparameterdarstellung.
3. Definieren Sie das generative Modell der einfaktoriellen Varianzanalyse in Effektdarstellung.
4. Erläutern Sie die Beziehung zwischen dem klassischen Modell der einfaktoriellen Varianzanalyse in Erwartungswertparameterdarstellung und dem generativen Modell der einfaktoriellen Varianzanalyse in Effektdarstellung.
5. Definieren Sie Null- und Alternativhypothese der einfaktoriellen Varianzanalyse.
6. Definieren Sie die Total Sum of Squares, die Between-Group Sum of Squares, und die Within-Group Sum of Squares der einfaktoriellen Varianzanalyse.
7. Erläutern Sie die intuitive Bedeutung der Total Sum of Squares, der Between-Group Sum of Squares, und der Within-Group Sum of Squares in der einfaktoriellen Varianzanalyse.
8. Geben Sie die Quadratsummenzerlegung der einfaktoriellen Varianzanalyse wieder.
9. Definieren Sie die F-Teststatistik der einfaktoriellen Varianzanalyse.
10. Geben Sie das Theorem zur Verteilung der F-Teststatistik bei Zutreffen der Nullhypothese einer einfaktoriellen Varianzanalyse wieder.
11. Definieren Sie den Level- α_0 -Test der einfaktoriellen Varianzanalyse.
12. Erläutern Sie das praktische Vorgehen bei Durchführung einer einfaktoriellen Varianzanalyse.

Selbstkontrollfragen

Zweifaktorielle Varianzanalysen

1. Erläutern Sie das Anwendungsszenario für zweifaktorielle Varianzanalysen.
2. Erläutern Sie die Begriffe Faktor und Faktorlevel bei zweifaktoriellen Varianzanalysen.
3. Wieviele Level besitzt der zweite experimentelle Faktor (Faktor B) einer 2×4 ANOVA?
4. Erläutern Sie die intuitive Bedeutung von Haupt- und Interaktionseffekten einer 2×2 ANOVA.
5. Definieren Sie die klassische Modell der zweifaktoriellen Varianzanalyse in Erwartungswertparameterdarstellung.
6. Definieren Sie das klassische Modell der zweifaktoriellen Varianzanalyse in Effektdarstellung.
7. Definieren Sie das generative Modell der zweifaktoriellen Varianzanalyse in Effektdarstellung.
8. Erläutern Sie die Beziehung zwischen dem klassischen Modell der zweifaktoriellen Varianzanalyse in Erwartungswertparameterdarstellung und dem klassischen Modell der zweifaktoriellen Varianzanalyse in Effektdarstellung.
9. Definieren Sie Null- und Alternativhypothesen einer zweifaktoriellen Varianzanalyse für Haupteffektparameter.
10. Definieren Sie Null- und Alternativhypothesen einer zweifaktoriellen Varianzanalyse für Interaktionsparameter.
11. Definieren Sie die F-Teststatistiken einer zweifaktoriellen Varianzanalyse für Haupteffekte.
12. Geben Sie das Theorem zur Verteilung der F-Teststatistiken für Haupteffekte bei Zutreffen der Nullhypothesen einer zweifaktoriellen Varianzanalyse wieder.
13. Definieren Sie die F-Teststatistiken einer zweifaktoriellen Varianzanalyse für Interaktionseffekte.
14. Geben Sie das Theorem zur Verteilung der F-Teststatistiken für Interaktionseffekte bei Zutreffen der Nullhypothese einer zweifaktoriellen Varianzanalyse wieder.
15. Definieren Sie die Level- α_0 -Tests einer zweifaktoriellen Varianzanalyse.
16. Erläutern Sie das praktische Vorgehen bei Durchführung einer zweifaktoriellen Varianzanalyse.

References

- DeGroot, Morris H., and Mark J. Schervish. 2012. *Probability and Statistics*. 4th ed. Boston: Addison-Wesley.
- Fahrmeir, Ludwig, Christian Heumann, Rita Künstler, Iris Pigeot, and Gerhard Tutz. 2016. *Statistik*. Springer-Lehrbuch. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-50372-0>.