



Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2023/24

Prof. Dr. Dirk Ostwald

Erste Begrifflichkeiten

Wahrscheinlichkeitstheorie

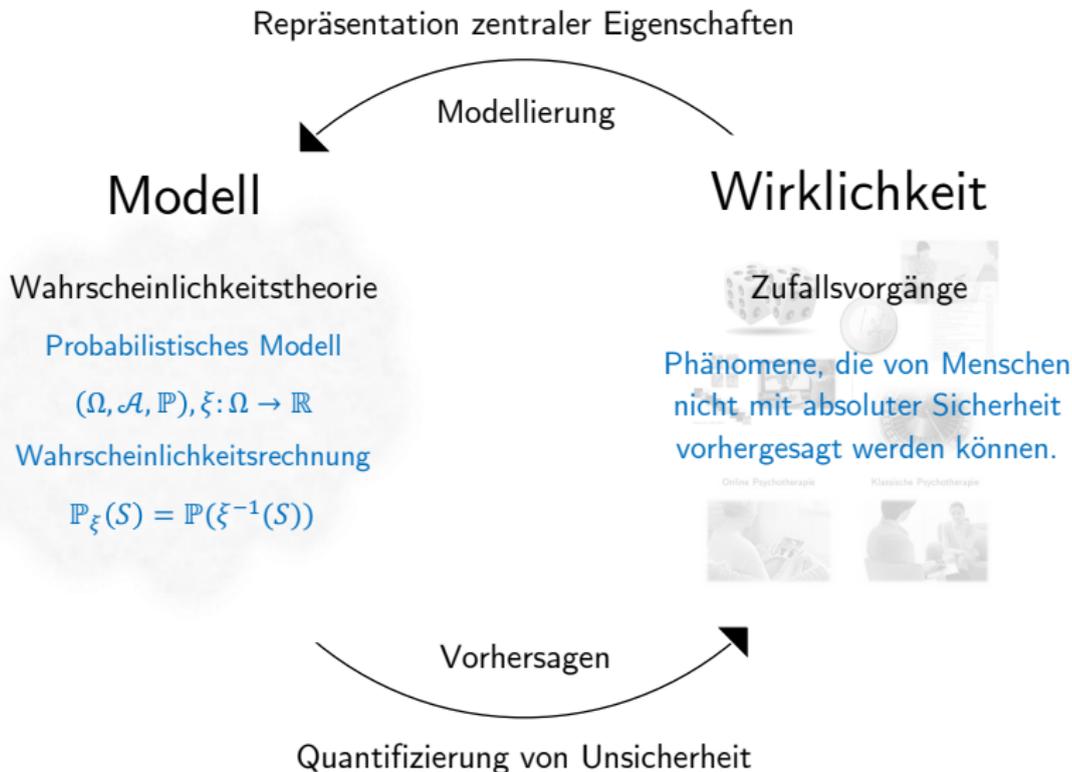
- Ein mathematisches Modell zur Beschreibung von Zufallsvorgängen
- Ein mathematische Modell zum quantitativen Schlussfolgern über Zufallsvorgänge

Zufallsvorgang

- Ein Phänomen der Wirklichkeit, das nicht mit absoluter Sicherheit vorhersagbar ist
- Ein Phänomen der Wirklichkeit, das für Menschen mit Unsicherheit behaftet sind

Grundannahmen der Wahrscheinlichkeitstheorie

- Zufallsvorgänge können durch *Wahrscheinlichkeitsräume* modelliert werden
- Mathematik kann zur Vorhersage zufälliger Ereignisse genutzt werden



Vorbemerkungen zur Wahrscheinlichkeitstheorie

Jede Augenzahl kommt im Mittel gleich häufig vor
Ich denke, jede Augenzahl hat die gleiche Wahrscheinlichkeit

Modell

Wahrscheinlichkeitstheorie

$$\Omega := \{1,2,3,4,5,6\}$$

$$\mathcal{A} := \mathcal{P}(\Omega)$$

$$\mathbb{P}(\{\omega\}) := 1/6$$

Wirklichkeit

Zufallsvorgang

Werfen eines fairen Würfels



Modellierung

Vorhersagen

Wenn ich nicht weiß, ob eine Augenzahl größer als Drei gefallen ist,
dann ist die Wahrscheinlichkeit, dass die Augenzahl gerade ist $1/2$.
Wenn ich weiß, dass eine Augenzahl größer als Drei gefallen ist, dann
ist die Wahrscheinlichkeit dafür, dass die Augenzahl gerade ist $2/3$.

Vorbemerkungen zur Wahrscheinlichkeitstheorie

Zur Interpretation des Wahrscheinlichkeitsbegriffs

In der Wahrscheinlichkeitstheorie sind Wahrscheinlichkeiten anhand ihrer mathematischen Eigenschaften definiert. Die Interpretation des mathematischen Wahrscheinlichkeitsbegriffs ist dabei nicht eindeutig.

Es gibt mindestens zwei unterschiedliche Interpretationen:

Nach der *Frequentistischen Interpretation* ist die Wahrscheinlichkeit eines Ereignisses die idealisierte relative Häufigkeit, mit der ein Ereignis unter den gleichen äußeren Bedingungen einzutreten pflegt. Zum Beispiel ist die Frequentistische Interpretation der Aussage "Mit einer Wahrscheinlichkeit von $1/6$ zeigt der Würfel im nächsten Wurf eine Zwei" die folgende: "Wenn man einen Würfel unendlich oft werfen würde und dabei die relative Häufigkeit des Ereignisses, dass der Würfel eine Zwei zeigt, bestimmen würde, dann wäre diese relative Häufigkeit gleich $1/6$ ".

Nach der *Bayesianischen Interpretation* ist die Wahrscheinlichkeit eines Ereignisses der Grad der Sicherheit, den eine Beobachter:in aufgrund ihrer subjektiven Einschätzung der Lage dem Eintreten eines Ereignisses zumisst. Zum Beispiel ist die Bayesianische Interpretation der Aussage "Mit einer Wahrscheinlichkeit von $1/6$ zeigt der Würfel im nächsten Wurf eine Zwei" die folgende: "Basierend auf meiner eigenen und der tradierten Erfahrung mit dem Werfen eines Würfels bin ich mir zu 16.6% sicher, dass der Würfel beim nächsten Wurf eine Zwei zeigt."

In Modellen von tatsächlich zumindest unter ähnlichen Umständen wiederholbaren Zufallsvorgängen wie dem Werfen eines Würfels ist der Unterschied zwischen Frequentistischer und Bayesianischer Interpretation oft eher subtil. Es gibt aber viele Zufallsvorgänge, die mit Wahrscheinlichkeiten beschrieben werden können, bei denen aufgrund ihrer Einmaligkeit eine Frequentistische Interpretation nicht angemessen ist. Zum Beispiel machen Aussagen der Form "Die Wahrscheinlichkeit dafür, dass die weltweiten Hitzerekorde im Jahr 2023 nicht auf den Klimawandel zurückzuführen sind, ist kleiner als 0.01" nur unter der Bayesianischen Interpretation Sinn, da es sich bei den Temperaturaufzeichnungen des Jahres 2023 nicht um ein wiederholbares Phänomen handelt.

(1) Wahrscheinlichkeitsräume

Definition

Erste Eigenschaften

Wahrscheinlichkeitsfunktionen

Beispiele

Selbstkontrollfragen

Definition

Erste Eigenschaften

Wahrscheinlichkeitsfunktionen

Beispiele

Selbstkontrollfragen

Definition (Wahrscheinlichkeitsraum)

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

- Ω eine beliebige nichtleere Menge von *Ergebnissen* ω ist und *Ergebnismenge* heißt,
- \mathcal{A} eine σ -*Algebra* auf Ω ist und *Ereignissystem* heißt,
- \mathbb{P} eine Abbildung der Form $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ mit den Eigenschaften
 - *Nicht-Negativität* $\mathbb{P}(A) \geq 0$ für alle $A \in \mathcal{A}$,
 - *Normiertheit* $\mathbb{P}(\Omega) = 1$ und
 - σ -*Additivität* $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ für paarweise disjunkte $A_i \in \mathcal{A}$ist und *Wahrscheinlichkeitsmaß* heißt.

Das Tuple (Ω, \mathcal{A}) aus Ergebnismenge und Ereignissystem wird als *Messraum* bezeichnet.

Bemerkung

- Die Definition benutzt den Begriff der σ -*Algebra*.

Definition (σ -Algebra)

Ω sei eine Menge und \mathcal{A} sei eine Menge von Teilmengen von Ω . \mathcal{A} heißt σ -Algebra auf Ω , wenn

- $\Omega \in \mathcal{A}$ ist,
- \mathcal{A} abgeschlossen unter der Bildung von Komplementärmenge ist, also wenn für alle $A \in \mathcal{A}$ gilt, dass auch $A^c := \Omega \setminus A \in \mathcal{A}$ ist,
- \mathcal{A} abgeschlossen unter der abzählbaren Vereinigung von Ereignissen ist, also wenn aus $A_1, A_2, \dots \in \mathcal{A}$ folgt, dass auch $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ ist.

Bemerkungen

- Eine σ -Algebra ist eine Menge von Mengen.
- Eine als bekannt vorausgesetzte andere Menge von Mengen ist die *Potenzmenge*.
- Mengen von Mengen heißen auch *Mengensysteme*.

Definition

ERGEBNISSE DER MATHEMATIK
UND IHRER GRENZGEBIETE
HERAUSGEGEBEN VON DR. SCHRIFFLSTUNG
UND
„ZENTRALBLATT FÜR MATHEMATIK“
ZWEITER BAND

GRUNDBEGRIFFE DER WAHRSCHEINLICHKEITS- RECHNUNG

VON
A. KOLMOGOROFF



BERLIN
VERLAG VON JULIUS SPRINGER
1933

§ 3. Terminologische Vorbemerkungen.

erreichlich. 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 noch keineswegs, daß bei einer sehr großen Anzahl 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. Des unmöglichen Ereignis (der 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 Nichtwerden der Wahrscheinlichkeit nur, daß bei einer ständigen Realisation der Bedingungen ξ 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 wie sehr klein wird (er kann n B. gleich 1) n setze).

§ 3. Terminologische Vorbemerkungen.

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

Mengenlehre.

1. A und B sind disjunkt, d. h. $AB = \emptyset$.
2. $AB \dots N = \emptyset$.
3. $AB \dots N = X$.

In 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 Realisation mindestens eines anderen Ereignisses A, B, \dots, N .
5. Das ereignisgegenwärtige Ereignis \bar{A} besteht in der Nichtrealisation des Ereignisses A .
6. A ist unmöglich.
7. A ist notwendig vorkommend.

* Vgl. § 13. Formel (3).

§ 4. Die elementare Wahrscheinlichkeitsrechnung.

5. Ein System \mathfrak{E} der Mengen A_1, A_2, \dots, A_n bildet eine Zerlegung der Menge E , wenn $A_1 + A_2 + \dots + A_n = E$ ist (das setzt bereits voraus, daß die Mengen A_i paarweise disjunkt sind).
6. B ist eine Untermenge von A : $B \subset A$.
7. Ein Versuch \mathfrak{E} 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 \mathfrak{E} .
8. Aus der Realisation des Ereignisses B folgt notwendig dieselbe von A .

§ 4. Unmittelbare Folgerungen aus den Axiomen, betreffend Wahrscheinlichkeiten, der Satz von Bayes.

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

- (1) $P(A) + P(A') = 1$,
 - (2) $P(A') = 1 - P(A)$.
- Da $E = 0$ ist, erhält man insbesondere
- (3) $P(\emptyset) = 0$.

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

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

(der Additionssatz).

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

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

die bedingte Wahrscheinlichkeit des Ereignisses B unter der Bedingung A . Aus (5) folgt unmittelbar

$$(6) \quad P(AB) = P(A)P_A(B)$$

Ein Induktionschluß ergibt sodann die allgemeine Formel

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

(der Multiplikationssatz).

Man beweist auch leicht folgende Formeln:

- (8) $P(A) \geq 0$,
- (9) $P_A(B) \geq 0$,
- (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 \mathfrak{E} 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 [Mengen, Abbildungen] einzuordnen.“

„Wir haben die eigentlichen Objekte unserer weiteren Betrachtungen - die zufälligen Ereignisse - als Mengen definiert.“

Kolmogoroff (1933) [*1903 †1987]

Ergebnismenge Ω

- Wir betrachten zunächst *endliche Wahrscheinlichkeitsräume* mit $|\Omega| < \infty$.
- Ω habe also nur endlich viele (“diskrete”) Elemente.
- Zum Modellieren des Werfen eines Würfels definiert man z.B. $\Omega := \{1, 2, 3, 4, 5, 6\}$.

Die stillschweigende Mechanik des Wahrscheinlichkeitsraummodells

- Wir stellen uns sequentielle *Durchgänge* eines *Zufallsvorgangs* vor.
- In jedem Durchgang wird genau ein ω aus Ω mit Wahrscheinlichkeit $\mathbb{P}(\{\omega\})$ *realisiert*.
- $\mathbb{P}(\{\omega\})$ bestimmt, mit welcher Wahrscheinlichkeit ω in einem Durchgang aus Ω realisiert wird.
- Beim Modell des Werfens eines fairen Würfels gilt etwa $\mathbb{P}(\{\omega\}) = 1/6$ für alle $\omega \in \Omega$.
- Im 1. Durchgang wird z.B. “4” realisiert, im 2. Durchgang “1”, im 3. Durchgang “5”, usw.

Definition

Ereignisse $A \in \mathcal{A}$

- *Ereignisse* stellt man sich am besten als Zusammenfassung (ein oder) mehrerer Ergebnisse vor.
- Beim Werfen eines Würfels sind mögliche Ereignisse zum Beispiel

Es fällt eine gerade Augenzahl, das heißt $\omega \in \{2, 4, 6\}$

Es fällt eine Augenzahl größer als Zwei, das heißt $\omega \in \{3, 4, 5, 6\}$

Es fällt eine Eins oder eine Fünf, das heißt $\omega \in \{1, 5\}$

- Natürlich sind auch die Ergebnisse $\omega \in \Omega$ mögliche Ereignisse zum Beispiel

Es fällt eine Eins, das heißt $\omega \in \{1\}$

Es fällt eine Sechs, das heißt $\omega \in \{6\}$

- Betrachtet man $\omega \in \Omega$ als Ereignis, so nennt man es *Elementarereignis* und schreibt $\{\omega\}$.

“Wir haben die eigentlichen Objekte unserer weiteren Betrachtungen - die zufälligen Ereignisse - als Mengen definiert.”

Kolmogoroff (1933) [*1903 †1987]

Definition

Ereignissystem \mathcal{A}

- Sinn des Ereignissystems ist es, alle Ereignisse, die sich basierend auf einer gegebenen Ergebnismenge bei Auswahl eines $\omega \in \Omega$ ergeben können, mathematisch zu repräsentieren.
- Das Ereignissystem \mathcal{A} ist die vollständige Menge aller möglichen Ereignisse bei gegebenem Ω .
- Die Forderung, dass \mathcal{A} die σ -Algebra Kriterien erfüllt, begründet sich wie folgt
 - Es soll sichergestellt sein, dass $\omega \in \Omega$ für beliebiges ω , dass also irgendein Ergebnis realisiert wird, eines der möglichen Ereignisse ist. Dies entspricht $\Omega \in \mathcal{A}$. Zu jedem Ereignis soll es auch möglich sein, dass dieses Ereignis gerade nicht eintritt. Dies entspricht, dass aus $A \in \mathcal{A}$ folgen soll, dass $A^c = \Omega \setminus A$ auch in \mathcal{A} ist. Dies impliziert auch, dass $\emptyset = \Omega \setminus \Omega \in \mathcal{A}$. Ein Ereignis ist also, dass kein Elementarereignis eintritt, allerdings passiert dies nur mit Wahrscheinlichkeit Null, $\mathbb{P}(\emptyset) = 0$. Es tritt also sicher immer ein Elementarereignis ein. Die Kombination von Ereignissen soll auch immer ein Ereignis sein, z.B. "Es fällt eine gerade Zahl" und/oder "Es fällt eine Zahl größer 2". Dies entspricht, dass aus $A_1, A_2, \dots \in \mathcal{A}$ folgen soll, dass auch $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.
- Für endliches Ω und für $\Omega := \mathbb{R}$ sind passende Ereignissysteme schon lange bekannt.

Ω ist endlich \Rightarrow Man wählt für \mathcal{A} die Potenzmenge von Ω

Ω ist \mathbb{R} \Rightarrow Man wählt für \mathcal{A} die *Borelsche σ -Algebra* $\mathcal{B}(\mathbb{R})$

Ω ist \mathbb{R}^n \Rightarrow Man wählt für \mathcal{A} die *Borelsche σ -Algebra* $\mathcal{B}(\mathbb{R}^n)$

Theorem (Ereignissystem bei endlicher Ergebnismenge)

$\Omega := \{\omega_1, \omega_2, \dots, \omega_n\}$ mit $n \in \mathbb{N}$ sei eine endliche Menge. Dann ist die Potenzmenge $\mathcal{P}(\Omega)$ von Ω eine σ -Algebra auf Ω und damit ein geeignetes Ereignissystem im Wahrscheinlichkeitsraummodell.

Beweis

Die Potenzmenge von Ω ist die Menge aller Teilmengen von Ω . Wir überprüfen die σ -Algebra Eigenschaften. Zunächst gilt, dass Ω selbst eine der Teilmengen von Ω ist, also ist die erste σ -Algebra Eigenschaft erfüllt. Sei nun A eine Teilmenge von Ω . Dann ist auch $A^c = \Omega \setminus A$ eine Teilmenge von Ω und somit ist auch die zweite σ -Algebra Eigenschaft erfüllt. Schließlich betrachten wir die Vereinigung von n Teilmengen $A_1, A_2, \dots, A_n \subseteq \Omega$. Dann ist $\bigcup_{i=1}^n A_i$ die Menge der $\omega \in \Omega$ für die gilt, dass $\omega \in A_1$ und/oder $\omega \in A_2$... und/oder $\omega \in A_n$. Da für alle diese ω gilt, dass $\omega \in \Omega$ ist also auch $\bigcup_{i=1}^n A_i$ eine Teilmenge von Ω und damit auch die dritte σ -Algebra Eigenschaft erfüllt.

Wahrscheinlichkeitsmaß \mathbb{P}

- (Ω, \mathcal{A}) ist die *strukturelle Basis* eines Wahrscheinlichkeitsraummodells.
- \mathbb{P} repräsentiert die probabilistischen Charakteristika eines Wahrscheinlichkeitsraummodells.
- \mathbb{P} entspricht also der *funktionellen Basis* eines Wahrscheinlichkeitsraummodells.
- Es gilt $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$, \mathbb{P} ordnet also (nur) Mengen Wahrscheinlichkeiten zu.
- Mit $\{\omega\} \in \mathcal{A} \forall \omega \in \Omega$ ordnet \mathbb{P} auch den Elementarereignissen Wahrscheinlichkeiten zu.
- Wahrscheinlichkeiten sind Zahlen in $[0, 1]$, nicht Prozente (20%) oder Verhältnisse (50:50).
- $\mathbb{P}(\Omega) = 1$ entspricht der Tatsache, dass in jedem Durchgang sicher $\omega \in \Omega$ gilt.
- In jedem Durchgang eines Zufallsvorgangs tritt also zumindest ein Elementarereignis ein.

σ -Additivität des Wahrscheinlichkeitsmaßes \mathbb{P}

- Die σ -Additivität von \mathbb{P} erlaubt es, aus bereits bekannten Ereigniswahrscheinlichkeiten die Wahrscheinlichkeiten anderer Ereignisse zu berechnen.
- Die σ -Additivität ist also die Grundlage für das Rechnen mit Wahrscheinlichkeiten, das heißt für die *Wahrscheinlichkeitsrechnung*.
- Man kann basierend auf einer Definition von Ω , \mathcal{A} und \mathbb{P} also Wahrscheinlichkeiten für alle möglichen Ereignisse eines Wahrscheinlichkeitsraummodells berechnen. Ob diese Wahrscheinlichkeiten aber tatsächlich etwas mit den realen Ereignissen in einem Zufallsvorgang zu tun haben, kommt darauf an, ob die Modellierung einigermaßen gelungen ist oder nicht.
- Die hergeleiteten Wahrscheinlichkeiten werden zumindest nach den Regeln der Vernunft, also der Logik und der Mathematik, d.h. rational bestimmt.
- Insgesamt erlaubt das Wahrscheinlichkeitsmodell also das modellbasierte schlussfolgernde Denken über mit Unsicherheit behaftete Phänomene

Probability Theory \Leftrightarrow Reasoning with Uncertainty

Definition

Erste Eigenschaften

Wahrscheinlichkeitsfunktionen

Beispiele

Selbstkontrollfragen

Theorem (Wahrscheinlichkeit des unmöglichen Ereignisses)

$(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum. Dann gilt

$$\mathbb{P}(\emptyset) = 0. \quad (1)$$

Beweis

Für $i = 1, 2, \dots$ sei $A_i := \emptyset$. Dann ist A_1, A_2, \dots eine Folge disjunkter Ereignisse, weil gilt, dass $\emptyset \cap \emptyset = \emptyset$ und es ist $\cup_{i=1}^{\infty} A_i = \emptyset$. Mit der σ -Additivität von \mathbb{P} folgt dann, dass

$$\mathbb{P}(\emptyset) = \mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) = \sum_{i=1}^{\infty} \mathbb{P}(\emptyset) \quad (2)$$

Das unendliche Aufaddieren der Zahl $\mathbb{P}(\emptyset) \in [0, 1]$ soll also wieder $\mathbb{P}(\emptyset)$ ergeben. Dies ist aber nur möglich, wenn $\mathbb{P}(\emptyset) = 0$.

□

Theorem (σ -Additivität bei endlichen Folgen disjunkter Ereignisse)

$(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum und A_1, \dots, A_n sei eine endliche Folge paarweise disjunkter Ereignisse. Dann gilt

$$\mathbb{P}(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i). \quad (3)$$

Beweis

Wir betrachten eine unendliche Folge von paarweise disjunkten Ereignissen A_1, A_2, \dots wobei für ein $n \in \mathbb{N}$ gelten soll, dass $A_i := \emptyset$ für $i > n$. Dann gilt mit der σ -Additivität von \mathbb{P} zunächst, dass

$$\mathbb{P}(\cup_{i=1}^n A_i) = \mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) = \sum_{i=1}^n \mathbb{P}(A_i) + \sum_{i=n+1}^{\infty} \mathbb{P}(A_i). \quad (4)$$

Mit $\mathbb{P}(A_i) = \mathbb{P}(\emptyset) = 0$ für $i = n + 1, n + 2, \dots$ folgt dann direkt

$$\mathbb{P}(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i) + 0 = \sum_{i=1}^n \mathbb{P}(A_i). \quad (5)$$

□

Definition

Erste Eigenschaften

Wahrscheinlichkeitsfunktionen

Beispiele

Selbstkontrollfragen

Definition (Wahrscheinlichkeitsfunktion)

Ω sei eine endliche Menge. Dann heißt eine Funktion $\pi : \Omega \rightarrow [0, 1]$ *Wahrscheinlichkeitsfunktion*, wenn gilt, dass

$$\sum_{\omega \in \Omega} \pi(\omega) = 1. \quad (6)$$

Sei weiterhin \mathbb{P} ein Wahrscheinlichkeitsmaß. Dann heißt die durch

$$\pi : \Omega \rightarrow [0, 1], \omega \mapsto \pi(\omega) := \mathbb{P}(\{\omega\}) \quad (7)$$

definierte Funktion *Wahrscheinlichkeitsfunktion* von \mathbb{P} auf Ω .

Bemerkungen

- Wahrscheinlichkeitsfunktion erlauben im Falle endlicher Ergebnismengen das Festlegen von Wahrscheinlichkeitsmaßen durch die Definition der Wahrscheinlichkeiten der Elementarereignisse.
- Über alle Eingabewerte $\omega \in \Omega$ summieren die Funktionswerte $\pi(\omega)$ zu 1.
- Weil \mathbb{P} per Definition σ -additiv ist, gilt insbesondere auch

$$\mathbb{P}(\Omega) = \mathbb{P}(\cup_{\omega \in \Omega} \{\omega\}) = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} \pi(\omega) = 1. \quad (8)$$

Theorem (Definition eines W-Maßes durch eine W-Funktion)

$(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum mit endlicher Ergebnismenge und $\pi : \Omega \rightarrow [0, 1]$ sei eine Wahrscheinlichkeitsfunktion. Dann existiert ein Wahrscheinlichkeitsmaß \mathbb{P} auf Ω mit π als Wahrscheinlichkeitsfunktion von \mathbb{P} . Dieses Wahrscheinlichkeitsmaß ist definiert als

$$\mathbb{P} : \mathcal{A} \rightarrow [0, 1], A \mapsto \mathbb{P}(A) := \sum_{\omega \in A} \pi(\omega). \quad (9)$$

Bemerkung

- Bei endlichem Ω können die Wahrscheinlichkeiten aller möglichen Ereignisse aus den Wahrscheinlichkeiten der Elementarereignisse $\pi(\omega)$ berechnet werden.

Beweis

Wir überprüfen zunächst die Wahrscheinlichkeitsmaßeigenschaften von \mathbb{P} . Weil $\pi(\omega) \in [0, 1]$ für alle $\omega \in \Omega$, gilt auch immer $\sum_{\omega \in A} \pi(\omega) \geq 0$ und damit die Nicht-Negativität von \mathbb{P} . Ferner folgt wie oben gesehen mit der Normiertheit von π direkt die Normiertheit von \mathbb{P} . Seien nun $A_1, A_2, \dots \in \mathcal{A}$. Dann gilt

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{\omega \in \bigcup_{i=1}^{\infty} A_i} \pi(\omega) = \sum_{i=1}^{\infty} \sum_{\omega \in A_i} \pi(\omega) = \sum_{i=1}^{\infty} \mathbb{P}(A_i). \quad (10)$$

und damit die σ -Additivität von \mathbb{P} . □

Definition

Erste Eigenschaften

Wahrscheinlichkeitsfunktionen

Beispiele

Selbstkontrollfragen

Beispiele

Aus dem bis hierin Gesagtem lässt sich nun zusammenfassend folgendes Vorgehen zur Modellierung eines Zufallsvorganges mithilfe eines Wahrscheinlichkeitsraums $(\Omega, \mathcal{A}, \mathbb{P})$ festhalten:

- (1) In einem ersten Schritt überlegt man sich eine sinnvolle Definition der Ergebnismenge Ω , also der Ergebnisse bzw. Elementarereignisse die in jedem Durchgang des Zufallsvorgangs realisiert werden sollen.
- (2) In einem zweiten Schritt wählt man dann ein geeignetes Ereignissystem; im Falle einer endlichen Ergebnismenge bietet sich die Potenzmenge $\mathcal{P}(\Omega)$ an, im Falle der überabzählbaren Ergebnismenge $\Omega := \mathbb{R}$ bietet sich die Borelsche σ -Algebra $\mathcal{B}(\mathbb{R})$ an.
- (3) Schließlich definiert man ein Wahrscheinlichkeitsmaß \mathbb{P} , dass die Wahrscheinlichkeiten für das Auftreten aller möglichen Ereignisse repräsentiert. Im Falle einer endlichen Ergebnismenge gelingt dies insbesondere wie oben beschrieben durch Definition der Wahrscheinlichkeiten der Elementarereignisse. In der Folge verdeutlichen wir dieses Vorgehen anhand von Beispielen. Im Falle der überabzählbaren Ergebnismenge $\Omega := \mathbb{R}$ bietet sich die Definition von \mathbb{P} mithilfe von Wahrscheinlichkeitsdichtefunktionen an, wie wir später sehen werden.

Beispiele

Würfeln mit einem Würfel

Wir modellieren das Würfeln mit einem Würfel. Es ist sicherlich sinnvoll, die Ergebnismenge als $\Omega := \{1, 2, 3, 4, 5, 6\}$ zu definieren. Allerdings wäre auch die Definition von $\Omega := \{., \dots, \dots, \dots, \dots, \dots\}$ in äquivalenter Weise möglich.

Da es sich um eine endliche Ergebnismenge handelt, wählen wir als σ -Algebra \mathcal{A} die Potenzmenge $\mathcal{P}(\Omega)$. \mathcal{A} enthält dann automatisch alle möglichen Ereignisse. Die Kardinalität von $\mathcal{A} := \mathcal{P}(\Omega)$ ist $|\mathcal{P}(\Omega)| = 2^{|\Omega|} = 2^6 = 64$. In untenstehender Tabelle listen wir sechs dieser 64 Ereignisse in ihrer verbalen Beschreibung und als Teilmenge A von Ω .

Table 1: Ausgewählte Ereignisse beim Modell des Werfens eines Würfels

Beschreibung	Mengenform
Es fällt eine beliebige Augenzahl	$\omega \in A = \Omega$
Keine Augenzahl fällt	$\omega \in A = \emptyset$
Es fällt eine Augenzahl größer als 4	$\omega \in A = \{5, 6\}$
Es fällt eine gerade Augenzahl	$\omega \in A = \{2, 4, 6\}$
Es fällt eine Sechs	$\omega \in A = \{6\}$
Eine Eins, eine Drei oder eine Sechs fällt	$\omega \in A = \{1, 3, 6\}$

Damit ist die Definition des Messraum (Ω, \mathcal{A}) in der Modellierung des Werfens eines Würfels abgeschlossen.

Würfeln mit einem Würfel (fortgesetzt)

Wie oben beschrieben kann das Wahrscheinlichkeitsmaß \mathbb{P} durch Festlegung von $\mathbb{P}(\{\omega\})$ für alle $\omega \in \Omega$ festgelegt werden. Für das Modell eines unverfälschten Würfels würde man

$$\mathbb{P}(\{\omega\}) := \frac{1}{|\Omega|} := 1/6 \text{ für alle } \omega \in \Omega \quad (11)$$

wählen. Für ein Modell eines verfälschten Würfels, der das Werfen einer Sechs bevorzugt, könnte man zum Beispiel definieren, dass

$$\mathbb{P}(\{\omega\}) = \frac{1}{8} \text{ für } \omega \in \{1, 2, 3, 4, 5\} \text{ und } \mathbb{P}(\{6\}) = \frac{3}{8}. \quad (12)$$

Im Fall des unverfälschten Würfels ergibt sich die Wahrscheinlichkeit für das Ereignis "Es fällt eine gerade Augenzahl" mit der σ -Additivität von \mathbb{P} zu

$$\mathbb{P}(\{2, 4, 6\}) = \mathbb{P}(\{2\} \cup \{4\} \cup \{6\}) = \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6}. \quad (13)$$

Im Fall des obigen Modells eines verfälschten Würfels ergibt sich für das gleiche Ereignis die Wahrscheinlichkeit

$$\mathbb{P}(\{2, 4, 6\}) = \mathbb{P}(\{2\} \cup \{4\} \cup \{6\}) = \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = \frac{1}{8} + \frac{1}{8} + \frac{3}{8} = \frac{5}{8}. \quad (14)$$

Beispiele

Gleichzeitiges Würfeln mit einem blauen und einem roten Würfel

Wir wollen nun das gleichzeitige Werfen eines blauen und eines roten Würfels modellieren. Dazu ist es sinnvoll, die Ergebnismenge als

$$\Omega := \{(r, b) | r \in \{1, 2, 3, 4, 5, 6\}, b \in \{1, 2, 3, 4, 5, 6\}\} \quad (15)$$

mit Kardinalität $|\Omega| = 36$ zu definieren, wobei r die Augenzahl des blauen Würfels und b die Augenzahl des roten Würfels repräsentieren soll.

Wiederum bietet sich die Wahl der Potenzmenge von Ω als σ -Algebra an, wir definieren also wieder $\mathcal{A} := \mathcal{P}(\Omega)$. Die Anzahl der in diesem Modell möglichen Ereignisse ergibt sich zu $|\mathcal{A}| = 2^{|\Omega|} = 2^{36} = 68.719.476.736$. In untenstehender Tabelle listen wir sechs dieser Ereignisse in ihrer verbalen Beschreibung und als Teilmenge A von Ω .

Table 2: Ausgewählte Ereignisse beim Modell des Werfens eines roten und eines blauen Würfels

Beschreibung	Mengenform
Auf dem roten Würfel fällt eine Drei	$\omega \in A = \{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\}$
Auf dem blauen Würfel fällt eine Drei	$\omega \in A = \{(1, 3), (2, 3), (3, 3), (4, 3), (5, 3), (6, 3)\}$
Auf beiden Würfeln fällt eine Drei	$\omega \in A = \{(3, 3)\}$
Es fällt eine Pasch	$\omega \in A = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$
Die Summe der gefallen Zahlen ist Vier	$\omega \in A = \{(1, 3), (3, 1), (2, 2)\}$

Gleichzeitiges Würfeln mit einem blauen und einem roten Würfel (fortgesetzt)

Die Definition des Messraum (Ω, \mathcal{A}) ist damit abgeschlossen. Ein Wahrscheinlichkeitsmaß \mathbb{P} kann wiederum durch Definition von $\mathbb{P}(\{\omega\})$ für alle $\omega \in \Omega$ festgelegt werden. Für das Modell zweier unverfälschter Würfel würde man

$$\mathbb{P}(\{\omega\}) := \frac{1}{|\Omega|} = \frac{1}{36} \text{ für alle } \omega \in \Omega \quad (16)$$

wählen. Unter diesem Wahrscheinlichkeitsmaße ergibt sich dann zum Beispiel die Wahrscheinlichkeit für das Ereignis “Die Summe der gefallen Zahlen ist Vier” mit der σ -Additivät von \mathbb{P} zu

$$\begin{aligned} \mathbb{P}(\{(1, 3), (3, 1), (2, 2)\}) &= \mathbb{P}(\{(1, 3)\} \cup \{(3, 1)\} \cup \{(2, 2)\}) \\ &= \mathbb{P}(\{(1, 3)\}) + \mathbb{P}(\{(3, 1)\}) + \mathbb{P}(\{(2, 2)\}) \\ &= 1/36 + 1/36 + 1/36 \\ &= 1/12. \end{aligned}$$

Werfen einer Münze

Wir modellieren das Werfen einer Münze, deren eine Seite Kopf (heads) und deren andere Seite Zahl (tails) zeigt. Es ist sinnvoll, die Ergebnismenge als $\Omega := \{H, T\}$ zu definieren, wobei H "Heads" und T "Tails" repräsentiert. Allerdings wäre auch jede andere binäre Definition von Ω möglich, z.B. $\Omega := \{0, 1\}$, $\Omega := \{-1, 1\}$ oder $\Omega := \{1, 2\}$.

Die Potenzmenge $\mathcal{A} := \mathcal{P}(\Omega)$ enthält alle möglichen Ereignisse. In diesem Fall können wir das gesamte Mengensystem \mathcal{A} sehr leicht komplett auflisten.

Table 3: Ereignissystem \mathcal{A} beim Modell des Werfens einer Münze

Beschreibung	Mengenform
Weder Kopf noch Zahl fällt	$\omega \in A = \emptyset$
Kopf fällt	$\omega \in A = \{H\}$
Zahl fällt	$\omega \in A = \{T\}$
Kopf oder Zahl fällt	$\omega \in A = \{H, T\}$

Die Definition des Messraums (Ω, \mathcal{A}) ist damit abgeschlossen.

Werfen einer Münze (fortgesetzt)

Ein Wahrscheinlichkeitsmaß \mathbb{P} kann wiederum durch Definition von $\mathbb{P}(\{\omega\})$ für alle $\omega \in \Omega$ festgelegt werden. Die Normiertheit von Ω bedingt hier insbesondere, dass

$$\mathbb{P}(\Omega) = 1 \Leftrightarrow \mathbb{P}(\{H\}) + \mathbb{P}(\{T\}) = 1 \Leftrightarrow \mathbb{P}(\{T\}) = 1 - \mathbb{P}(\{H\}). \quad (17)$$

Bei Festlegung der Wahrscheinlichkeit des Elementarereignisses $\{H\}$ wird also die Wahrscheinlichkeit des Elementarereignis $\{T\}$ sofort mit festgelegt (andersherum natürlich ebenso). Für das Modell einer fairen Münze würde man $\mathbb{P}(\{H\}) = \mathbb{P}(\{T\}) = 1/2$ wählen. Die Wahrscheinlichkeiten aller möglichen Ereignisse ergeben in diesem Fall zu

$$\mathbb{P}(\emptyset) = 0, \mathbb{P}(\{H\}) = 1/2, \mathbb{P}(\{T\}) = 1/2 \text{ und } \mathbb{P}(\{H, T\}) = 1. \quad (18)$$

Definition

Erste Eigenschaften

Wahrscheinlichkeitsfunktionen

Beispiele

Selbstkontrollfragen

Selbstkontrollfragen

1. Erläutern Sie den Begriff des Zufallsvorgangs.
2. Geben Sie die Definition des Begriffs der σ -Algebra wieder.
3. Geben Sie die Definition des Begriffs des Wahrscheinlichkeitsmaßes wieder.
4. Geben Sie die Definition des Begriffs des Wahrscheinlichkeitsraums wieder.
5. Erläutern Sie den Begriff der Ergebnismenge Ω .
6. Erläutern Sie die stillschweigende Mechanik des Wahrscheinlichkeitsraummodells.
7. Erläutern Sie den Begriff eines Ereignisses $A \in \mathcal{A}$.
8. Erläutern Sie den Begriff des Ereignissystems \mathcal{A} .
9. Welche σ -Algebra wählt man sinnvoller Weise für einen Wahrscheinlichkeitsraum mit endlicher Ergebnismenge?
10. Erläutern Sie den Begriff des Wahrscheinlichkeitsmaßes \mathbb{P} .
11. Geben Sie die Definition des Begriffs der Wahrscheinlichkeitsfunktion wieder.
12. Warum ist der Begriff der Wahrscheinlichkeitsfunktion bei der Modellierung eines Zufallsvorgangs durch einen Wahrscheinlichkeitsraum mit endlicher Ergebnismenge hilfreich?
13. Erläutern Sie die Modellierung des Werfens eines Würfels mithilfe eines Wahrscheinlichkeitsraums.
14. Erläutern Sie die Modellierung des gleichzeitigen Werfens eines roten und eines blauen Würfels mithilfe eines Wahrscheinlichkeitsraums.

Selbstkontrollfragen - Lösungen

1. Ein Zufallsvorgang ist ein Phänomen der Wirklichkeit, das nicht mit absoluter Sicherheit vorhersagbar ist bzw. für Menschen mit Unsicherheit behaftet ist. Siehe Folie 2 [Erste Begrifflichkeiten](#).
2. Siehe Definition (σ -Algebra).
3. Siehe Definition (Wahrscheinlichkeitsraum) bezüglich \mathbb{P} .
4. Siehe Definition (Wahrscheinlichkeitsraum).
5. Die Ergebnismenge enthält die in einem Durchgang des modellierten Zufallsvorgangs möglichen Elementarereignisse, siehe Folie 12 [Ergebnismenge \$\Omega\$](#) und [Die stillschweigende Mechanik des Wahrscheinlichkeitsraummodells](#).
6. Siehe Folie 12 [Die stillschweigende Mechanik des Wahrscheinlichkeitsraummodells](#).
7. Siehe Folie 13 [Ereignisse \$A \in \mathcal{A}\$](#) .
8. Siehe Folie 14 [Ereignissystem \$\mathcal{A}\$](#) .
9. Die Potenzmenge der Ergebnismenge, sie enthält alle prinzipiell möglichen Ereignisse, siehe auch Theorem (Ereignissystem bei endlicher Ergebnismenge).
10. Siehe Folie 16 [Wahrscheinlichkeitsmaß \$\mathbb{P}\$](#) .
11. Siehe Definition (Wahrscheinlichkeitsfunktion).
12. Durch Definition der Funktionswerte einer Wahrscheinlichkeitsfunktion kann ein Wahrscheinlichkeitsmaß festgelegt werden, siehe auch die [Beispiele](#) auf Folien 25 - 32.
13. Siehe Folien 26 und 27 [Würfeln mit einem Würfel](#).
14. Siehe Folien 28 und 29 [Gleichzeitiges Würfeln mit einem blauen und einem roten Würfel](#).

Kolmogoroff, A. 1933. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin, Heidelberg: Springer Berlin Heidelberg.
<https://doi.org/10.1007/978-3-642-49888-6>.



Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2023/24

Prof. Dr. Dirk Ostwald

(2) Elementare Wahrscheinlichkeiten

Gemeinsame Wahrscheinlichkeiten

Bedingte Wahrscheinlichkeiten

Unabhängige Ereignisse

Selbstkontrollfragen

Gemeinsame Wahrscheinlichkeiten

Bedingte Wahrscheinlichkeiten

Unabhängige Ereignisse

Selbstkontrollfragen

Definition (Gemeinsame Wahrscheinlichkeit)

$(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum und es seien $A, B \in \mathcal{A}$. Dann heißt

$$\mathbb{P}(A \cap B) \tag{1}$$

die *gemeinsame Wahrscheinlichkeit von A und B*.

Bemerkungen

- Zur Abgrenzung nennt man $\mathbb{P}(A)$ auch manchmal auch *totale Wahrscheinlichkeit von A*.
- Intuitiv entspricht $\mathbb{P}(A \cap B)$ der Wahrscheinlichkeit, dass A und B gleichzeitig eintreten.
- In der Mechanik des W-Raummodells ist $\mathbb{P}(A \cap B)$ die Wahrscheinlichkeit, dass in einem Durchgang des Zufallsvorgang ein ω mit sowohl $\omega \in A$ und $\omega \in B$ realisiert wird.

Gemeinsame Wahrscheinlichkeiten

Beispiel

- Wir betrachten das Wahrscheinlichkeitsraummodells des Werfens eines fairen Würfels.
- Wir betrachten die Ereignisse

$A := \{2, 4, 6\}$ Es fällt eine gerade Augenzahl

$B := \{4, 5, 6\}$ Es fällt eine Augenzahl größer als Drei

- Dann gilt $A \cap B = \{4, 6\}$.
- Die Interpretation von $A \cap B = \{4, 6\}$ ist

“Es fällt eine gerade Augenzahl und diese Augenzahl ist größer als Drei.”

- Mit $\mathbb{P}(\{4\}) = \mathbb{P}(\{6\}) = \frac{1}{6}$ und der σ -Additivität von \mathbb{P} ergibt sich

$$\begin{aligned}\mathbb{P}(A \cap B) &= \mathbb{P}(\{2, 4, 6\} \cap \{4, 5, 6\}) \\ &= \mathbb{P}(\{4, 6\}) \\ &= \mathbb{P}(\{4\} \cup \{6\}) \\ &= \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) && (2) \\ &= \frac{1}{6} + \frac{1}{6} \\ &= \frac{1}{3}.\end{aligned}$$

Interpretation von $\mathbb{P}(A \cup B)$

- $\mathbb{P}(A \cap B)$ und $\mathbb{P}(A \cup B)$ sollten nicht verwechselt werden.
- \cup entspricht dem *inkluisiven oder*, also *und/oder*.
- Δ entspricht dem *exklusiven oder*, also *entweder..., oder ..., aber nicht beides*.
- $A \cup B$ entspricht also dem Ereignis, dass A und/oder B eintritt.
- $\omega \in A \cup B$ ist also schon erfüllt, wenn "nur" $\omega \in A$ oder "nur" $\omega \in B$ gilt.
- Für obiges Beispiel gilt

$$A \cup B = \{2, 4, 6\} \cup \{4, 5, 6\} = \{2, 4, 5, 6\} \quad (3)$$

mit der Interpretation

"Es fällt eine gerade Augenzahl oder es fällt eine Augenzahl größer als Drei oder es fällt eine gerade Augenzahl und diese Augenzahl ist größer als Drei".

und der Wahrscheinlichkeit

$$\mathbb{P}(\{2, 4, 5, 6\}) = 4/6 = 2/3. \quad (4)$$

Theorem (Weitere Eigenschaften von Wahrscheinlichkeiten)

$(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum und es seien $A, B \in \mathcal{A}$ Ereignisse. Dann gelten

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

Bemerkungen

- Die Eigenschaften sind in der Analyse probabilistischer Modelle oft nützlich.

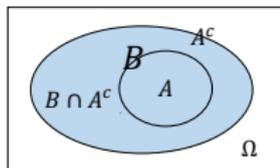
Gemeinsame Wahrscheinlichkeiten

Beweis

Die zweite, dritte, und vierte Aussage dieses Theorems basieren auf elementaren mengentheoretischen Aussagen und der σ -Additivität von \mathbb{P} . Wir wollen diese elementaren mengentheoretischen Aussagen hier nicht beweisen, sondern verweisen jeweils auf die Intuition, die durch die Venn-Diagramme in untenstehender Abbildung vermittelt wird.

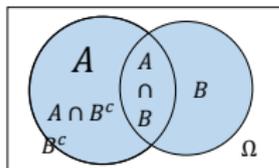
A

$$A \subset B \Rightarrow B = A \cup (B \cap A^c)$$
$$A \cap (B \cap A^c) = \emptyset$$



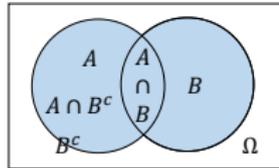
B

$$(A \cap B) \cap (A \cap B^c) = \emptyset$$
$$A = (A \cap B) \cup (A \cap B^c)$$



C

$$B \cap (A \cap B^c) = \emptyset$$
$$A \cup B = B \cup (A \cap B^c)$$



Zu 1.: Wir halten zunächst fest, dass aus $A^c := \Omega \setminus A$ folgt, dass $A^c \cup A = \Omega$ und dass $A^c \cap A = \emptyset$. Mit der Normiertheit und der σ -Additivität von \mathbb{P} folgt dann

$$\mathbb{P}(\Omega) = 1 \Leftrightarrow \mathbb{P}(A^c \cup A) = 1 \Leftrightarrow \mathbb{P}(A^c) + \mathbb{P}(A) = 1 \Leftrightarrow \mathbb{P}(A^c) = 1 - \mathbb{P}(A). \quad (5)$$

Gemeinsame Wahrscheinlichkeiten

Beweis (fortgeführt)

Zu 2.: Zunächst gilt (vgl. Abbildung A)

$$A \subset B \Rightarrow B = A \cup (B \cap A^c) \text{ mit } A \cap (B \cap A^c) = \emptyset. \quad (6)$$

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

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c). \quad (7)$$

Mit $\mathbb{P}(B \cap A^c) \geq 0$ folgt dann $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Zu 3.: Zunächst gilt (vgl. Abbildung B)

$$(A \cap B) \cap (A \cap B^c) = \emptyset \text{ und } A = (A \cap B) \cup (A \cap B^c). \quad (8)$$

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

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) \Leftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A \cap B^c) \quad (9)$$

Zu 4.: Zunächst gilt (vgl. Abbildung C)

$$B \cap (A \cap B^c) = \emptyset \text{ und } A \cup B = B \cup (A \cap B^c). \quad (10)$$

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

$$\mathbb{P}(A \cup B) = \mathbb{P}(B) + \mathbb{P}(A \cap B^c). \quad (11)$$

Mit 3. folgt dann

$$\mathbb{P}(A \cup B) = \mathbb{P}(B) + \mathbb{P}(A) - \mathbb{P}(A \cap B) \quad (12)$$

Zu 5.: Die Aussage folgt direkt aus 4. mit $\mathbb{P}(A \cap B) = \emptyset$ und $\mathbb{P}(\emptyset) = 0$. □

Gemeinsame Wahrscheinlichkeiten

Bedingte Wahrscheinlichkeiten

Unabhängige Ereignisse

Selbstkontrollfragen

Definition (Bedingte Wahrscheinlichkeit)

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

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (13)$$

Weiterhin heißt das für ein festes $B \in \mathcal{A}$ mit $\mathbb{P}(B) > 0$ definierte Wahrscheinlichkeitsmaß

$$\mathbb{P}(\cdot|B) : \mathcal{A} \rightarrow [0, 1], A \mapsto \mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad (14)$$

die *bedingte Wahrscheinlichkeit gegeben Ereignis B* .

Bedingte Wahrscheinlichkeiten

Bemerkungen

- $\mathbb{P}(A|B)$ ist die mit $1/\mathbb{P}(B)$ skalierte gemeinsame Wahrscheinlichkeit $\mathbb{P}(A \cap B)$.
- Die Festlegung der gemeinsamen Wahrscheinlichkeit $\mathbb{P}(A \cap B)$ legt $\mathbb{P}(A|B)$ schon fest.
- Im Gegensatz zu $\mathbb{P}(\cdot \cap B)$ definiert $\mathbb{P}(\cdot|B)$ ein Wahrscheinlichkeitsmaß für alle $A \in \mathcal{A}$.
- Es gelten also
 - $\mathbb{P}(A|B) \geq 0$ für alle $A \in \mathcal{A}$,
 - $\mathbb{P}(\Omega|B) = 1$ und
 - $\mathbb{P}(\cup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} \mathbb{P}(A_i|B)$ für paarweise disjunkte $A_1, A_2, \dots \in \mathcal{A}$.
- \Rightarrow Die Rechenregeln der Wahrscheinlichkeitstheorie 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.

Bedingte Wahrscheinlichkeiten

Beispiel

Wir betrachten erneut das Modell $(\Omega, \mathcal{A}, \mathbb{P})$ des fairen Würfels. Wir wollen die bedingte Wahrscheinlichkeit für das Ereignis "Es fällt eine gerade Augenzahl" gegeben das Ereignis "Es fällt eine Zahl größer als Drei" berechnen. Wir haben oben bereits gesehen, 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 Drei" der Teilmenge $B := \{4, 5, 6\}$ von Ω entspricht. Weiterhin haben wir gesehen, dass unter der Annahme, dass der modellierte Würfel fair ist, gilt, dass

$$\mathbb{P}(\{2, 4, 6\}) = \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2} \quad (15)$$

$$\mathbb{P}(\{4, 5, 6\}) = \mathbb{P}(\{4\}) + \mathbb{P}(\{5\}) + \mathbb{P}(\{6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}. \quad (16)$$

Schließlich hatten wir auch gesehen, dass das Ereignis $A \cap B$, also das Ereignis "Es fällt eine gerade Augenzahl, die größer als Drei ist", die Wahrscheinlichkeit

$$\mathbb{P}(A \cap B) = \mathbb{P}(\{2, 4, 6\} \cap \{4, 5, 6\}) = \mathbb{P}(\{4, 6\}) = \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3} \quad (17)$$

hat. Nach Definition der bedingten Wahrscheinlichkeit ergibt sich dann direkt

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\{4, 6\})}{\mathbb{P}(\{4, 5, 6\})} = \frac{\mathbb{P}(\{4\} \cup \{6\})}{\mathbb{P}(\{4\} \cup \{5\} \cup \{6\})} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{1}{3} \cdot \frac{2}{1} = \frac{2}{3}. \quad (18)$$

Wenn man also weiß, dass eine Augenzahl größer als Drei 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) $1/2$. Dies Beispiel verdeutlicht insbesondere, dass das Bedingen auf einem Ereignis der Inkorporation von neuem Wissen in wahrscheinlichkeitstheoretische Modellen entspricht.

Theorem (Gemeinsame und bedingte Wahrscheinlichkeiten)

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

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A). \quad (19)$$

Beweis

Mit der Definition der jeweiligen bedingten Wahrscheinlichkeit folgen direkt

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \Leftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) \quad (20)$$

und

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} \Leftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A). \quad (21)$$

Bemerkung

- Gemeinsame Wahrscheinlichkeiten können aus bedingten und totalen Wahrscheinlichkeiten berechnet werden.
- Definition von $\mathbb{P}(A)$ und $\mathbb{P}(B|A)$ definiert $\mathbb{P}(A \cap B)$ implizit mit.

Bedingte Wahrscheinlichkeiten

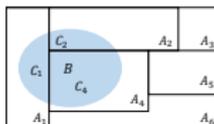
Theorem (Gesetz der totalen Wahrscheinlichkeit)

$(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum und A_1, \dots, A_k sei eine Partition von Ω . Dann gilt für jedes $B \in \mathcal{A}$, dass

$$\mathbb{P}(B) = \sum_{i=1}^k \mathbb{P}(B|A_i)\mathbb{P}(A_i). \quad (22)$$

Beweis

Für $i = 1, \dots, k$ sei $C_i := B \cap A_i$, so dass $\cup_{i=1}^k C_i = B$ und $C_i \cap C_j = \emptyset$ für $1 \leq i, j \leq k, i \neq j$, visuell



Also gilt

$$\mathbb{P}(B) = \sum_{i=1}^k \mathbb{P}(C_i) = \sum_{i=1}^k \mathbb{P}(B \cap A_i) = \sum_{i=1}^k \mathbb{P}(B|A_i)\mathbb{P}(A_i). \quad (23)$$

Bemerkung

- Totale Wahrscheinlichkeiten können aus bedingten und totalen Wahrscheinlichkeiten berechnet werden.
- $\mathbb{P}(B)$ entspricht der gewichteten Summe der bedingten Wahrscheinlichkeiten $\mathbb{P}(B|A_i)$ mit Gewichten $\mathbb{P}(A_i)$.

Theorem (Theorem von Bayes)

$(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum und A_1, \dots, A_k sei eine Partition von Ω mit $\mathbb{P}(A_i) > 0$ für alle $i = 1, \dots, k$. Wenn $\mathbb{P}(B) > 0$ gilt, dann gilt für jedes $i = 1, \dots, k$, dass

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{i=1}^k \mathbb{P}(B|A_i)\mathbb{P}(A_i)}. \quad (24)$$

Beweis

Mit der Definition der bedingten Wahrscheinlichkeit und dem Gesetz der totalen Wahrscheinlichkeit gilt

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{i=1}^k \mathbb{P}(B|A_i)\mathbb{P}(A_i)}. \quad (25)$$

□

Bemerkungen

- Das Theorem von Bayes ist eine zu $\mathbb{P}(A_i \cap B)/\mathbb{P}(B)$ alternative Formel, um die bedingte Wahrscheinlichkeit $\mathbb{P}(A_i|B)$ zu berechnen.
- Das Theorem von Bayes gilt unabhängig von der Frequentistischen oder Bayesianischen Interpretation von Wahrscheinlichkeiten.
- Im Rahmen der *Frequentistischen Inferenz* wird das Theorem von Bayes recht selten benutzt; hier steht vor allem die Tatsache $\mathbb{P}(A) = \mathbb{P}(A|B)$ bei Unabhängigkeit von A und B im Vordergrund.
- Im Rahmen der *Bayesianischen Inferenz* ist das Theorem von Bayes zentral; hier wird $\mathbb{P}(A_i)$ oft *Prior Wahrscheinlichkeit* und $\mathbb{P}(A_i|B)$ oft *Posterior Wahrscheinlichkeit des Ereignisses A_i* genannt. Wie oben erläutert entspricht $\mathbb{P}(A_i|B)$ dann der aktualisierten Wahrscheinlichkeit von A_i , wenn man um das Eintreten von B weiß.

Gemeinsame Wahrscheinlichkeiten

Bedingte Wahrscheinlichkeiten

Unabhängige Ereignisse

Selbstkontrollfragen

Definition (Unabhängige Ereignisse)

Zwei Ereignisse $A \in \mathcal{A}$ and $B \in \mathcal{A}$ heißen *unabhängig*, wenn

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \quad (26)$$

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

$$\mathbb{P}(\cap_{j \in J} A_j) = \prod_{j \in J} \mathbb{P}(A_j). \quad (27)$$

Bemerkungen

- Die Unabhängigkeit bestimmter Ereignissen kann in der Definition eines probabilistischen Modells vorausgesetzt werden, so zum Beispiel die Unabhängigkeit von Fehlervariablen in statistischen Modellen.
- 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 $A_i, i \in I$.
- Der Sinn der Produkteigenschaft bei Unabhängigkeit erschließt sich im Kontext *bedingter Wahrscheinlichkeiten*.

Theorem (Bedingte Wahrscheinlichkeit unter Unabhängigkeit)

$(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{A}$ seien unabhängige Ereignisse mit $\mathbb{P}(B) > 0$.
Dann gilt

$$\mathbb{P}(A|B) = \mathbb{P}(A). \quad (28)$$

Beweis

Unter den Voraussetzungen des Theorems gilt

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A). \quad (29)$$

□

Bemerkungen

- Bei Unabhängigkeit von A und B ist es irrelevant, ob auch B eintritt, $\mathbb{P}(A)$ bleibt gleich.
- Unabhängigkeit wird also als $\mathbb{P}(A)\mathbb{P}(B)$ modelliert, damit $\mathbb{P}(A|B) = \mathbb{P}(A)$ gilt.
- Die Unabhängigkeit zweier Ereignisse bedeutet also, dass das Wissen um das Eintreten eines der beiden Ereignisse die Wahrscheinlichkeit für das Eintreten des anderen Ereignisses nicht ändert. Andersherum bedeutet Abhängigkeit, dass das Wissen um das Eintreten eines der beiden Ereignisse die Wahrscheinlichkeit für das Eintreten des anderen Ereignisses verändert, also erhöht oder verringert.

Gemeinsame Wahrscheinlichkeiten

Bedingte Wahrscheinlichkeiten

Unabhängige Ereignisse

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die Definition der gemeinsamen Wahrscheinlichkeit zweier Ereignisse wieder.
2. Erläutern Sie die intuitive Bedeutung der gemeinsamen Wahrscheinlichkeit zweier Ereignisse.
3. Geben Sie das Theorem zu weiteren Eigenschaften von Wahrscheinlichkeiten wieder.
4. Geben Sie die Definition der bedingten Wahrscheinlichkeit eines Ereignisses wieder.
5. Geben Sie die Definition der bedingten Wahrscheinlichkeit wieder.
6. Geben Sie das Theorem zu gemeinsamen und bedingten Wahrscheinlichkeiten wieder.
7. Geben Sie das Gesetz der totalen Wahrscheinlichkeit wieder.
8. Geben Sie das Theorem von Bayes wieder.
9. Geben Sie den Beweis des Theorems von Bayes wieder.
10. Geben Sie die Definition der Unabhängigkeit zweier Ereignisse wieder.
11. Geben Sie das Theorem zur bedingten Wahrscheinlichkeit unter Unabhängigkeit wieder.
12. Geben Sie den Beweis des Theorems zur bedingten Wahrscheinlichkeit unter Unabhängigkeit wieder.
13. Erläutern Sie das Theorem zur bedingten Wahrscheinlichkeit unter Unabhängigkeit.

Selbstkontrollfragen - Lösungen

1. Siehe Definition (Gemeinsame Wahrscheinlichkeit).
2. Siehe Bemerkungen zu Definition (Gemeinsame Wahrscheinlichkeit).
3. Siehe Theorem (Weitere Eigenschaften von Wahrscheinlichkeiten).
4. Siehe Definition (Bedingte Wahrscheinlichkeit). Die bedingte Wahrscheinlichkeit eines Ereignisses gegeben ein anderes Ereignis ist ein Wert in $[0, 1]$.
5. Siehe Definition (Bedingte Wahrscheinlichkeit). Die bedingte Wahrscheinlichkeit gegeben ein Ereignis ist ein Wahrscheinlichkeitsmaß.
6. Siehe Theorem (Gemeinsame und bedingte Wahrscheinlichkeit).
7. Siehe Theorem (Gesetz der totalen Wahrscheinlichkeit).
8. Siehe Theorem (Theorem von Bayes).
9. Siehe Beweis zum Theorem (Theorem von Bayes).
10. Siehe Definition (Unabhängige Ereignisse).
11. Siehe Theorem (Bedingte Wahrscheinlichkeit unter Unabhängigkeit).
12. Siehe Beweis zum Theorem (Bedingte Wahrscheinlichkeit unter Unabhängigkeit).
13. Siehe Bemerkungen zum Theorem (Bedingte Wahrscheinlichkeit unter Unabhängigkeit).



Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2023/24

Prof. Dr. Dirk Ostwald

(3) Zufallsvariablen

Konstruktion, Definition, Notation, Intuition

Wahrscheinlichkeitsmassfunktionen

Wahrscheinlichkeitsdichtefunktionen

Kumulative Verteilungsfunktionen

Selbstkontrollfragen

Konstruktion, Definition, Notation, Intuition

Wahrscheinlichkeitsmassfunktionen

Wahrscheinlichkeitsdichtefunktionen

Kumulative Verteilungsfunktionen

Selbstkontrollfragen

Konstruktion von Zufallsvariablen und Verteilungen

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

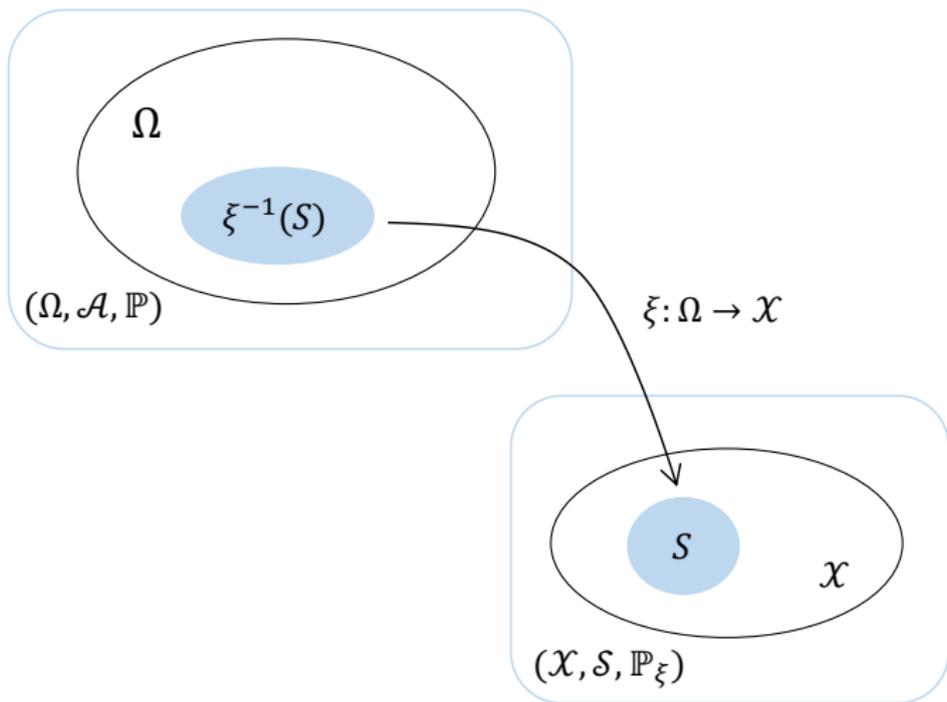
$$\xi^{-1}(S) := \{\omega \in \Omega \mid \xi(\omega) \in S\}. \quad (1)$$

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

$$\mathbb{P}_\xi : \mathcal{S} \rightarrow [0, 1], S \mapsto \mathbb{P}_\xi(S) := \mathbb{P}(\xi^{-1}(S)) = \mathbb{P}(\{\omega \in \Omega \mid \xi(\omega) \in S\}) \quad (2)$$

zugeordnet werden.

- ξ heißt nun *Zufallsvariable* und \mathbb{P}_ξ heißt *Bildmaß* oder *Verteilung von ξ* .
- $(\mathcal{X}, \mathcal{S}, \mathbb{P}_\xi)$ ist ein Wahrscheinlichkeitsraum.
- Mit $\mathcal{X} := \mathbb{R}$ und $\mathcal{S} := \mathcal{B}(\mathbb{R})$ rückt der Wahrscheinlichkeitsraum $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_\xi)$ ins Zentrum.



$$\mathbb{P}(\xi^{-1}(S)) = \mathbb{P}(\{\omega \in \Omega \mid \xi(\omega) \in S\}) =: \mathbb{P}_\xi(S)$$

Definition (Zufallsvariable)

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

$$\{\omega \in \Omega \mid \xi(\omega) \in S\} \in \mathcal{A} \text{ für alle } S \in \mathcal{S}. \quad (3)$$

Bemerkungen

- ZVen sind weder “zufällig” noch “Variablen”.
- Intuitiv wird $\omega \in \Omega$ “zufällig” anhand von \mathbb{P} gezogen und $\xi(\omega)$ realisiert.
- Wir nennen \mathcal{X} den *Ergebnisraum der ZV* ξ .
- 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 eines roten und eines blauen Würfels)

- Für das Werfen eines roten und eines blauen Würfels ist ein sinnvolles W-Raummodell
 - $\Omega := \{(r, b) | r \in \mathbb{N}_6, b \in \mathbb{N}_6\}$
 - $\mathcal{A} := \mathcal{P}(\Omega)$.
 - $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ mit $\mathbb{P}(\{(r, b)\}) = \frac{1}{36}$ für alle $(r, b) \in \Omega$.
- Die Augenzahl-Summenbildung wird dann sinnvoller Weise durch die Zufallsvariable

$$\xi : \Omega \rightarrow \mathcal{X}, (r, b) \mapsto \xi((r, b)) := r + b. \quad (4)$$

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}_ξ von ξ für alle Elementarereignisse $\{x\} \in \mathcal{S}$ berechnen, wie auf der nächsten Folie gezeigt.
- Wir haben damit ein weiteres Wahrscheinlichkeitsraum-Modell $(\mathcal{X}, \mathcal{S}, \mathbb{P}_\xi)$ konstruiert.

Beispiel (Summe eines roten und eines blauen Würfels)

Bestimmung der Verteilung \mathbb{P}_ξ von ξ durch Bestimmung der Elementarereigniswahrscheinlichkeiten

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

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

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

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

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

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

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

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

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

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

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

Definition (Notation für Zufallsvariablen)

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

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

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

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

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

Aus diesen Konventionen folgen exemplarisch die folgenden Konventionen für Verteilungen:

$$\mathbb{P}_\xi(\xi \in S) = \mathbb{P}(\{\xi \in S\}) = \mathbb{P}(\{\omega \in \Omega \mid \xi(\omega) \in S\}), S \subset \mathcal{X}$$

$$\mathbb{P}_\xi(\xi \leq x) = \mathbb{P}(\{\xi \leq x\}) = \mathbb{P}(\{\omega \in \Omega \mid \xi(\omega) \leq x\}), x \in \mathcal{X}.$$

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

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

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

Definition (Realisierung einer Zufallsvariable)

$(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum $(\mathcal{X}, \mathcal{S})$ sei ein Messraum und $\xi : \Omega \rightarrow \mathcal{X}$ sei eine Zufallsvariable. Dann heißt $\xi(\omega) \in \mathcal{X}$ auch *Realisierung der Zufallsvariable*.

- In der Datenanalyse werden Daten typischerweise als Realisierungen von Zufallsvariablen modelliert.
- Da die Auswahl eines $\omega \in \Omega$ in einem Zufallsvorgang zufällig ist, erscheint $\xi(\omega)$ zufällig.

Simulation von Zufallsvariablenrealisierungen (Summe zweier Würfel)

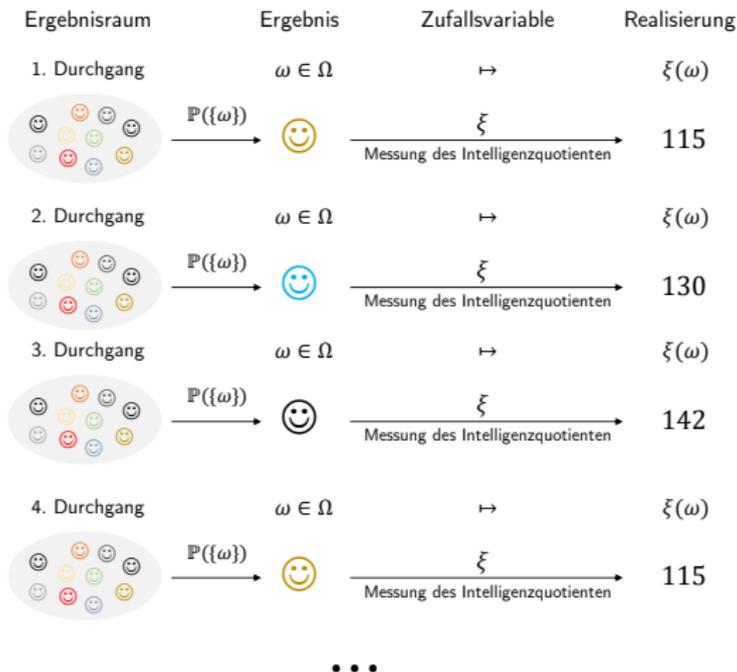
```
# Wahrscheinlichkeitsraummodell
Omega = list() # Ergebnisrauminitialisierung
idx = 1 # Ergebnisindexinitialisierung
for(r in 1:6){ # Ergebnisse roter Würfel
  for(b in 1:6){ # Ergebnisse blauer Würfel
    Omega[[idx]] = c(r,b) # \omega \in \Omega
    idx = idx + 1 } # Ergebnisindexupdate
K = length(Omega) # Kardinalität von \Omega
pi = rep(1/K,1,K) # Wahrscheinlichkeitsfunktion \pi

# Zufallsvorgang
omega = Omega[[which(rmultinom(1,1,pi) == 1)]] # Auswahl von \omega anhand \mathbb{P}(\{\omega\})

# Auswertung der Zufallsvariable
xi_omega = sum(omega) # \xi(\omega)

omega : 1 3
xi(omega) : 4
```

Zufallsvariablen als Modelle von Messvorgängen



Theorem (Arithmetik reeller Zufallsvariablen)

$(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum, $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ sei der reelle Messraum, $\xi : \Omega \rightarrow \mathbb{R}$, $v : \Omega \rightarrow \mathbb{R}$ seien reellwertige Zufallsvariablen und $c \in \mathbb{R}$ sei eine Konstante. Weiterhin seien

$$\begin{aligned}\xi + c : \Omega &\rightarrow \mathbb{R}, \omega \mapsto (\xi + c)(\omega) := \xi(\omega) + c \text{ für } c \in \mathbb{R} \\ c\xi : \Omega &\rightarrow \mathbb{R}, \omega \mapsto (c\xi)(\omega) := c\xi(\omega) \text{ für } c \in \mathbb{R} \\ \xi + v : \Omega &\rightarrow \mathbb{R}, \omega \mapsto (\xi + v)(\omega) := \xi(\omega) + v(\omega) \\ \xi v : \Omega &\rightarrow \mathbb{R}, \omega \mapsto (\xi v)(\omega) := \xi(\omega)v(\omega)\end{aligned}\tag{5}$$

die Addition einer Konstante zu einer reellwertigen Zufallsvariable, die Multiplikation einer reellwertigen Zufallsvariable mit einer Konstante, die Addition zweier reellwertiger Zufallsvariablen und die Multiplikation zweier reellwertigen Zufallsvariablen, respektive. Dann sind auch $\xi + c$, $c\xi$, $\xi + v$ und ξv reellwertige Zufallsvariablen.

Bemerkungen

- Intuitiv ergibt die Addition einer zufälligen Größe zu einer konstanten Größe eine zufällige Größe.
- Intuitiv ergibt die Multiplikation einer zufälligen Größe mit einer Konstante eine zufällige Größe.
- Intuitiv ergibt die Addition zweier zufälliger Größen wiederum eine zufällige Größe.
- Intuitiv ergibt die Multiplikation zweier zufälliger Größen wiederum eine zufällige Größe.
- Für einen Beweis, siehe zum Beispiel Hesse (2009), Seite 33-34.

Konstruktion, Definition, Notation, Intuition

Wahrscheinlichkeitsmassfunktionen

Wahrscheinlichkeitsdichtefunktionen

Kumulative Verteilungsfunktionen

Selbstkontrollfragen

Definition (Diskrete ZV, Wahrscheinlichkeitsmassfunktion)

Eine Zufallsvariable ξ heißt *diskret*, wenn ihr Ergebnisraum \mathcal{X} endlich oder abzählbar ist und eine Funktion der Form

$$p : \mathcal{X} \rightarrow [0, 1], x \mapsto p(x) \quad (6)$$

existiert, für die gilt

(1) $\sum_{x \in \mathcal{X}} p(x) = 1$ und

(2) $\mathbb{P}_\xi(\xi = x) = p(x)$ für alle $x \in \mathcal{X}$.

Eine entsprechende Funktion p heißt *Wahrscheinlichkeitsmassfunktion (WMF)* von ξ .

Bemerkungen

- Eine Menge heißt abzählbar, wenn sie bijektiv auf \mathbb{N} abgebildet werden kann.
- WMFen heißen im Deutschen auch *W-Funktionen* oder *Zähldichten*.
- WMFen heißen auf Englisch *probability mass functions (PMFs)*.
- Die Eigenschaft $\sum_{x \in \mathcal{X}} p(x) = 1$ nennt man auch *Normiertheit* von p .
- Zur Parallelität mit PMFs und WDFs bevorzugt wird den Begriff WMF.

Definition (Bernoulli-Zufallsvariable)

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

$$p : \mathcal{X} \rightarrow [0, 1], x \mapsto p(x) := \mu^x(1 - \mu)^{1-x} \text{ mit } \mu \in [0, 1]. \quad (7)$$

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

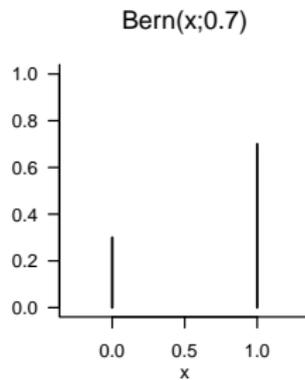
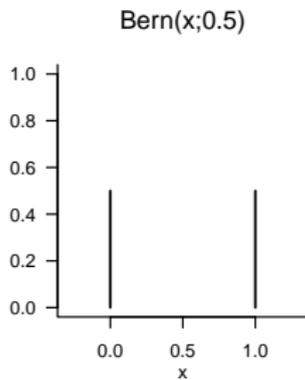
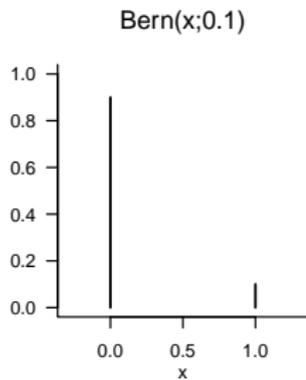
$$\text{Bern}(x; \mu) := \mu^x(1 - \mu)^{1-x}. \quad (8)$$

Bemerkungen

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

$$\mathbb{P}(\xi = 1) = \mu^1(1 - \mu)^{1-1} = \mu. \quad (9)$$

Bernoulli-Zufallsvariable



Definition (Binomial-Zufallsvariable)

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

$$p : \mathcal{X} \rightarrow [0, 1], x \mapsto p(x) := \binom{n}{x} \mu^x (1 - \mu)^{n-x} \text{ für } \mu \in [0, 1]. \quad (10)$$

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

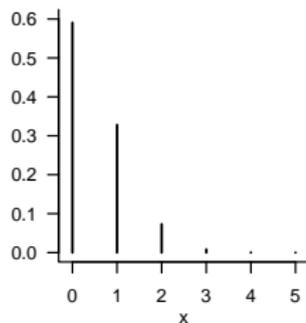
$$\text{Bin}(x; \mu, n) := \binom{n}{x} \mu^x (1 - \mu)^{n-x}. \quad (11)$$

Bemerkung

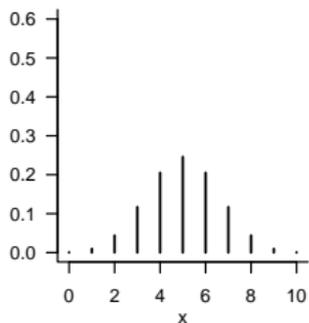
- Es gilt $\text{Bin}(x; \mu, 1) = \text{Bern}(x; \mu)$.

Binomial-Zufallsvariable

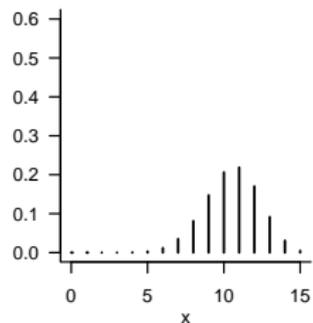
Bin($x;0.1,5$)



Bin($x;0.5,10$)



Bin($x;0.7,15$)



Definition (Diskret-gleichverteilte Zufallsvariable)

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

$$p : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}, x \mapsto p(x) := \frac{1}{|\mathcal{X}|}. \quad (12)$$

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

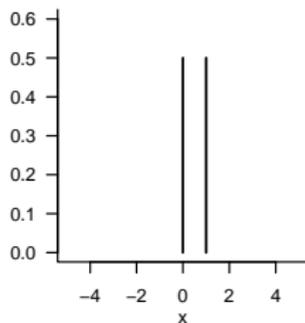
$$U(x; |\mathcal{X}|) := \frac{1}{|\mathcal{X}|}. \quad (13)$$

Bemerkungen

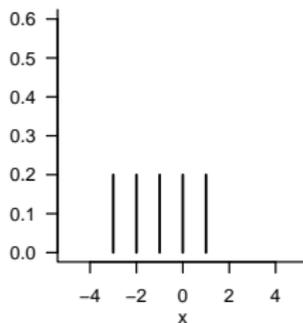
- $\text{Bern}(x; 0.5) = U(x; |\mathcal{X}|)$ für $\mathcal{X} = \{0, 1\}$.

Diskret-gleichverteilte Zufallsvariable

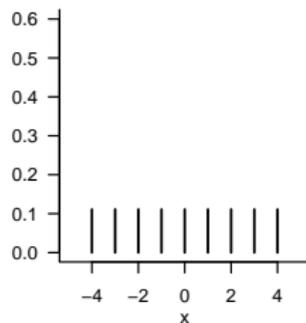
$U(x; X = \{0,1\})$



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



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



Konstruktion, Definition, Notation, Intuition

Wahrscheinlichkeitsmassfunktionen

Wahrscheinlichkeitsdichtefunktionen

Kumulative Verteilungsfunktionen

Selbstkontrollfragen

Definition (Kontinuierliche ZV, Wahrscheinlichkeitsdichtefunktion)

Eine Zufallsvariable ξ heißt *kontinuierlich*, wenn \mathbb{R} der Ergebnisraum von ξ ist und eine Funktion

$$p : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, x \mapsto p(x) \quad (14)$$

existiert, für die gilt

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

(2) $\mathbb{P}_{\xi}(\xi \in [a, b]) = \int_a^b p(x) dx$ für alle $a, b \in \mathbb{R}$ mit $a \leq b$.

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

Bemerkungen

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

Definition (Normalverteilte und standardnormalverteilte Zufallsvariablen)

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

$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, x \mapsto p(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right). \quad (15)$$

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

$$N(x; \mu, \sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right). \quad (16)$$

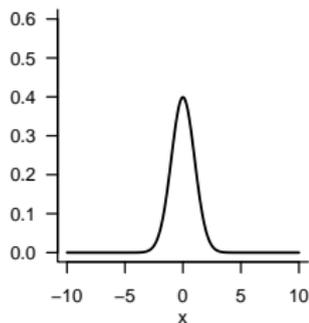
Eine normalverteilte Zufallsvariable mit $\mu = 0$ und $\sigma^2 = 1$ heißt *standardnormalverteilte Zufallsvariable* und wird oft als *Z-Zufallsvariable* bezeichnet.

Bemerkungen

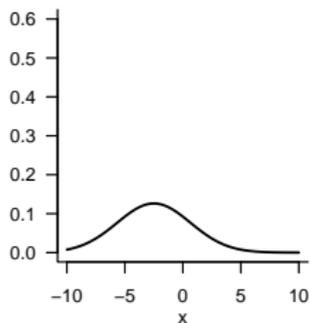
- Der Parameter μ entspricht dem Wert höchster Wahrscheinlichkeitsdichte.
- Der Parameter σ^2 spezifiziert die Breite der WDF.

Normalverteilte Zufallsvariablen

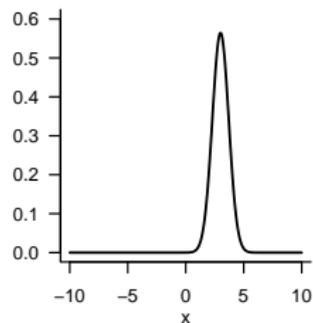
$N(x; 0,1)$



$N(x; -2.5,10)$



$N(x; 3,0.5)$



Definition (Gamma-Zufallsvariable)

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

$$p : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}, x \mapsto p(x) := \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right), \quad (17)$$

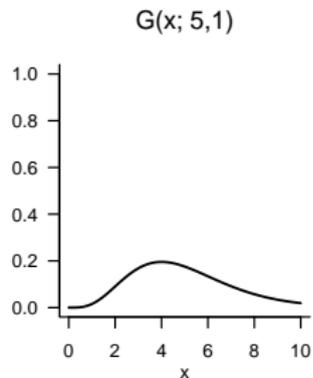
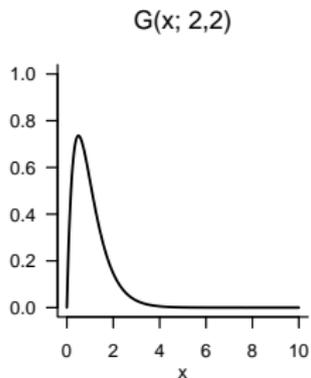
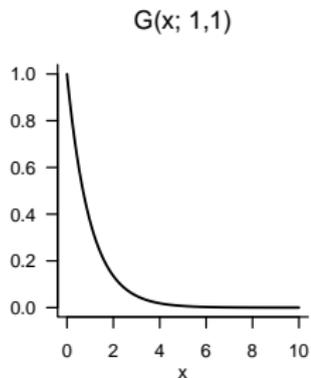
wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass ξ einer *Gammaverteilung mit Formparameter $\alpha > 0$ und Skalenparameter $\beta > 0$* unterliegt und nennen ξ eine *gammaverteilte Zufallsvariable*. Wir kürzen dies mit $\xi \sim G(\alpha, \beta)$ ab. Die WDF einer gammaverteilten Zufallsvariable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

$$G(x; \alpha, \beta) := \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right). \quad (18)$$

Bemerkung

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

Gamma-Zufallsvariablen



Definition (Beta-Zufallsvariable)

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

$$p : \mathcal{X} \rightarrow [0, 1], x \mapsto p(x) := \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \text{ mit } \alpha, \beta \in \mathbb{R}_{>0}, \quad (19)$$

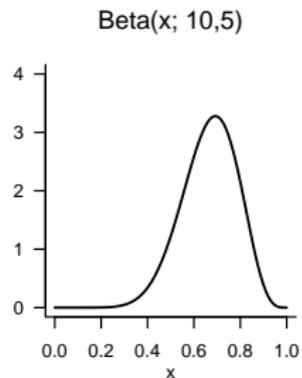
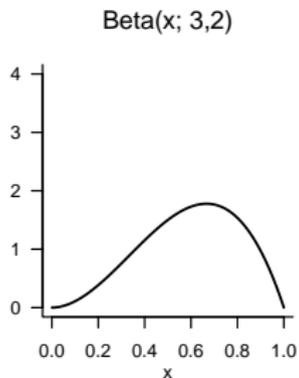
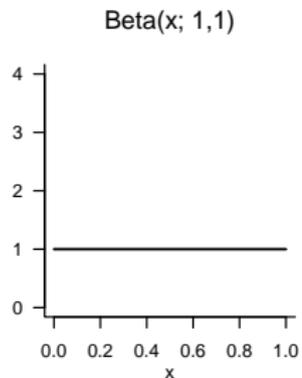
wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass ξ einer *Beta-Verteilung* mit Parametern $\alpha > 0$ und $\beta > 0$ unterliegt, und nennen ξ eine *beta-verteilte Zufallsvariable*. Wir kürzen dies mit $\xi \sim \text{Beta}(\alpha, \beta)$ ab. Die WDF einer beta-verteilten Zufallsvariable bezeichnen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

$$\text{Beta}(x; \alpha, \beta) := \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}. \quad (20)$$

Bemerkung

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

Beta-Zufallsvariable



Definition (Gleichverteilte Zufallsvariable)

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

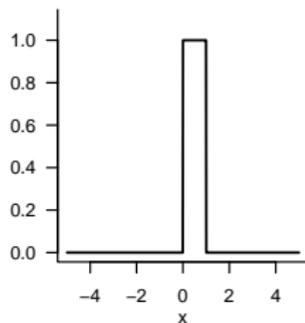
$$p : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, x \mapsto p(x) := \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & x \notin [a, b] \end{cases}. \quad (21)$$

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

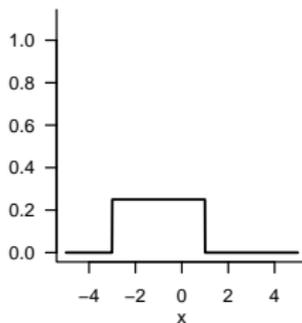
$$U(x; a, b) := \frac{1}{b-a}. \quad (22)$$

Gleichverteilte Zufallsvariablen

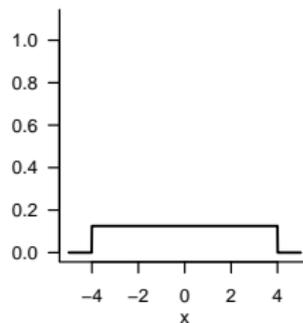
$U(x; 0,1)$



$U(x; -3,1)$



$U(x; -4,4)$



Konstruktion, Definition, Notation, Intuition

Wahrscheinlichkeitsmassefunktionen

Wahrscheinlichkeitsdichtefunktionen

Kumulative Verteilungsfunktionen

Selbstkontrollfragen

Definition (Kumulative Verteilungsfunktion)

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

$$P : \mathbb{R} \rightarrow [0, 1], x \mapsto P(x) := \mathbb{P}(\xi \leq x). \quad (23)$$

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

$$\mathbb{P}(\xi > x) = 1 - P(x) \text{ für alle } x \in \mathcal{X}. \quad (24)$$

Beweis

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

$$\Omega = \{\omega \in \Omega \mid \xi(\omega) > x\} \cup \{\omega \in \Omega \mid \xi(\omega) \leq x\} = \{\xi > x\} \cup \{\xi \leq x\}. \quad (25)$$

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

$$\begin{aligned} \mathbb{P}(\Omega) &= 1 \\ \Leftrightarrow \mathbb{P}(\{\xi > x\} \cup \{\xi \leq x\}) &= 1 \\ \Leftrightarrow \mathbb{P}(\{\xi > x\}) + \mathbb{P}(\{\xi \leq x\}) &= 1 \\ \Leftrightarrow \mathbb{P}(\{\xi > x\}) &= 1 - \mathbb{P}(\{\xi \leq x\}) \\ \Leftrightarrow \mathbb{P}(\{\xi > x\}) &= 1 - P(x). \end{aligned} \quad (26)$$

□

Theorem (Intervallwahrscheinlichkeiten)

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

$$\mathbb{P}(\xi \in]x_1, x_2]) = P(x_2) - P(x_1) \text{ für alle } x_1, x_2 \in \mathcal{X} \text{ mit } x_1 < x_2. \quad (27)$$

Beweis

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

$$\{\xi \leq x_1\} \cap \{x_1 < \xi \leq x_2\} = \emptyset \text{ und } \{\xi \leq x_1\} \cup \{x_1 < \xi \leq x_2\} = \{\xi \leq x_2\}. \quad (28)$$

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

$$\begin{aligned} \mathbb{P}(\{\xi \leq x_1\} \cup \{x_1 < \xi \leq x_2\}) &= \mathbb{P}(\{\xi \leq x_2\}) \\ \Leftrightarrow \mathbb{P}(\{\xi \leq x_1\}) + \mathbb{P}(\{x_1 < \xi \leq x_2\}) &= \mathbb{P}(\{\xi \leq x_2\}) \\ \Leftrightarrow \mathbb{P}(\{x_1 < \xi \leq x_2\}) &= \mathbb{P}(\{\xi \leq x_2\}) - \mathbb{P}(\{\xi \leq x_1\}) \\ \Leftrightarrow \mathbb{P}(\{x_1 < \xi \leq x_2\}) &= P(x_2) - P(x_1) \\ \Leftrightarrow \mathbb{P}(\xi \in]x_1, x_2]) &= P(x_2) - P(x_1). \end{aligned} \quad (29)$$

□

Theorem (Eigenschaften von kumulative Verteilungsfunktionen)

Es sei ξ 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.

Kumulative Verteilungsfunktionen

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 $x_1 < x_2$,

$$\{\xi \leq x_1\} = \{\omega \in \Omega \mid \xi(\omega) \leq x_1\} \subset \{\omega \in \Omega \mid \xi(\omega) \leq x_2\} = \{\xi \leq x_2\} \quad (30)$$

Also gilt

$$\mathbb{P}(\{\xi \leq x_1\}) \leq \mathbb{P}\{\xi \leq x_2\} \Rightarrow P(x_1) \leq P(x_2). \quad (31)$$

- (2) Wir verzichten auf einen Beweis

- (3) Wir definieren

$$P(x^+) = \lim_{y \rightarrow x, y > x} P(y). \quad (32)$$

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

$$\{\xi \leq x\} = \bigcap_{n=1}^{\infty} \{\xi \leq y_n\}. \quad (33)$$

Es gilt also

$$P(x) = \mathbb{P}(\{\xi \leq x\}) = \mathbb{P}(\bigcap_{n=1}^{\infty} \{\xi \leq y_n\}) = \lim_{n \rightarrow \infty} \mathbb{P}(\{\xi \leq y_n\}) = P(x^+), \quad (34)$$

wobei wir die dritte Gleichung unbegründet stehen lassen.

Kumulative Verteilungsfunktionen von diskreten Zufallsvariablen

- Wenn $a < b$ und $\mathbb{P}(a < \xi < b) = 0$, dann ist P konstant horizontal auf $]a, b[$.
- An jedem Punkt x mit $\mathbb{P}(\xi = x) > 0$ springt die KVF um den Betrag $\mathbb{P}(\xi = x)$.
- \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

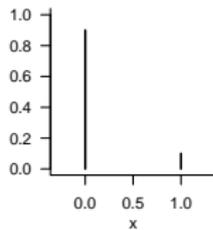
$$P : \mathbb{R} \rightarrow [0, 1], x \mapsto P(x) := \sum_{k=0}^{\lfloor x \rfloor} \mathbb{P}(\xi = k) \quad (35)$$

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

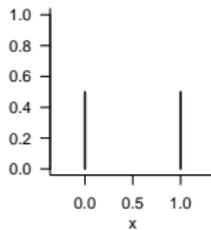
Kumulative Verteilungsfunktionen

Bernoulli-Zufallsvariablen

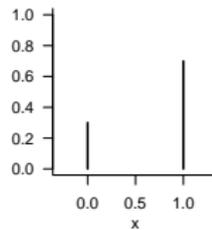
WMF von Bern($x;0.1$)



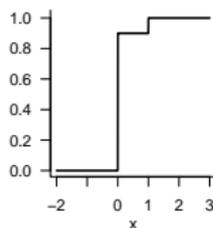
WMF von Bern($x;0.5$)



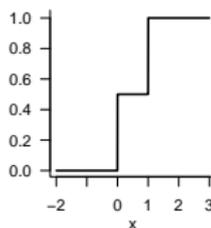
WMF von Bern($x;0.7$)



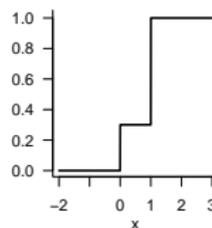
KVF von Bern($x;0.1$)



KVF von Bern($x;0.5$)



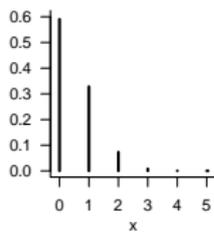
KVF von Bern($x;0.7$)



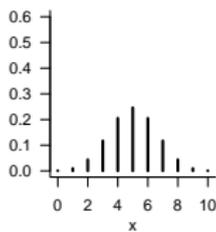
Kumulative Verteilungsfunktionen

Binomial-Zufallsvariablen

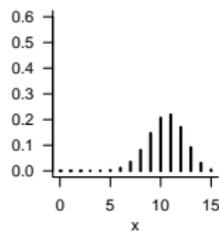
WMF von $\text{Bin}(x;0.1,5)$



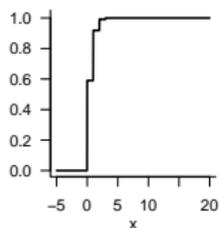
WMF von $\text{Bin}(x;0.5,10)$



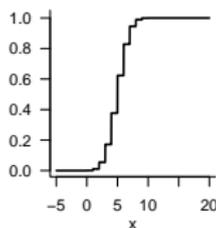
WMF von $\text{Bin}(x;0.7,15)$



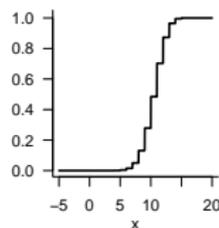
KVF von $\text{Bin}(x;5,0.1)$



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



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



Theorem (Kumulative Verteilungsfunktionen von kontinuierlichen ZVen)

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

$$P(x) = \int_{-\infty}^x p(s) ds \text{ und } p(x) = P'(x). \quad (36)$$

Beweis

Wir halten zunächst fest, dass weil $\mathbb{P}(\xi = x) = 0$ für alle $x \in \mathbb{R}$ gilt, die KVF von ξ 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.
- KVFFen von kontinuierlichen ZV heißen auch kumulative Dichtefunktionen (KDFen).

Beispiel (Normalverteilung)

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

- Die WDF von ξ 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 ξ ist

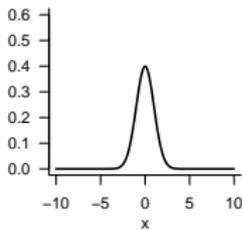
$$P : \mathbb{R} \rightarrow]0, 1[, x \mapsto P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x \exp\left(-\frac{1}{2\sigma^2}(s - \mu)^2\right) ds.$$

- Die KVF von $\xi \sim N(\mu, \sigma^2)$ 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.

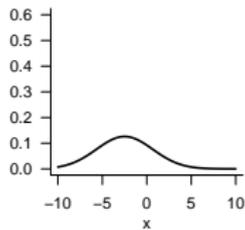
Kumulative Verteilungsfunktionen

Normalverteilte Zufallsvariablen

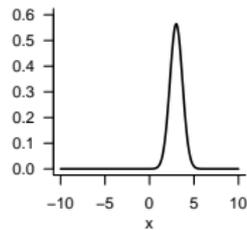
WDF von $N(0,1)$



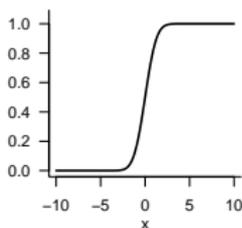
WDF von $N(-2.5,10)$



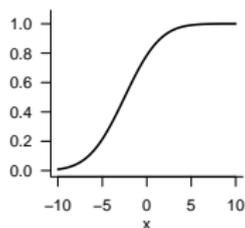
WDF von $N(3,0.5)$



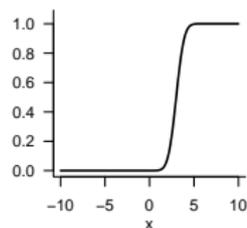
KDF von $N(0,1)$



KDF von $N(-2.5,10)$



KDF von $N(3,0.5)$



Definition (Inverse Kumulative Verteilungsfunktion)

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

$$P^{-1} :]0, 1[\rightarrow \mathbb{R}, q \mapsto P^{-1}(q) := \{x \in \mathbb{R} \mid P(x) = q\} \quad (37)$$

die *inverse kumulative Verteilungsfunktion* von ξ .

Bemerkungen

- P^{-1} ist die Inverse von P , d.h. $P^{-1}(P(x)) = x$.
- Offenbar gilt $P(x) = q \Leftrightarrow \mathbb{P}(\xi \leq x) = q$.
- Für $q \in]0, 1[$ ist also $P^{-1}(q)$ der Wert x von ξ , so dass $\mathbb{P}(\xi \leq x) = q$ 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 $\xi \sim N(\mu, \sigma^2)$.

- Die KVF von ξ ist

$$P : \mathbb{R} \rightarrow]0, 1[, x \mapsto \mathbb{P}(\xi \leq x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x \exp\left(-\frac{1}{2\sigma^2}(s - \mu)^2\right) ds \quad (38)$$

- Die inverse KVF von ξ ist

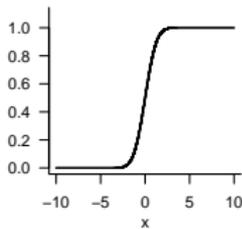
$$P^{-1} :]0, 1[\rightarrow \mathbb{R}, q \mapsto P^{-1}(q) = \{x \in \mathbb{R} | P(x) = q\}. \quad (39)$$

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

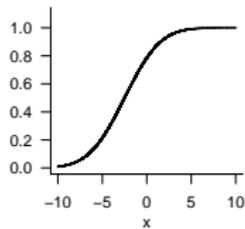
Kumulative Verteilungsfunktionen

Normalverteilte Zufallsvariablen

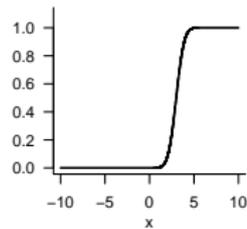
KDF von $N(0,1)$



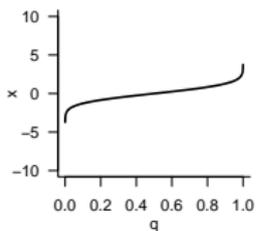
KDF von $N(-2.5,10)$



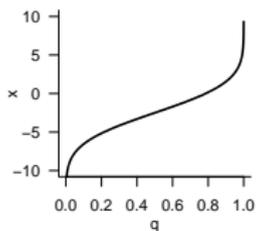
KDF von $N(3,0.5)$



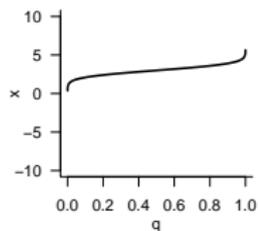
Inverse KDF von $N(0,1)$



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



Inverse KDF von $N(3,0.5)$



Konstruktion, Definition, Notation, Intuition

Wahrscheinlichkeitsmassfunktionen

Wahrscheinlichkeitsdichtefunktionen

Kumulative Verteilungsfunktionen

Selbstkontrollfragen

Konstruktion, Definition, Notation, Intuition

Wahrscheinlichkeitsmassfunktionen

Wahrscheinlichkeitsdichtefunktionen

Kumulative Verteilungsfunktionen

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die Definition des Begriffs der Zufallsvariable wieder.
2. Erläutern Sie die Gleichung $\mathbb{P}_\xi(\xi = x) = \mathbb{P}(\{\xi = x\})$.
3. Erläutern Sie die intuitive Bedeutung von $\mathbb{P}(\xi = x)$.
4. Geben Sie die Definition des Begriffs der Wahrscheinlichkeitsmassefunktion wieder.
5. Geben Sie die Definition des Begriffs der Wahrscheinlichkeitsdichtefunktion wieder.
6. Geben Sie die Definition der WDF einer normalverteilten Zufallsvariable wieder.
7. Geben Sie die Definition des Begriffs der kumulativen Verteilungsfunktion wieder.
8. Schreiben sie die Intervallwahrscheinlichkeit einer Zufallsvariable mithilfe ihrer KVF.
9. Schreiben Sie den Wert $P(x)$ der KVF einer kontinuierlichen Zufallsvariable mithilfe ihrer WDF p .
10. Schreiben Sie den Wert $p(x)$ der WDF einer kontinuierlichen Zufallsvariable mithilfe ihrer KVF P .
11. Geben Sie die Definition der KVF einer normalverteilten Zufallsvariable wieder.
12. Geben Sie die Definition des Begriffs der inversen Verteilungsfunktion wieder.

Selbstkontrollfragen - Lösungen

1. Siehe Definition (Zufallsvariable).
2. Die Gleichung besagt, dass der von der Verteilung \mathbb{P}_ξ der Zufallsvariable ξ zugeordnete Wahrscheinlichkeitswert für das Ereignis $\xi = x$ in \mathcal{X} definitionsgemäß gleich dem Wert des von dem Wahrscheinlichkeitsmaß \mathbb{P} dem Urbild des Ereignisses $\xi = x$, also der Menge $\{\xi = x\} = \{\omega \in \Omega \mid \xi(\omega) = x\}$ in \mathcal{A} , ist. Siehe Definition (Notation von Zufallsvariablen).
3. Intuitiv ist $\mathbb{P}(\xi = x)$ die Wahrscheinlichkeit dafür, dass die Zufallsvariable ξ den Wert x annimmt.
4. Siehe Definition (Diskrete Zufallsvariable, Wahrscheinlichkeitsmassefunktion).
5. Siehe Definition (Kontinuierliche Zufallsvariable, Wahrscheinlichkeitsdichtefunktion).
6. Siehe Definition (Normalverteilte und standardnormalverteilte Zufallsvariable).
7. Siehe Definition (Kumulative Verteilungsfunktion).
8. Siehe Theorem (Intervallwahrscheinlichkeiten).
9. Siehe Theorem (Kumulative Verteilungsfunktionen von kontinuierlichen Zufallsvariablen). Es gilt

$$P(x) = \int_{-\infty}^x p(s) ds. \quad (40)$$

10. Siehe Theorem (Kumulative Verteilungsfunktionen von kontinuierlichen Zufallsvariablen). Es gilt

$$p(x) = \frac{d}{dx} P(x) = P'(x). \quad (41)$$

11. Siehe Kumulative Verteilungsfunktionen, Beispiel (Normalverteilung). Für eine normalverteilte Zufallsvariable ξ mit Erwartungswertparameter μ und Varianzparameter $\sigma^2 > 0$ hat die KVF die Form

$$P : \mathbb{R} \rightarrow]0, 1[, x \mapsto P(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x \exp\left(-\frac{1}{2\sigma^2}(s - \mu)^2\right) ds. \quad (42)$$

12. Siehe Definition (Inverse Kumulative Verteilungsfunktion).

Hesse, Christian. 2009. *Wahrscheinlichkeitstheorie*. 2nd ed. Wiesbaden: Vieweg + Teubner.



Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2023/24

Prof. Dr. Dirk Ostwald

(4) Zufallsvektoren

Definition

Multivariate Verteilungen

Marginalverteilungen

Bedingte Verteilungen

Unabhängige Zufallsvariablen

Selbstkontrollfragen

Definition

Multivariate Verteilungen

Marginalverteilungen

Bedingte Verteilungen

Unabhängige Zufallsvariablen

Selbstkontrollfragen

Definition (Zufallsvektor)

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

$$\xi : \Omega \rightarrow \mathcal{X}, \omega \mapsto \xi(\omega) := \begin{pmatrix} \xi_1(\omega) \\ \vdots \\ \xi_n(\omega) \end{pmatrix} \quad (1)$$

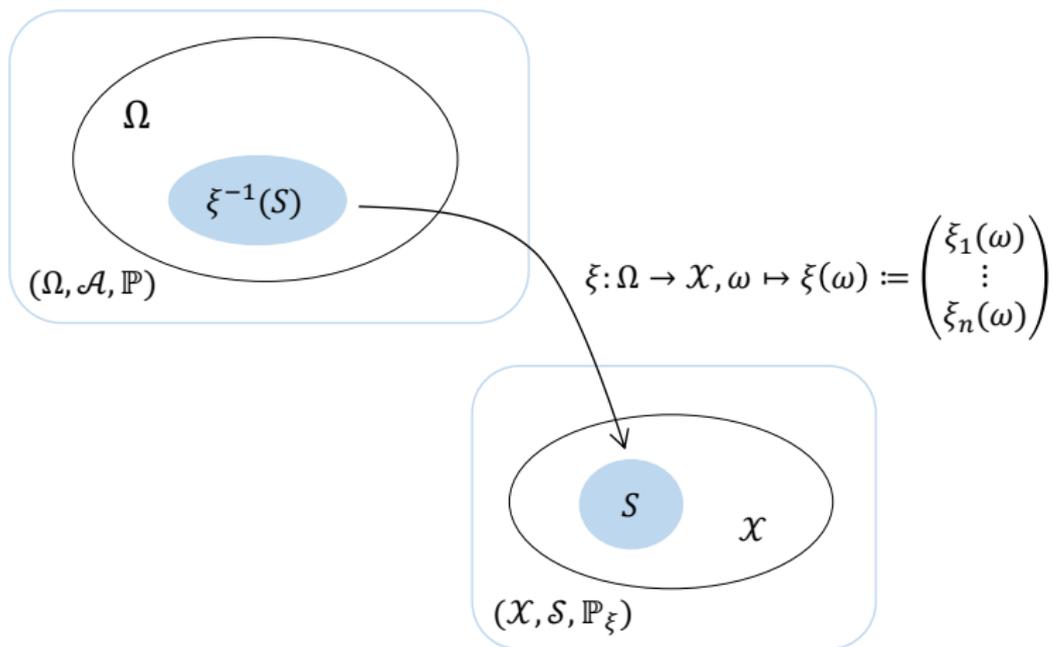
mit der *Messbarkeitseigenschaft*

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

Bemerkungen

- ξ ist hier eine univariate, vektorwertige Abbildung.
- Das Standardbeispiel für $(\mathcal{X}, \mathcal{S})$ ist $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.
- Wir verzichten auf eine explizite Einführung n -dimensionaler σ -Algebren wie $\mathcal{B}(\mathbb{R}^n)$.
- Ohne Beweis halten wir fest, dass ξ messbar ist, wenn die Funktionen ξ_1, \dots, ξ_n messbar sind.
- Die Komponentenfunktionen eines Zufallsvektors sind Zufallsvariablen.
- Ein n -dimensionaler Zufallsvektor ist die Konkatenation von n Zufallsvariablen.
- Für $n := 1$ ist ein Zufallsvektor eine Zufallsvariable.
- Für einen Zufallsvektor schreiben wir auch häufig $\xi := (\xi_1, \dots, \xi_n)$.

Definition



$$\mathbb{P}(\xi^{-1}(S)) = \mathbb{P}(\{\omega \in \Omega \mid \xi(\omega) \in S\}) =: \mathbb{P}_\xi(S)$$

Definition

Multivariate Verteilungen

Marginalverteilungen

Bedingte Verteilungen

Unabhängige Zufallsvariablen

Selbstkontrollfragen

Definition (Multivariate Verteilung)

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

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

sei ein Zufallsvektor. Dann heißt das Wahrscheinlichkeitsmaß \mathbb{P}_ξ , definiert durch

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

die *multivariate Verteilung des Zufallsvektor* ξ .

Bemerkungen

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

$$\begin{aligned} \mathbb{P}_\xi(\xi \in S) &:= \mathbb{P}(\{\xi \in S\}) = \mathbb{P}(\{\omega \in \Omega \mid \xi(\omega) \in S\}) \\ \mathbb{P}_\xi(\xi = x) &:= \mathbb{P}(\{\xi = x\}) = \mathbb{P}(\{\omega \in \Omega \mid \xi(\omega) = x\}) \\ \mathbb{P}_\xi(\xi \leq x) &:= \mathbb{P}(\{\xi \leq x\}) = \mathbb{P}(\{\omega \in \Omega \mid \xi(\omega) \leq x\}) \end{aligned} \quad (5)$$

$$\mathbb{P}_\xi(x_1 \leq \xi \leq x_2) := \mathbb{P}(\{x_1 \leq \xi \leq x_2\}) = \mathbb{P}(\{\omega \in \Omega \mid x_1 \leq \xi(\omega) \leq x_2\})$$

- Relationsoperatoren wie \leq werden hier *komponentenweise* verstanden.
- 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 (Diskreter Zufallsvektor, Multivariate WMF)

ξ sei ein Zufallsvektor mit Ergebnisraum \mathcal{X} . ξ heißt *diskreter Zufallsvektor* wenn der Ergebnisraum \mathcal{X} endlich oder abzählbar ist und eine Funktion

$$p_\xi : \mathcal{X} \rightarrow [0, 1], x \mapsto p_\xi(x) \quad (6)$$

existiert, für die gilt

(1) $\sum_{x \in \mathcal{X}} p_\xi(x) = 1$ und

(2) $\mathbb{P}_\xi(\xi = x) = p_\xi(x)$ für alle $x \in \mathcal{X}$.

Ein entsprechende Funktion p heißt *multivariate Wahrscheinlichkeitsmassenfunktion (WMF)* von ξ .

Bemerkungen

- Der Begriff der multivariaten WMF ist analog zum Begriff der WMF.
- Man spricht oft einfach von der WMF eines Zufallsvektors.
- Wie univariate WMFen sind multivariate WMFen nicht-negativ und normiert.

Beispiel (Multivariate Wahrscheinlichkeitsmassefunktion)

Wir betrachten einen zweidimensionalen Zufallsvektor $\xi := (\xi_1, \xi_2)$ 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. Dann entspricht der Ergebnisraum von ξ der in untenstehender Tabelle spezifizierten Menge an Tupeln (x_1, x_2) .

(x_1, x_2)	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$
$x_1 = 1$	(1, 1)	(1, 2)	(1, 3)	(1, 4)
$x_1 = 2$	(2, 1)	(2, 2)	(2, 3)	(2, 4)
$x_1 = 3$	(3, 1)	(3, 2)	(3, 3)	(3, 4)

Beispiel (Multivariate Wahrscheinlichkeitsmassefunktion)

Wir betrachten einen zweidimensionalen Zufallsvektor $\xi := (\xi_1, \xi_2)$ 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

$$p_\xi : \{1, 2, 3\} \times \{1, 2, 3, 4\} \rightarrow [0, 1], (x_1, x_2) \mapsto p_\xi(x_1, x_2) \quad (7)$$

ist dann durch nachfolgende Tabelle definiert:

$p_\xi(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$
$x_1 = 1$	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

$$\sum_{x_1=1}^3 \sum_{x_2=1}^4 p_\xi(x_1, x_2) = 1. \quad (8)$$

Definition (Kontinuierlicher Zufallsvektor, Multivariate WDF)

Ein Zufallsvektor ξ heißt *kontinuierlich*, wenn \mathbb{R}^n der Ergebnisraum von ξ ist und eine Funktion

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

existiert, für die gilt

$$(1) \int_{\mathbb{R}^n} p_\xi(x) dx = 1 \text{ und}$$

$$(2) \mathbb{P}_\xi(x_1 \leq \xi \leq x_2) = \int_{x_{1_1}}^{x_{2_1}} \dots \int_{x_{1_n}}^{x_{2_n}} p_\xi(s_1, \dots, s_n) ds_1 \dots ds_n.$$

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

Bemerkungen

- Der Begriff der multivariaten WDF ist analog zum Begriff der WDF.
- Man spricht häufig auch einfach von der WDF eines Zufallsvektors
- Wie univariate WDFen sind multivariate WDFen nicht-negativ und normiert.
- Wie für kontinuierliche Zufallsvariablen gilt für kontinuierliche Zufallsvektoren

$$\mathbb{P}_\xi(\xi = x) = \mathbb{P}_\xi(x \leq \xi \leq x) = \int_{x_1}^{x_1} \dots \int_{x_n}^{x_n} p_\xi(s_1, \dots, s_n) ds_1 \dots ds_n = 0 \quad (10)$$

- Das Standardbeispiel ist die multivariate Normalverteilung (SoSe 2024).

Definition

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, $\xi : \Omega \rightarrow \mathcal{X}$ sei ein Zufallsvektor, \mathbb{P}_ξ sei die Verteilung von ξ , $\mathcal{X}_i \subset \mathcal{X}$ sei der Ergebnisraum der i ten Komponente ξ_i von ξ , und \mathcal{S}_i sei eine σ -Algebra auf \mathcal{X}_i . Dann heißt die durch

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

definierte Verteilung die *ite univariate Marginalverteilung* von ξ .

Bemerkungen

- Univariate Marginalverteilungen sind die Verteilungen der Komponenten eines Zufallsvektors.
- Univariate Marginalverteilungen sind Verteilungen von Zufallsvariablen.
- Die Festlegung der multivariaten Verteilung von ξ legt auch die Verteilungen der ξ_i fest.

Theorem (Marginale Wahrscheinlichkeitsmasse- und -dichtefunktionen)

(1) $\xi = (\xi_1, \dots, \xi_n)$ sei ein n -dimensionaler diskreter Zufallsvektor mit Wahrscheinlichkeitsmassefunktion p_ξ und Komponentenergebnisräumen $\mathcal{X}_1, \dots, \mathcal{X}_n$. Dann ergibt sich die Wahrscheinlichkeitsmassefunktion der i ten Komponente ξ_i von ξ als

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

(2) $\xi = (\xi_1, \dots, \xi_n)$ sei ein n -dimensionaler kontinuierlicher Zufallsvektor mit Wahrscheinlichkeitsdichtefunktion p_ξ und Komponentenergebnisraum \mathbb{R} . Dann ergibt sich die Wahrscheinlichkeitsdichtefunktion der i ten Komponente ξ_i von ξ als

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

Bemerkungen

- Wir verzichten auf einen Beweis.
- Die WMFen univariater Marginalverteilungen diskreter Zufallsvektoren ergeben sich durch Summation.
- Die WDFen univariater Marginalverteilungen kontinuierlicher Zufallsvektoren ergeben sich durch Integration.

Marginalverteilungen

Beispiel (Marginale Wahrscheinlichkeitsmassenfunktionen)

Wir betrachten erneut den zweidimensionalen Zufallsvektor $\xi := (\xi_1, \xi_2)$ 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 marginale WMFen p_{ξ_1} und p_{ξ_2} :

$p_{\xi}(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$	$p_{\xi_1}(x_1)$
$x_1 = 1$	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
$p_{\xi_2}(x_2)$	0.2	0.3	0.3	0.2	

Man beachte, dass notwendigerweise gilt, dass

$$1 = \sum_{x_1=1}^3 \sum_{x_2=1}^4 p_{\xi}(x_1, x_2) = \sum_{x_1=1}^3 p_{\xi_1}(x_1) \quad \text{und} \quad 1 = \sum_{x_2=1}^4 \sum_{x_1=1}^3 p_{\xi}(x_1, x_2) = \sum_{x_2=1}^4 p_{\xi_2}(x_2) \quad (14)$$

dass aus der Normiertheit von p_{ξ} die Normiertheit von p_{ξ_1} und p_{ξ_2} also direkt folgt.

Beispiel (Marginale Wahrscheinlichkeitsmassenfunktionen)

Ein Realisierungsbeispiel mithilfe relativer Häufigkeiten mag den Begriff der marginalen WMF intuitiv verdeutlichen. Nehmen wir an, wir hätten $n = 100$ (unabhängige) Realisierungen von ξ vorliegen.

Um die Wahrscheinlichkeiten $p_\xi(x_1, x_2)$ zu schätzen, würden wir die Anzahl der Realisierungen von (x_1, x_2) zählen und durch n teilen. Hätten wir beispielsweise 12 Realisierungen von $(3, 2)$ vorliegen, so würden wir $p_\xi(3, 2) \approx 12/100 = 0.12$ schätzen.

Die Frage nach der marginalen Wahrscheinlichkeit von $x_2 = 2$ entspräche dann der Frage, wie oft unter den Realisierungen zu finden sind, bei denen $x_2 = 2$ ist, irrespektive des Wertes von x_1 . Dies wäre gerade die Anzahl der Realisierungen der Form $(1, 2)$, $(2, 2)$ und $(3, 2)$. Gäbe es von diesen beispielsweise 0, 22 und 12 respektive, so würde man die Wahrscheinlichkeit $p_{\xi_2}(2)$ natürlicherweise durch

$$\frac{0 + 22 + 12}{100} = \frac{0}{100} + \frac{22}{100} + \frac{12}{100} = 0.00 + 0.22 + 0.12 = 0.34 \quad (15)$$

schätzen. Anstelle der Wahrscheinlichkeiten $p_\xi(1, 2)$, $p_\xi(2, 2)$, $p_\xi(3, 2)$ addiert man hier also die entsprechenden relativen Häufigkeiten.

Definition

Multivariate Verteilungen

Marginalverteilungen

Bedingte Verteilungen

Unabhängige Zufallsvariablen

Selbstkontrollfragen

Bedingte Verteilungen

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

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (16)$$

Analog wird für zwei Zufallsvariablen ξ_1, ξ_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 ξ_1 gegeben ξ_2 mithilfe der Ereignisse

$$A := \{\xi_1 \in S_1\} \text{ und } B := \{\xi_2 \in S_2\} \quad (17)$$

definiert.

So ergibt sich zum Beispiel die bedingte Wahrscheinlichkeit, dass $\xi_1 \in S_1$ gegeben dass $\xi_2 \in S_2$ unter der Annahme, dass $\mathbb{P}(\{\xi_2 \in S_2\}) > 0$, zu

$$\mathbb{P}(\{\xi_1 \in S_1\}|\{\xi_2 \in S_2\}) = \frac{\mathbb{P}(\{\xi_1 \in S_1\} \cap \{\xi_2 \in S_2\})}{\mathbb{P}(\{\xi_2 \in S_2\})}. \quad (18)$$

Wir betrachten zunächst durch WMFen/WDFen zweidimensionaler Zufallsvektoren definierte bedingte Verteilungen.

Definition (Bedingte WMF, diskrete bedingte Verteilung)

$\xi := (\xi_1, \xi_2)$ sei ein diskreter Zufallsvektor mit Ergebnisraum $\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2$, WMF $p_\xi = p_{\xi_1, \xi_2}$ und marginalen WMFen p_{ξ_1} und p_{ξ_2} . Die bedingte WMF von ξ_1 gegeben $\xi_2 = x_2$ ist dann für $p_{\xi_2}(x_2) > 0$ definiert als

$$p_{\xi_1|\xi_2=x_2} : \mathcal{X}_1 \rightarrow [0, 1], x_1 \mapsto p_{\xi_1|\xi_2=x_2}(x_1|x_2) := \frac{p_{\xi_1, \xi_2}(x_1, x_2)}{p_{\xi_2}(x_2)} \quad (19)$$

Analog ist für $p_{\xi_1}(x_1) > 0$ die bedingte WMF von ξ_2 gegeben $\xi_1 = x_1$ definiert als

$$p_{\xi_2|\xi_1=x_1} : \mathcal{X}_2 \rightarrow [0, 1], x_2 \mapsto p_{\xi_2|\xi_1=x_1}(x_2|x_1) := \frac{p_{\xi_1, \xi_2}(x_1, x_2)}{p_{\xi_1}(x_1)} \quad (20)$$

Die bedingten Verteilungen mit WMFen $p_{\xi_1|\xi_2=x_2}$ und $p_{\xi_2|\xi_1=x_1}$ heißen dann die *diskreten bedingten Verteilungen* von ξ_1 gegeben $\xi_2 = x_2$ und ξ_2 gegeben $\xi_1 = x_1$, respektive.

Bemerkungen

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

$$p_{\xi_1|\xi_2}(x_1|x_2) = \frac{p_{\xi_1, \xi_2}(x_1, x_2)}{p_{\xi_2}(x_2)} = \frac{\mathbb{P}(\{\xi_1 = x_1\} \cap \{\xi_2 = x_2\})}{\mathbb{P}(\{\xi_2 = x_2\})}. \quad (21)$$

- Bedingte Verteilungen sind (lediglich) normalisierte gemeinsame Verteilungen.

Beispiel (Bedingte Wahrscheinlichkeitsmassfunktionen)

Wir betrachten erneut den zweidimensionalen Zufallsvektor $\xi := (\xi_1, \xi_2)$ 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 $p_{\xi_2|\xi_1=x_1}$

$p_{\xi_2 \xi_1}(x_2 x_1)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$
$p_{\xi_2 \xi_1=1}(x_2 x_1 = 1)$	$\frac{0.1}{0.4} = 0.25$	$\frac{0.0}{0.4} = 0.00$	$\frac{0.2}{0.4} = 0.50$	$\frac{0.1}{0.4} = 0.25$
$p_{\xi_2 \xi_1=2}(x_2 x_1 = 2)$	$\frac{0.1}{0.3} = 0.3\bar{3}$	$\frac{0.2}{0.3} = 0.6\bar{6}$	$\frac{0.0}{0.3} = 0.00$	$\frac{0.0}{0.3} = 0.00$
$p_{\xi_2 \xi_1=3}(x_2 x_1 = 3)$	$\frac{0.0}{0.3} = 0.00$	$\frac{0.1}{0.3} = 0.3\bar{3}$	$\frac{0.1}{0.3} = 0.3\bar{3}$	$\frac{0.1}{0.3} = 0.3\bar{3}$

Bemerkungen

- Man beachte, dass $\sum_{x_2=1}^4 p_{\xi_2|\xi_1=x_1}(x_2|x_1) = 1$ für alle $x_1 \in \mathcal{X}_1$.
- Man beachte die qualitative Ähnlichkeit der WMFen $p_{\xi_1, \xi_2}(x_1, x_2)$ und $p_{\xi_2|\xi_1}(x_2|x_1)$.
- Bedingte Verteilungen sind (lediglich) normalisierte gemeinsame Verteilungen.

Definition (Bedingte WDF, kontinuierliche bedingte Verteilungen)

$\xi := (\xi_1, \xi_2)$ sei ein kontinuierlicher Zufallsvektor mit Ergebnisraum \mathbb{R}^2 , WDF $p_\xi = p_{\xi_1, \xi_2}$ und marginalen WDFen p_{ξ_1} und p_{ξ_2} . Die bedingte WDF von ξ_1 gegeben $\xi_2 = x_2$ ist dann für $p_{\xi_2}(x_2) > 0$ definiert als

$$p_{\xi_1|\xi_2=x_2} : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, x_1 \mapsto p_{\xi_1|\xi_2=x_2}(x_1|x_2) := \frac{p_{\xi_1, \xi_2}(x_1, x_2)}{p_{\xi_2}(x_2)} \quad (22)$$

Analog ist für $p_{\xi_1}(x_1) > 0$ die bedingte WMF von ξ_2 gegeben $\xi_1 = x_1$ definiert als

$$p_{\xi_2|\xi_1=x_1} : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, x_2 \mapsto p_{\xi_2|\xi_1=x_1}(x_2|x_1) := \frac{p_{\xi_1, \xi_2}(x_1, x_2)}{p_{\xi_1}(x_1)} \quad (23)$$

Die Verteilungen mit WDFen $p_{\xi_1|\xi_2=x_2}$ und $p_{\xi_2|\xi_1=x_1}$ heißen dann die *kontinuierlichen bedingten Verteilungen* von ξ_1 gegeben $\xi_2 = x_2$ und ξ_2 gegeben $\xi_1 = x_1$, respektive.

Bemerkung

- Im kontinuierlichen Fall gilt zwar $\mathbb{P}(\xi = x) = 0$, aber nicht notwendig auch $p_\xi(x) = 0$.

Definition

Multivariate Verteilungen

Marginalverteilungen

Bedingte Verteilungen

Unabhängige Zufallsvariablen

Selbstkontrollfragen

Definition (Unabhängige Zufallsvariablen)

$(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum und $\xi := (\xi_1, \xi_2)$ ein zweidimensionaler Zufallsvektor. Die Zufallsvariablen ξ_1, ξ_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

$$\mathbb{P}_\xi(\xi_1 \in S_1, \xi_2 \in S_2) = \mathbb{P}_{\xi_1}(\xi_1 \in S_1)\mathbb{P}_{\xi_2}(\xi_2 \in S_2). \quad (24)$$

Bemerkungen

- Die Definition besagt, dass die Ereignisse $\{\xi_1 \in S_1\}$ und $\{\xi_2 \in S_2\}$ unabhängig sind.
- Es gilt also auch, dass $\mathbb{P}(\{\xi_1 \in S_1\}|\{\xi_2 \in S_2\}) = \mathbb{P}(\{\xi_1 \in S_1\})$.
- Wissen um das Ereignis $\{\xi_2 \in S_2\}$ verändert die Wahrscheinlichkeit von $\{\xi_1 \in S_1\}$ nicht.
- Einen formaleren Zugang bietet das Konzept der *Produktwahrscheinlichkeitsräume*.

Theorem (Unabhängigkeit und Faktorisierung der WMF/WDF)

(1) $\xi := (\xi_1, \xi_2)$ sei ein diskreter Zufallsvektor mit Ergebnisraum $\mathcal{X}_1 \times \mathcal{X}_2$, WMF p_ξ und marginalen WMFen p_{ξ_1}, p_{ξ_2} . Dann gilt

ξ_1 und ξ_2 sind unabhängige Zufallsvariablen \Leftrightarrow

$$p_\xi(x_1, x_2) = p_{\xi_1}(x_1)p_{\xi_2}(x_2) \text{ für alle } (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2. \quad (25)$$

(2) $\xi := (\xi_1, \xi_2)$ sei ein kontinuierlicher Zufallsvektor mit Ergebnisraum \mathbb{R}^2 , WDF p_ξ und marginalen WDFen p_{ξ_1}, p_{ξ_2} . Dann gilt

ξ_1 und ξ_2 sind unabhängige Zufallsvariablen \Leftrightarrow

$$p_\xi(x_1, x_2) = p_{\xi_1}(x_1)p_{\xi_2}(x_2) \text{ für alle } (x_1, x_2) \in \mathbb{R}^2. \quad (26)$$

Bemerkungen

- Wir verzichten auf einen Beweis.
- Die Produkteigenschaft $p_\xi(x_1, x_2) = p_{\xi_1}(x_1)p_{\xi_2}(x_2)$ heißt auch *Faktorisierung*.
- Unabhängigkeit zweier ZVen entspricht der Faktorisierung ihrer gemeinsamen WMF/WDF.

Unabhängige Zufallsvariablen

Beispiel (Unabhängige diskrete Zufallsvariablen)

Wir betrachten erneut den zweidimensionalen Zufallsvektor $\xi := (\xi_1, \xi_2)$, der Werte in $\{1, 2, 3\} \times \{1, 2, 3, 4\}$ annimmt, und dessen gemeinsame und marginale WMFen die untenstehende Form haben

$p_\xi(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$	$p_{\xi_1}(x_1)$
$x_1 = 1$	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
$p_{\xi_2}(x_2)$	0.20	0.30	0.30	0.20	

Da hier gilt, dass

$$p_\xi(1, 1) = 0.10 \neq 0.08 = 0.40 \cdot 0.20 = p_{\xi_1}(1)p_{\xi_2}(1) \quad (27)$$

sind die Zufallsvariablen ξ_1 und ξ_2 nicht unabhängig.

Unabhängige Zufallsvariablen

Beispiel (Unabhängige diskrete Zufallsvariablen)

Die gemeinsame Verteilung von ξ_1 und ξ_2 unter der Annahme der Unabhängigkeit von ξ_1 und ξ_2 bei gleichen Marginalverteilungen ergibt sich zu

$p_{\xi}(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$	$p_{\xi_1}(x_1)$
$x_1 = 1$	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
$p_{\xi_2}(x_2)$	0.20	0.30	0.30	0.20	

Weiterhin ergeben sich im Falle der Unabhängigkeit von ξ_1 und ξ_2 zum Beispiel die bedingten Wahrscheinlichkeitsmassefunktion $p_{\xi_2|\xi_1}$ zu

$p_{\xi_1 \xi_2}(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$
$p_{\xi_2 \xi_1=1}(x_2 x_1 = 1)$	$\frac{0.08}{0.40} = 0.2$	$\frac{0.12}{0.40} = 0.3$	$\frac{0.12}{0.40} = 0.3$	$\frac{0.08}{0.40} = 0.2$
$p_{\xi_2 \xi_1=2}(x_2 x_1 = 2)$	$\frac{0.06}{0.30} = 0.2$	$\frac{0.09}{0.30} = 0.3$	$\frac{0.09}{0.30} = 0.3$	$\frac{0.06}{0.30} = 0.2$
$p_{\xi_2 \xi_1=3}(x_2 x_1 = 3)$	$\frac{0.06}{0.30} = 0.2$	$\frac{0.09}{0.30} = 0.3$	$\frac{0.09}{0.30} = 0.3$	$\frac{0.06}{0.30} = 0.2$

Im Falle der Unabhängigkeit von ξ_1 und ξ_2 ändert sich die Verteilung von ξ_2 gegeben (oder im Wissen um) den Wert von ξ_1 also nicht und entspricht jeweils der Marginalverteilung von ξ_2 . Dies entspricht natürlich der Intuition der Unabhängigkeit von Ereignissen im Kontext elementarer Wahrscheinlichkeiten.

Definition (n unabhängige Zufallsvariablen)

$\xi := (\xi_1, \dots, \xi_n)$ sei ein n -dimensionaler Zufallsvektor mit Ergebnisraum $\mathcal{X} = \times_{i=1}^n \mathcal{X}_i$. Die n Zufallsvariablen ξ_1, \dots, ξ_n heißen *unabhängig*, wenn für alle $S_i \in \mathcal{X}_i, i = 1, \dots, n$ gilt, dass

$$\mathbb{P}_\xi(\xi_1 \in S_1, \dots, \xi_n \in S_n) = \prod_{i=1}^n \mathbb{P}_{\xi_i}(\xi_i \in S_i). \quad (28)$$

Wenn der Zufallsvektor eine n -dimensionale WMF oder WDF p_ξ mit marginalen WMFen oder WDFen $p_{\xi_i}, i = 1, \dots, n$ besitzt, dann ist die Unabhängigkeit von ξ_1, \dots, ξ_n gleichbedeutend mit der Faktorisierung der gemeinsamen WMF oder WDF, also mit

$$p_\xi(\xi_1, \dots, \xi_n) = \prod_{i=1}^n p_{\xi_i}(x_i). \quad (29)$$

Bemerkung

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

Definition (Unabhängig und identisch verteilte Zufallsvariablen)

n Zufallsvariablen ξ_1, \dots, ξ_n heißen *unabhängig und identisch verteilt (u.i.v.)*, wenn

- (1) ξ_1, \dots, ξ_n unabhängige Zufallsvariablen sind, und
- (2) die Marginalverteilungen der ξ_i übereinstimmen, also gilt, dass

$$\mathbb{P}_{\xi_i} = \mathbb{P}_{\xi_j} \text{ für alle } 1 \leq i, j \leq n. \quad (30)$$

Wenn die Zufallsvariablen ξ_1, \dots, ξ_n unabhängig und identisch verteilt sind und die *ite* Marginalverteilung $\mathbb{P}_\xi := \mathbb{P}_{\xi_i}$ ist, so schreibt man auch

$$\xi_1, \dots, \xi_n \sim \mathbb{P}_\xi. \quad (31)$$

Bemerkungen

- Man sagt kurz, dass ξ_1, \dots, ξ_n u.i.v. sind.
- Im Englischen spricht man von *independent and identically distributed (i.i.d)* Zufallsvariablen.
- In der Statistik werden Fehlerterme meist durch u.i.v. Zufallsvariablen modelliert.
- n u.i.v. normalverteilte ZVen werden als $\xi_1, \dots, \xi_n \sim N(\mu, \sigma^2)$ geschrieben.

Definition

Multivariate Verteilungen

Marginalverteilungen

Bedingte Verteilungen

Unabhängige Zufallsvariablen

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die Definition des Begriffs des Zufallsvektors wieder.
2. Geben Sie die Definition des Begriffs der multivariaten Verteilung eines Zufallsvektors wieder.
3. Geben Sie die Definition des Begriffs der multivariaten WMF wieder.
4. Geben Sie die Definition des Begriffs der multivariaten WDF wieder.
5. Geben Sie die Definition des Begriffs der univariaten Marginalverteilung eines Zufallsvektors wieder
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. Geben Sie die Definition des Begriffs der Unabhängigkeit zweier Zufallsvariablen wieder.
9. Wie erkennt man an der gemeinsamen WMF oder WDF eines zweidimensionalen Zufallsvektors, ob die Komponenten des Zufallsvektors unabhängig sind oder nicht?
10. Geben Sie die Definition des Begriffs der Unabhängigkeit von n Zufallsvariablen wieder.
11. Geben Sie die Definition des Begriffs der unabhängig und identisch verteilten Zufallsvariablen wieder.

Selbstkontrollfragen - Lösungen

1. Siehe Definition (Zufallsvektor).
2. Siehe Definition (Multivariate Verteilung).
3. Siehe Definition (Diskreter Zufallsvektor, Multivariate WMF).
4. Siehe Definition (Kontinuierlicher Zufallsvektor, Multivariate WDF).
5. Siehe Definition (Univariate Marginalverteilung).
6. Siehe Theorem (Marginale Wahrscheinlichkeitsmasse- und -dichtefunktionen). Man erhält die WMF der i ten Komponente eines diskreten Zufallsvektors durch Summation der Werte der WMF des Zufallsvektors über alle Komponenten des Zufallsvektors *außer* der i ten Komponente.
7. Siehe Theorem (Marginal Wahrscheinlichkeitsmasse- und dichtefunktionen). Man erhält die WDF der i ten Komponente eines kontinuierlichen Zufallsvektors durch Integration der Werte der WDF des Zufallsvektors über alle Komponenten des Zufallsvektors *außer* der i ten Komponente.
8. Siehe Definition (Unabhängige Zufallsvariablen).
9. Man prüft, ob sich alle Werte der gemeinsamen WMF oder WDF $p_{\xi}(x_1, x_2)$ des Zufallsvektors $\xi(\xi_1, \xi_2)$ durch Multiplikation der entsprechenden marginalen WMF oder WDF Werte $p_{\xi_1}(x_1)$ und $p_{\xi_2}(x_2)$ ergeben. Ist dies für alle Werte $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$ der Fall, so sind ξ_1 und ξ_2 unabhängig, ist dies nicht für alle Werte $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$, so sind ξ_1 und ξ_2 nicht unabhängig.
10. Siehe Definition (n unabhängige Zufallsvariablen).
11. Siehe Definition (Unabhängig und identisch verteilte Zufallsvariablen).



Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2023/24

Prof. Dr. Dirk Ostwald

(5) Erwartungswerte

Erwartungswert

Varianz und Standardabweichung

Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung

Kovarianz und Korrelation

Selbstkontrollfragen

Erwartungswert

Varianz und Standardabweichung

Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung

Kovarianz und Korrelation

Selbstkontrollfragen

Definition (Erwartungswert)

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

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

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

Bemerkungen

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

Beispiel (Erwartungswert einer Bernoulli Zufallsvariable)

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

Beweis

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

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

□

Erwartungswert

Beispiel (Erwartungswert einer normalverteilten Zufallsvariable)

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

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

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

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

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

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

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

□

Theorem (Einige Erwartungswerte und Varianzen)

Zufallsvariable	$\mathbb{E}(\xi)$
$\xi \sim B(\mu)$	μ
$\xi \sim \text{Bin}(\mu, n)$	$n\mu$
$\xi \sim N(\mu, \sigma^2)$	μ
$\xi \sim G(\alpha, \beta)$	$\alpha\beta$
$\xi \sim \text{Beta}(\alpha, \beta)$	$\frac{\alpha}{\alpha+\beta}$
$\xi \sim U(a, b)$	$\frac{a+b}{2}$

Bemerkung

- Wir verzichten auf einen Beweis.

Definition (Erwartungswert einer Funktion einer Zufallsvariable)

$(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum, ξ sei eine Zufallsvariable mit Ergebnisraum \mathcal{X} und $f : \mathcal{X} \rightarrow \mathcal{Z}$ sei eine Funktion mit Zielmenge \mathcal{Z} . Dann ist der *Erwartungswert der Funktion f der Zufallsvariable ξ* definiert als

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

Bemerkung

- Der Erwartungswert einer Zufallsvariable ist der Spezialfall mit

$$f : \mathcal{X} \rightarrow \mathcal{Z}, x \mapsto f(x) := x. \tag{9}$$

Definition (Erwartungswert eines Zufallsvektors)

ξ sei ein n -dimensionaler Zufallsvektor. Dann ist der *Erwartungswert* von ξ definiert als der n -dimensionale reelle Vektor

$$\mathbb{E}(\xi) := \begin{pmatrix} \mathbb{E}(\xi_1) \\ \vdots \\ \mathbb{E}(\xi_n) \end{pmatrix}. \quad (10)$$

Bemerkung

- Der Erwartungswert eines Zufallsvektors ist der Vektor der Erwartungswerte seiner Komponenten.

Definition (Erwartungswert einer Funktion eines Zufallsvektors)

$(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum, ξ sei ein Zufallsvektor mit Ergebnisraum \mathcal{X} und $f : \mathcal{X} \rightarrow \mathcal{Z}$ sei eine Funktion mit Zielmenge \mathcal{Z} . Dann ist der *Erwartungswert der Funktion f des Zufallsvektors ξ* definiert als

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

Bemerkung

- Die Definition ist analog zur Definition des Erwartungswerts einer Funktion einer Zufallsvariable.

Theorem (Eigenschaften des Erwartungswerts)

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

$$\mathbb{E}(a\xi + b) = a\mathbb{E}(\xi) + b. \quad (11)$$

(2) (Linearkombination) Für Zufallsvariablen ξ_1, \dots, ξ_n und $a_1, \dots, a_n \in \mathbb{R}$ gilt

$$\mathbb{E}\left(\sum_{i=1}^n a_i \xi_i\right) = \sum_{i=1}^n a_i \mathbb{E}(\xi_i). \quad (12)$$

(3) (Faktorisierung bei Unabhängigkeit) Für unabhängige Zufallsvariablen ξ_1, \dots, ξ_n gilt

$$\mathbb{E}\left(\prod_{i=1}^n \xi_i\right) = \prod_{i=1}^n \mathbb{E}(\xi_i). \quad (13)$$

Bemerkung

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

Beweis

Eigenschaft (1) folgt aus den Linearitätseigenschaften von Summen und Integralen. Wir betrachten nur den Fall einer kontinuierlichen Zufallsvariable ξ mit WDF p_ξ genauer und definieren zunächst $v := a\xi + b$. Dann gilt

$$\begin{aligned}\mathbb{E}(v) &= \mathbb{E}(a\xi + b) \\ &= \int_{-\infty}^{\infty} (ax + b)p_\xi(x) dx \\ &= \int_{-\infty}^{\infty} axp_\xi(x) + bp_\xi(x) dx \\ &= a \int_{-\infty}^{\infty} xp_\xi(x) dx + b \int_{-\infty}^{\infty} p_\xi(x) dx \\ &= a\mathbb{E}(\xi) + b.\end{aligned}\tag{14}$$

Erwartungswert

Beweis (fortgeführt)

Eigenschaft (2) folgt gleichfalls aus den Linearitätseigenschaften von Summen und Integralen. Wir betrachten nur den Fall von zwei kontinuierlichen Zufallsvariablen ξ_1 und ξ_2 mit bivariater WDF p_{ξ_1, ξ_2} genauer. In diesem Fall gilt

$$\begin{aligned}\mathbb{E}\left(\sum_{i=1}^2 a_i \xi_i\right) &= \mathbb{E}(a_1 \xi_1 + a_2 \xi_2) \\ &= \iint_{\mathbb{R}^2} (a_1 x_1 + a_2 x_2) p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2 \\ &= \iint_{\mathbb{R}^2} a_1 x_1 p_{\xi_1, \xi_2}(x_1, x_2) + a_2 x_2 p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2 \\ &= a_1 \iint_{\mathbb{R}^2} x_1 p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2 + a_2 \iint_{\mathbb{R}^2} x_2 p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2 \\ &= a_1 \int_{-\infty}^{\infty} x_1 \left(\int_{-\infty}^{\infty} p_{\xi_1, \xi_2}(x_1, x_2) dx_2 \right) dx_1 + a_2 \int_{-\infty}^{\infty} x_2 \left(\int_{-\infty}^{\infty} p_{\xi_1, \xi_2}(x_1, x_2) dx_1 \right) dx_2 \\ &= a_1 \int_{-\infty}^{\infty} x_1 p_{\xi_1}(x_1) dx_1 + a_2 \int_{-\infty}^{\infty} x_2 p_{\xi_2}(x_2) dx_2 \\ &= a_1 \mathbb{E}(\xi_1) + a_2 \mathbb{E}(\xi_2) \\ &= \sum_{i=1}^2 a_i \mathbb{E}(\xi_i).\end{aligned}\tag{15}$$

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_{ξ_1, \dots, ξ_n} . Weil als ξ_1, \dots, ξ_n unabhängig vorausgesetzt sind, gilt

$$p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \prod_{i=1}^n p_{\xi_i}(x_i). \quad (16)$$

Weiterhin gilt also

$$\begin{aligned} \mathbb{E} \left(\prod_{i=1}^n \xi_i \right) &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left(\prod_{i=1}^n x_i \right) p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{i=1}^n x_i \prod_{i=1}^n p_{\xi_i}(x_i) dx_1 \dots dx_n \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{i=1}^n x_i p_{\xi_i}(x_i) dx_1 \dots dx_n \\ &= \prod_{i=1}^n \int_{-\infty}^{\infty} x_i p_{\xi_i}(x_i) dx_i \\ &= \prod_{i=1}^n \mathbb{E}(\xi_i). \end{aligned} \quad (17)$$

□

Erwartungswert

Varianz und Standardabweichung

Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung

Kovarianz und Korrelation

Selbstkontrollfragen

Definition (Varianz und Standardabweichung)

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

$$\mathbb{V}(\xi) := \mathbb{E}((\xi - \mathbb{E}(\xi))^2), \quad (18)$$

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

$$\mathbb{S}(\xi) := \sqrt{\mathbb{V}(\xi)}. \quad (19)$$

Bemerkungen

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

Beispiel (Varianz einer Bernoulli Zufallsvariable)

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

$$\mathbb{V}(\xi) = \mu(1 - \mu). \quad (20)$$

Beweis

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

$$\begin{aligned} \mathbb{V}(\xi) &= \mathbb{E}((\xi - \mu)^2) \\ &= \sum_{x \in \{0,1\}} (x - \mu)^2 \text{Bern}(x; \mu) \\ &= (0 - \mu)^2 \mu^0 (1 - \mu)^{1-0} + (1 - \mu)^2 \mu^1 (1 - \mu)^{1-1} \\ &= \mu^2 (1 - \mu) + (1 - \mu)^2 \mu \\ &= (\mu^2 + (1 - \mu)\mu) (1 - \mu) \\ &= (\mu^2 + \mu - \mu^2) (1 - \mu) \\ &= \mu(1 - \mu). \end{aligned} \quad (21)$$

□

Theorem (Varianzverschiebungssatz)

Es sei ξ eine Zufallsvariable. Dann gilt

$$\mathbb{V}(\xi) = \mathbb{E}(\xi^2) - \mathbb{E}(\xi)^2. \quad (22)$$

Beweis

Mit der Definition der Varianz und der Linearität des Erwartungswerts gilt

$$\begin{aligned} \mathbb{V}(\xi) &= \mathbb{E}((\xi - \mathbb{E}(\xi))^2) \\ &= \mathbb{E}(\xi^2 - 2\xi\mathbb{E}(\xi) + \mathbb{E}(\xi)^2) \\ &= \mathbb{E}(\xi^2) - 2\mathbb{E}(\xi)\mathbb{E}(\xi) + \mathbb{E}(\mathbb{E}(\xi)^2) \\ &= \mathbb{E}(\xi^2) - 2\mathbb{E}(\xi)^2 + \mathbb{E}(\xi)^2 \\ &= \mathbb{E}(\xi^2) - \mathbb{E}(\xi)^2. \end{aligned} \quad (23)$$

□

Bemerkung

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

Beispiel (Varianz einer normalverteilten Zufallsvariable)

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

Beweis

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

$$\mathbb{V}(\xi) = \mathbb{E}(\xi^2) - \mathbb{E}(\xi)^2 = \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} x^2 \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx - \mu^2 \quad (24)$$

Mit der Substitutionsregel

$$\int_a^b f(g(x))g'(x) dx = \int_{g(a)}^{g(b)} f(x) dx \quad (25)$$

und der Definition von

$$g: \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \sqrt{2\sigma^2}x + \mu, g(-\infty) := -\infty, g(\infty) := \infty, \text{ with } g'(x) = \sqrt{2\sigma^2}, \quad (26)$$

kann das Integral auf der rechten Seite von Gleichung (24) dann als

Beweis (fortgeführt)

$$\begin{aligned} & \int_{-\infty}^{\infty} x^2 \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx \\ &= \int_{-\infty}^{\infty} (\sqrt{2\sigma^2}x + \mu)^2 \exp\left(-\frac{1}{2\sigma^2}((\sqrt{2\sigma^2}x + \mu) - \mu)^2\right) \sqrt{2\sigma^2} dx \\ &= \sqrt{2\sigma^2} \int_{-\infty}^{\infty} (\sqrt{2\sigma^2}x + \mu)^2 \exp\left(-\frac{2\sigma^2 x^2}{2\sigma^2}\right) dx \\ &= \sqrt{2\sigma^2} \int_{-\infty}^{\infty} (\sqrt{2\sigma^2}x + \mu)^2 \exp(-x^2) dx. \end{aligned} \tag{27}$$

geschrieben werden. Also gilt

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

Varianz und Standardabweichung

Beweis (fortgeführt)

Wir halten weiterhin ohne Beweis fest, dass

$$\int_{-\infty}^{\infty} x \exp(-x^2) dx = 0 \text{ und } \int_{-\infty}^{\infty} \exp(-x^2) dx = \sqrt{\pi}. \quad (28)$$

Es ergibt sich dann

$$\begin{aligned} \mathbb{V}(\xi) &= \frac{1}{\sqrt{\pi}} \left(2\sigma^2 \int_{-\infty}^{\infty} x^2 \exp(-x^2) dx + \mu^2 \sqrt{\pi} \right) - \mu^2 \\ &= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} x^2 \exp(-x^2) dx + \mu^2 - \mu^2 \\ &= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} x^2 \exp(-x^2) dx \end{aligned} \quad (29)$$

Mit der partiellen Integrationsregel

$$\int_a^b f'(x)g(x) dx = f(x)g(x)|_a^b - \int_a^b f(x)g'(x) dx \quad (30)$$

und der Definition von

$$f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto f(x) := \exp(-x^2) \text{ with } f'(x) = -2 \exp(-x^2) \quad (31)$$

Varianz und Standardabweichung

Beweis (fortgeführt)

und

$$g : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto g(x) := -\frac{1}{2}x \text{ with } g'(x) = -\frac{1}{2}, \quad (32)$$

so dass

$$f'(x)g(x) = -2 \exp(-x^2) \left(-\frac{1}{2}x\right) = x^2 \exp(-x^2), \quad (33)$$

gilt, ergibt sich dann

$$\begin{aligned} \mathbb{V}(\xi) &= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} x^2 \exp(-x^2) dx \\ &= \frac{2\sigma^2}{\sqrt{\pi}} \left(-\frac{1}{2}x \exp(-x^2) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \exp(-x^2) \left(-\frac{1}{2}\right) dx \right) \\ &= \frac{2\sigma^2}{\sqrt{\pi}} \left(-\frac{1}{2}x \exp(-x^2) \Big|_{-\infty}^{\infty} + \frac{1}{2} \int_{-\infty}^{\infty} \exp(-x^2) dx \right), \end{aligned} \quad (34)$$

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

$$\mathbb{V}(\xi) = \frac{2\sigma^2}{\sqrt{\pi}} \left(\frac{1}{2} \int_{-\infty}^{\infty} \exp(-x^2) dx \right) = \frac{\sigma^2}{\sqrt{\pi}} \sqrt{\pi} = \sigma^2. \quad (35)$$

□

Theorem (Einige Erwartungswerte und Varianzen)

Zufallsvariable	$\mathbb{E}(\xi)$	$\mathbb{V}(\xi)$
$\xi \sim B(\mu)$	μ	$\mu(1 - \mu)$
$\xi \sim \text{Bin}(\mu, n)$	$n\mu$	$n\mu(1 - \mu)$
$\xi \sim N(\mu, \sigma^2)$	μ	σ^2
$\xi \sim G(\alpha, \beta)$	$\alpha\beta$	$\alpha\beta^2$
$\xi \sim \text{Beta}(\alpha, \beta)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
$\xi \sim U(a, b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$

Bemerkung

- Wir verzichten auf einen Beweis.

Theorem (Eigenschaften der Varianz)

(1) (Nichtnegativität) ξ sei eine Zufallsvariable. Dann gilt

$$\mathbb{V}(\xi) \geq 0. \quad (36)$$

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

$$\mathbb{V}(a\xi + b) = a^2\mathbb{V}(\xi) \text{ und } \mathbb{S}(a\xi + b) = |a|\mathbb{S}(\xi). \quad (37)$$

(3) (Linearkombination bei Unabhängigkeit) ξ_1, \dots, ξ_n seien unabhängige Zufallsvariable und es seien $a_1, \dots, a_n \in \mathbb{R}$. Dann gilt

$$\mathbb{V}\left(\sum_{i=1}^n a_i \xi_i\right) = \sum_{i=1}^n a_i^2 \mathbb{V}(\xi_i). \quad (38)$$

Beweis

Hinsichtlich (1) betrachten wir den Fall einer kontinuierlichen Zufallsvariable. Dann gilt zunächst

$$(x - \mathbb{E}(\xi))^2 \geq 0 \text{ für alle } x \in \mathbb{R}. \quad (39)$$

Weiterhin gilt für die WDF p von ξ , dass

$$p(x) \geq 0 \text{ für alle } x \in \mathbb{R}. \quad (40)$$

Also folgt

$$p(x)(x - \mathbb{E}(\xi))^2 \geq 0 \text{ für alle } x \in \mathbb{R}. \quad (41)$$

Damit gilt dann aber

$$\mathbb{V}(\xi) = \int_{-\infty}^{\infty} p(x)(x - \mathbb{E}(\xi))^2 dx \geq 0, \quad (42)$$

denn das Riemann Integral jeder nicht-negativen Funktion ist nicht-negativ. Analog zeigt man die Nichtnegativität der Varianz bei diskreten Zufallsvariablen.

Beweis (fortgeführt)

Um Eigenschaft (2) zu zeigen, definieren wir zunächst $v := a\xi + b$ und halten fest, dass $\mathbb{E}(v) = a\mathbb{E}(\xi) + b$. Für die Varianz von v ergibt sich dann

$$\begin{aligned}\mathbb{V}(v) &= \mathbb{E}((v - \mathbb{E}(v))^2) \\ &= \mathbb{E}((a\xi + b - a\mathbb{E}(\xi) - b)^2) \\ &= \mathbb{E}((a\xi - a\mathbb{E}(\xi))^2) \\ &= \mathbb{E}((a(\xi - \mathbb{E}(\xi)))^2) \\ &= \mathbb{E}(a^2(\xi - \mathbb{E}(\xi))^2) \\ &= a^2\mathbb{E}((\xi - \mathbb{E}(\xi))^2) \\ &= a^2\mathbb{V}(\xi)\end{aligned}\tag{43}$$

Wurzelziehen ergibt dann das Resultat für die Standardabweichung.

Varianz und Standardabweichung

Beweis (fortgeführt)

Für Eigenschaft (3) betrachten wir den Fall zweier unabhängiger Zufallsvariablen ξ_1 und ξ_2 genauer. Wir halten zunächst fest, dass in diesem Fall gilt, dass

$$\mathbb{E}(a_1\xi_1 + a_2\xi_2) = a_1\mathbb{E}(\xi_1) + a_2\mathbb{E}(\xi_2). \quad (44)$$

Es ergibt sich also

$$\begin{aligned} \mathbb{V}\left(\sum_{i=1}^2 a_i \xi_i\right) &= \mathbb{V}(a_1\xi_1 + a_2\xi_2) \\ &= \mathbb{E}\left((a_1\xi_1 + a_2\xi_2 - \mathbb{E}(a_1\xi_1 + a_2\xi_2))^2\right) \\ &= \mathbb{E}\left((a_1\xi_1 + a_2\xi_2 - a_1\mathbb{E}(\xi_1) - a_2\mathbb{E}(\xi_2))^2\right) \\ &= \mathbb{E}\left((a_1\xi_1 - a_1\mathbb{E}(\xi_1) + a_2\xi_2 - a_2\mathbb{E}(\xi_2))^2\right) \\ &= \mathbb{E}\left(((a_1(\xi_1 - \mathbb{E}(\xi_1))) + (a_2(\xi_2 - \mathbb{E}(\xi_2))))^2\right) \\ &= \mathbb{E}\left((a_1(\xi_1 - \mathbb{E}(\xi_1)))^2 + 2(a_1(\xi_1 - \mathbb{E}(\xi_1)))(a_2(\xi_2 - \mathbb{E}(\xi_2))) + (a_2(\xi_2 - \mathbb{E}(\xi_2)))^2\right) \\ &= \mathbb{E}\left((a_1^2(\xi_1 - \mathbb{E}(\xi_1))^2 + 2a_1a_2(\xi_1 - \mathbb{E}(\xi_1))(\xi_2 - \mathbb{E}(\xi_2))) + a_2^2(\xi_2 - \mathbb{E}(\xi_2))^2\right) \\ &= a_1^2\mathbb{E}\left((\xi_1 - \mathbb{E}(\xi_1))^2\right) + 2a_1a_2\mathbb{E}\left((\xi_1 - \mathbb{E}(\xi_1))(\xi_2 - \mathbb{E}(\xi_2))\right) + a_2^2\mathbb{E}\left((\xi_2 - \mathbb{E}(\xi_2))^2\right) \\ &= a_1^2\mathbb{V}(\xi_1) + 2a_1a_2\mathbb{E}\left((\xi_1 - \mathbb{E}(\xi_1))(\xi_2 - \mathbb{E}(\xi_2))\right) + a_2^2\mathbb{V}(\xi_2) \\ &= \sum_{i=1}^2 a_i^2\mathbb{V}(\xi_i) + 2a_1a_2\mathbb{E}\left((\xi_1 - \mathbb{E}(\xi_1))(\xi_2 - \mathbb{E}(\xi_2))\right). \end{aligned} \quad (45)$$

Beweis (fortgeführt)

Weil ξ_1 und ξ_2 unabhängig sind, ergibt sich mit den Eigenschaften des Erwartungswerts für unabhängige Zufallsvariablen, dass

$$\begin{aligned}\mathbb{E}((\xi_1 - \mathbb{E}(\xi_1))(\xi_2 - \mathbb{E}(\xi_2))) &= \mathbb{E}((\xi_1 - \mathbb{E}(\xi_1))) \mathbb{E}((\xi_2 - \mathbb{E}(\xi_2))) \\ &= (\mathbb{E}(\xi_1) - \mathbb{E}(\xi_1))(\mathbb{E}(\xi_2) - \mathbb{E}(\xi_2)) \\ &= 0\end{aligned}\tag{46}$$

ist. Damit folgt also

$$\mathbb{V}\left(\sum_{i=1}^2 a_i \xi_i\right) = \sum_{i=1}^2 a_i^2 \mathbb{V}(\xi_i).\tag{47}$$

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

□

Erwartungswert

Varianz und Standardabweichung

Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung

Kovarianz und Korrelation

Selbstkontrollfragen

Definition (Stichprobenmittel, -varianz, -standardabweichung)

ξ_1, \dots, ξ_n seien Zufallsvariablen. Dann nennt man ξ_1, \dots, ξ_n auch eine *Stichprobe*.

- Das *Stichprobenmittel* von ξ_1, \dots, ξ_n ist definiert als der arithmetische Mittelwert

$$\bar{\xi} := \frac{1}{n} \sum_{i=1}^n \xi_i. \quad (48)$$

- Die *Stichprobenvarianz* von ξ_1, \dots, ξ_n ist definiert als

$$S^2 := \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \bar{\xi})^2. \quad (49)$$

- Die *Stichprobenstandardabweichung* ist definiert als

$$S := \sqrt{S^2}. \quad (50)$$

Bemerkungen

- $\mathbb{E}(\xi)$, $\mathbb{V}(\xi)$, und $\mathbb{S}(\xi)$ sind Kennzahlen einer Zufallsvariable ξ .
- $\bar{\xi}$, S^2 , und S sind Kennzahlen einer Stichprobe ξ_1, \dots, ξ_n .
- $\bar{\xi}$, S^2 , und S sind Zufallsvariablen, ihre Realisationen werden im Folgenden mit \bar{x} , s^2 und s bezeichnet.

Beispiel (Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung)

- Es seien $\xi_1, \dots, \xi_{10} \sim N(1, 2)$.
- 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

$$\bar{x} = \frac{1}{10} \sum_{i=1}^{10} x_i = \frac{6.88}{10} = 0.68. \quad (51)$$

- Die Stichprobenvarianzrealisation ist

$$s^2 = \frac{1}{9} \sum_{i=1}^{10} (x_i - \bar{x})^2 = \frac{1}{9} \sum_{i=1}^{10} (x_i - 0.68)^2 = \frac{25.37}{9} = 2.82. \quad (52)$$

- Die Stichprobenstandardabweichungrealisation ist

$$s = \sqrt{s^2} = \sqrt{2.82} = 1.68. \quad (53)$$

Zur Prävention von Verwechslungen

- Der *Erwartungswert* $\mathbb{E}(\xi)$ und die *Varianz* $\mathbb{V}(\xi)$ sind Kennzahlen von Zufallsvariablen und werden basierend auf der Verteilung (WMF/WDF) einer Zufallsvariable bestimmt.
- Das *Stichprobenmittel* $\bar{\xi}$ und die *Stichprobenvarianz* S^2 sind Kennzahlen von Stichproben und werden basierend auf Stichproben, insbesondere ihrer Realisierungen, berechnet.
- Der *Erwartungswertparameter* μ und der *Varianzparameter* σ^2 sind Parameter der Normalverteilung und werden im Rahmen der probabilistischen Inferenz definiert oder geschätzt.

Erwartungswert

Varianz und Standardabweichung

Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung

Kovarianz und Korrelation

Selbstkontrollfragen

Definition (Kovarianz und Korrelation)

Die *Kovarianz* zweier Zufallsvariablen ξ und v ist definiert als

$$\mathbb{C}(\xi, v) := \mathbb{E}((\xi - \mathbb{E}(\xi))(v - \mathbb{E}(v))). \quad (54)$$

Die *Korrelation* zweier Zufallsvariablen ξ und v ist definiert als

$$\rho(\xi, v) := \frac{\mathbb{C}(\xi, v)}{\sqrt{\mathbb{V}(\xi)}\sqrt{\mathbb{V}(v)}} = \frac{\mathbb{C}(\xi, v)}{\mathbb{S}(\xi)\mathbb{S}(v)}. \quad (55)$$

Bemerkungen

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

$$\mathbb{C}(\xi, \xi) = \mathbb{E}((\xi - \mathbb{E}(\xi))^2) = \mathbb{V}(\xi). \quad (56)$$

- $\rho(\xi, v)$ wird auch *Korrelationskoeffizient* von ξ und v genannt.
- Wenn $\rho(\xi, v) = 0$ ist, werden ξ und v *unkorreliert* genannt.
- Wir zeigen später mit der Cauchy-Schwarz Ungleichung, dass $-1 \leq \rho(\xi, v) \leq 1$.
- Die Begriffe der Stichprobenkovarianz und der Stichprobenkorrelation führen wir erst an späterer Stelle ein.

Beispiel (Kovarianz und Korrelation zweier diskreter Zufallsvariablen)

Es sei $\zeta := (\xi, \nu)$ ein Zufallsvektor mit WMF $p_{\xi, \nu}$ definiert durch

$p_{\xi, \nu}(x, y)$	$y = 1$	$y = 2$	$y = 3$	$p_{\xi}(x)$
$x = 1$	0.10	0.05	0.15	0.30
$x = 2$	0.60	0.05	0.05	0.70
$p_{\nu}(y)$	0.70	0.10	0.20	

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

$$\mathbb{E}(\xi) = \sum_{x=1}^2 xp_{\xi}(x) = 1 \cdot 0.3 + 2 \cdot 0.7 = 1.7 \quad (57)$$

und

$$\mathbb{E}(\nu) = \sum_{y=1}^3 yp_{\nu}(y) = 1 \cdot 0.7 + 2 \cdot 0.1 + 3 \cdot 0.2 = 1.5. \quad (58)$$

Mit der Definition der Kovarianz von ξ und ν , gilt dann

$$\begin{aligned}C(\xi, v) &= \mathbb{E}((\xi - \mathbb{E}(\xi))(v - \mathbb{E}(v))) \\&= \sum_{x=1}^2 \sum_{y=1}^3 (x - \mathbb{E}(\xi))(y - \mathbb{E}(v))p_{\xi, v}(x, y) \\&= \sum_{x=1}^2 \sum_{y=1}^3 (x - 1.7)(y - 1.5)p_{\xi, v}(x, y) \\&= \sum_{x=1}^2 (x - 1.7)(1 - 1.5)p_{\xi, v}(x, 1) \\&\quad + (x - 1.7)(2 - 1.5)p_{\xi, v}(x, 2) \\&\quad + (x - 1.7)(3 - 1.5)p_{\xi, v}(x, 3) \\&= (1 - 1.7)(1 - 1.5)p_{\xi, v}(1, 1) + (1 - 1.7)(2 - 1.5)p_{\xi, v}(1, 2) + (1 - 1.7)(3 - 1.5)p_{\xi, v}(1, 3) \\&\quad + (2 - 1.7)(1 - 1.5)p_{\xi, v}(2, 1) + (2 - 1.7)(2 - 1.5)p_{\xi, v}(2, 2) + (2 - 1.7)(3 - 1.5)p_{\xi, v}(2, 3) \\&= (-0.7) \cdot (-0.5) \cdot 0.10 + (-0.7) \cdot 0.5 \cdot 0.05 + (-0.7) \cdot 1.5 \cdot 0.15 \\&\quad + 0.3 \cdot (-0.5) \cdot 0.60 + 0.3 \cdot 0.5 \cdot 0.05 + 0.3 \cdot 1.5 \cdot 0.05 \\&= 0.035 - 0.0175 - 0.1575 - 0.09 + 0.0075 + 0.0225 \\&= -0.2.\end{aligned} \tag{59}$$

Theorem (Symmetrie von Kovarianz und Korrelation)

ξ und v seien zwei Zufallsvariablen. Dann gelten

$$\mathbb{C}(\xi, v) = \mathbb{C}(v, \xi) \text{ und } \rho(\xi, v) = \rho(v, \xi) \quad (60)$$

Beweis

Mit der Kommutativität der Multiplikation gelten

$$\mathbb{C}(\xi, v) = \mathbb{E}((\xi - \mathbb{E}(\xi))(v - \mathbb{E}(v))) = \mathbb{E}((v - \mathbb{E}(v))(\xi - \mathbb{E}(\xi))) \mathbb{C}(v, \xi) \quad (61)$$

und

$$\rho(\xi, v) = \frac{\mathbb{C}(\xi, v)}{\mathfrak{S}(\xi)v} = \frac{\mathbb{C}(v, \xi)}{\mathfrak{S}(v)\xi} = \rho(v, \xi) \quad (62)$$

Bemerkungen

- Man sagt auch, dass Kovarianz und Korrelation "nicht gerichtet" sind.

Theorem (Kovarianzverschiebungssatz)

ξ und v seien Zufallsvariablen. Dann gilt

$$\mathbb{C}(\xi, v) = \mathbb{E}(\xi v) - \mathbb{E}(\xi)\mathbb{E}(v). \quad (63)$$

Beweis

Mit der Definition der Kovarianz gilt

$$\begin{aligned} \mathbb{C}(\xi, v) &= \mathbb{E}((\xi - \mathbb{E}(\xi))(v - \mathbb{E}(v))) \\ &= \mathbb{E}(\xi v - \xi \mathbb{E}(v) - \mathbb{E}(\xi)v + \mathbb{E}(\xi)\mathbb{E}(v)) \\ &= \mathbb{E}(\xi v) - \mathbb{E}(\xi)\mathbb{E}(v) - \mathbb{E}(\xi)\mathbb{E}(v) + \mathbb{E}(\xi)\mathbb{E}(v) \\ &= \mathbb{E}(\xi v) - \mathbb{E}(\xi)\mathbb{E}(v). \end{aligned} \quad (64)$$

□

Bemerkungen

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

Theorem (Varianzen von Summen und Differenzen von Zufallsvariablen)

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

$$\mathbb{V}(a\xi + bv + c) = a^2\mathbb{V}(\xi) + b^2\mathbb{V}(v) + 2ab\mathbb{C}(\xi, v). \quad (65)$$

Speziell gelten

$$\mathbb{V}(\xi + v) = \mathbb{V}(\xi) + \mathbb{V}(v) + 2\mathbb{C}(\xi, v) \quad (66)$$

und

$$\mathbb{V}(\xi - v) = \mathbb{V}(\xi) + \mathbb{V}(v) - 2\mathbb{C}(\xi, v) \quad (67)$$

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

$$\mathbb{E}(a\xi + bv + c) = a\mathbb{E}(\xi) + b\mathbb{E}(v) + c. \quad (68)$$

Es ergibt sich also

$$\begin{aligned} \mathbb{V}(a\xi + bv + c) &= \mathbb{E} \left((a\xi + bv + c - a\mathbb{E}(\xi) - b\mathbb{E}(v) - c)^2 \right) \\ &= \mathbb{E} \left((a(\xi - \mathbb{E}(\xi)) + b(v - \mathbb{E}(v)))^2 \right) \\ &= \mathbb{E} \left(a^2(\xi - \mathbb{E}(\xi))^2 + 2ab(\xi - \mathbb{E}(\xi))(v - \mathbb{E}(v)) + b^2(v - \mathbb{E}(v))^2 \right) \\ &= a^2\mathbb{E} \left((\xi - \mathbb{E}(\xi))^2 \right) + b^2\mathbb{E} \left((v - \mathbb{E}(v))^2 \right) + 2ab\mathbb{E} \left((\xi - \mathbb{E}(\xi))(v - \mathbb{E}(v)) \right) \\ &= a^2\mathbb{V}(\xi) + b^2\mathbb{V}(v) + 2ab\mathbb{C}(\xi, v) \end{aligned} \quad (69)$$

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

□

Theorem (Korrelation und Unabhängigkeit)

ξ und v seien zwei Zufallsvariablen. Dann gelten:

- (1) Wenn ξ und v unabhängig sind, dann ist $\mathbb{C}(\xi, v) = 0$ und ξ und v sind unkorreliert.
- (2) Wenn $\mathbb{C}(\xi, v) = 0$ ist, also ξ und v unkorreliert sind, dann sind ξ und v nicht notwendigerweise unabhängig.

Beweis

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

$$\mathbb{E}(\xi v) = \mathbb{E}(\xi)\mathbb{E}(v). \quad (70)$$

Mit dem Kovarianzverschiebungssatz folgt dann

$$\mathbb{C}(\xi, v) = \mathbb{E}(\xi v) - \mathbb{E}(\xi)\mathbb{E}(v) = \mathbb{E}(\xi)\mathbb{E}(v) - \mathbb{E}(\xi)\mathbb{E}(v) = 0. \quad (71)$$

Mit der Definition des Korrelationskoeffizienten folgt dann, dass $\rho(\xi, v) = 0$ ist und ξ und v unkorreliert sind.

Kovarianz und Korrelation

Beweis (fortgeführt)

(2) Wir zeigen nun durch Angabe eines Beispiels, dass die Kovarianz von abhängigen Zufallsvariablen ξ und v null sein kann. Zu diesem Zweck betrachten wir den Fall zweier diskreter Zufallsvariablen ξ und v mit Ergebnisräumen $\mathcal{X} = \{-1, 0, 1\}$ und $v = \{0, 1\}$, marginaler WMF von ξ gegeben durch $p_\xi(x) := \frac{1}{3}$ für alle $x \in \mathcal{X}$ und der Definition von

$$v := \xi^2. \quad (72)$$

Die Definition von v impliziert dann die folgende bedingte WMF

$p_{v \xi}(y x)$	$x = -1$	$x = 0$	$x = 1$
$y = 0$	0	1	0
$y = 1$	1	0	1

Die marginale WMF p_ξ und die bedingte WMF $p_{v|\xi}$ implizieren dann die gemeinsame WMF

$p_{\xi,v}(x, y)$	$x = -1$	$x = 0$	$x = 1$	$p_v(y)$
$y = 0$	0	1/3	0	1/3
$y = 1$	1/3	0	1/3	2/3
$p_\xi(x)$	1/3	1/3	1/3	

Es gilt also zum Beispiel

$$p_{\xi,v}(-1, 0) = 0 \neq \frac{1}{9} = \frac{1}{3} \cdot \frac{1}{3} = p_\xi(-1)p_v(0). \quad (73)$$

Die gemeinsame WMF von ξ und v faktorisiert also nicht und ξ und v sind nicht unabhängig.

Beweis (fortgeführt)

Allerdings folgt mit

$$\mathbb{E}(\xi) = \sum_{x \in \mathcal{X}} xp_{\xi}(x) = -1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = 0 \quad (74)$$

und

$$\mathbb{E}(\xi v) = \mathbb{E}(\xi \xi^2) = \mathbb{E}(\xi^3) = \sum_{x \in \mathcal{X}} x^3 p_{\xi}(x) = -1^3 \cdot \frac{1}{3} + 0^3 \cdot \frac{1}{3} + 1^3 \cdot \frac{1}{3} = 0, \quad (75)$$

sowie dem Kovarianzverschiebungssatz, dass

$$\mathbb{C}(\xi, v) = \mathbb{E}(\xi v) - \mathbb{E}(\xi)\mathbb{E}(v) = \mathbb{E}(\xi^3) - \mathbb{E}(\xi)\mathbb{E}(v) = 0 - 0 \cdot \mathbb{E}(v) = 0. \quad (76)$$

Die Kovarianz von ξ und v ist also null und ξ und v damit unkorreliert, obwohl ξ und v nicht unabhängig sind.

□

Erwartungswert

Varianz und Standardabweichung

Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung

Kovarianz und Korrelation

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die Definition des Erwartungswerts einer Zufallsvariable wieder.
2. Geben Sie die Interpretation der Erwartungswerts einer Zufallsvariable wieder
3. Berechnen Sie den Erwartungswert einer Bernoulli Zufallsvariable.
4. Geben Sie das Theorem zu den Eigenschaften des Erwartungswerts wieder.
5. Geben Sie die Definition der Varianz und der Standardabweichung einer Zufallsvariable wieder.
6. Geben Sie die Interpretation der Varianz einer Zufallsvariable wieder.
7. Berechnen Sie die Varianz einer Bernoulli Zufallsvariable.
8. Geben Sie das Theorem zum Varianzverschiebungssatz wieder.
9. Geben Sie das Theorem zu den Eigenschaften der Varianz wieder.
10. Geben Sie die Definition des Begriffs einer Stichprobe wieder.
11. Geben Sie die Definitionen von Stichprobenmittel, -varianz und -standardabweichung wieder.
12. Geben Sie die Definition von Kovarianz und Korrelation zweier Zufallsvariablen wieder.
13. Geben Sie das Theorem zum Kovarianzverschiebungssatz wieder.
14. Geben Sie das Theorem zu Varianzen von Summen und Differenzen von Zufallsvariablen wieder.
15. Geben Sie das Theorem zur Korrelation und Unabhängigkeit zweier Zufallsvariablen wieder.

Selbstkontrollfragen

1. Siehe Definition (Erwartungswert).
2. Siehe Bemerkungen zu Definition (Erwartungswert).
3. Siehe Beweis zu Beispiel (Erwartungswert einer Bernoulli Zufallsvariable).
4. Siehe Theorem (Eigenschaften des Erwartungswerts).
5. Siehe Definition (Varianz und Standardabweichung).
6. Siehe Bemerkungen zu Definition (Varianz und Standardabweichung).
7. Siehe Beweis zu Beispiel (Varianz einer Bernoulli Zufallsvariable).
8. Siehe Theorem (Varianzverschiebungssatz).
9. Siehe Theorem (Eigenschaften der Varianz).
10. Siehe Definition (Stichprobenmittel, -varianz, -standardabweichung).
11. Siehe Definition (Stichprobenmittel, -varianz, -standardabweichung).
12. Siehe Definition (Kovarianz und Korrelation).
13. Siehe Theorem (Kovarianzverschiebungssatz).
14. Siehe Theorem (Varianzen von Summen und Differenzen von Zufallsvariablen).
15. Siehe Theorem (Korrelation und Unabhängigkeit).



Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2023/24

Prof. Dr. Dirk Ostwald

(6) Ungleichungen und Grenzwerte

Wahrscheinlichkeitsungleichungen

Erwartungswertungleichungen

Gesetze der Großen Zahl

Zentrale Grenzwertsätze

Selbstkontrollfragen

Wahrscheinlichkeitsungleichungen

Erwartungswertungleichungen

Gesetze der Großen Zahl

Zentrale Grenzwertsätze

Selbstkontrollfragen

Theorem (Markov Ungleichung)

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

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

Bemerkungen

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

Wahrscheinlichkeitsungleichungen

Beweis

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

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

weil ξ nicht-negativ ist. Es folgt dann

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

Dabei gilt die erste Ungleichung weil

$$\int_0^x s p(s) ds \geq 0 \quad (4)$$

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

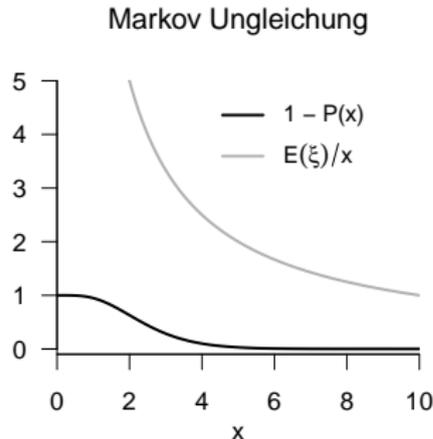
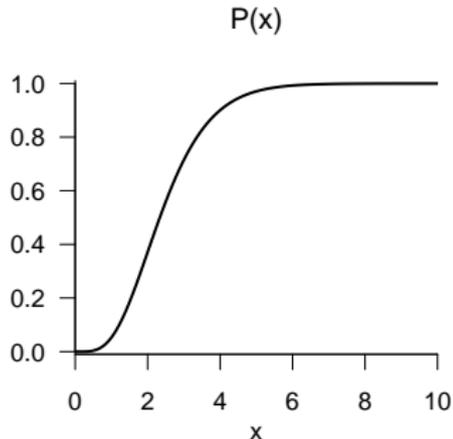
$$\mathbb{E}(\xi) \geq x \mathbb{P}(\xi \geq x) \Leftrightarrow \mathbb{P}(\xi \geq x) \leq \frac{\mathbb{E}(\xi)}{x}. \quad (5)$$

□

Wahrscheinlichkeitsungleichungen

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

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



Theorem (Chebyshev Ungleichung)

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

$$\mathbb{P}(|\xi - \mathbb{E}(\xi)| \geq x) \leq \frac{\mathbb{V}(\xi)}{x^2}. \quad (6)$$

Bemerkungen

- Die Chebyshev Ungleichung setzt Abweichungen vom Erwartungswert in Bezug zur Varianz.
- Zum Beispiel gilt

$$\mathbb{P}(|\xi - \mathbb{E}(\xi)| \geq 3\sqrt{\mathbb{V}(\xi)}) \leq \frac{\mathbb{V}(\xi)}{(3\sqrt{\mathbb{V}(\xi)})^2} = \frac{1}{9}. \quad (7)$$

Wahrscheinlichkeitsungleichungen

Beweis

Wir halten zunächst fest, dass für $a, b \in \mathbb{R}$ gilt, dass aus $a^2 \geq b^2$ folgt, dass $|a| \geq b$. Dazu betrachten wir die folgenden vier möglichen Fälle.

(1) $a^2 \geq b^2$ für $a \geq 0$ und $b \geq 0$. Dann gilt

$$a^2 \geq b^2 \Rightarrow \sqrt{a^2} \geq \sqrt{b^2} \Rightarrow a \geq b \Rightarrow |a| \geq b. \quad (8)$$

(2) $a^2 \geq b^2$ für $a \leq 0$ und $b \geq 0$. Dann gilt

$$a^2 \geq b^2 \Rightarrow \sqrt{a^2} \geq \sqrt{b^2} \Rightarrow -a \geq b \Rightarrow |a| \geq b. \quad (9)$$

(3) $a^2 \geq b^2$ für $a \geq 0$ und $b \leq 0$. Dann gilt

$$a^2 \geq b^2 \Rightarrow \sqrt{a^2} \geq \sqrt{b^2} \Rightarrow a \geq -b \geq b \Rightarrow |a| \geq b. \quad (10)$$

(4) $a^2 \geq b^2$ für $a \leq 0$ und $b \leq 0$. Dann gilt

$$a^2 \geq b^2 \Rightarrow \sqrt{a^2} \geq \sqrt{b^2} \Rightarrow -a \geq -b \geq b \Rightarrow |a| \geq b. \quad (11)$$

Als nächstes definieren wir $v := (\xi - \mathbb{E}(\xi))^2$. Dann folgt aus der Markov Ungleichung

$$\mathbb{P}(v \geq x^2) \leq \frac{\mathbb{E}(v)}{x^2} \Leftrightarrow \mathbb{P}((\xi - \mathbb{E}(\xi))^2 \geq x^2) \leq \frac{\mathbb{E}((\xi - \mathbb{E}(\xi))^2)}{x^2} \Leftrightarrow \mathbb{P}(|\xi - \mathbb{E}(\xi)| \geq x) \leq \frac{\mathbb{V}(\xi)}{x^2}. \quad (12)$$

□

Wahrscheinlichkeitsungleichungen

Erwartungswertungleichungen

Gesetze der Großen Zahl

Zentrale Grenzwertsätze

Selbstkontrollfragen

Theorem (Cauchy-Schwarz Ungleichung)

ξ und v seien zwei Zufallsvariablen und $\mathbb{E}(\xi v)$ sei endlich. Dann gilt

$$\mathbb{E}(\xi v)^2 \leq \mathbb{E}(\xi^2) \mathbb{E}(v^2). \quad (13)$$

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.
- Für einen Beweis verweisen wir auf DeGroot and Schervish (2012), Theorem 4.6.2.

Theorem (Korrelationsungleichung)

ξ und v seien Zufallsvariablen mit $\mathbb{V}(\xi), \mathbb{V}(v) > 0$. Dann gelten

$$\frac{\mathbb{C}(\xi, v)^2}{\mathbb{V}(\xi)\mathbb{V}(v)} \leq 1 \text{ und } -1 \leq \rho(\xi, v) \leq 1. \quad (14)$$

Bemerkung

- Korrelationen nehmen also immer Werte im Intervall $[-1, 1]$ an.

Erwartungswertungleichungen

Beweis

Mit der Cauchy-Schwarz Ungleichung für zwei Zufallsvariablen α und β gilt, dass

$$\mathbb{E}(\alpha\beta)^2 \leq \mathbb{E}(\alpha^2) \mathbb{E}(\beta^2). \quad (15)$$

Wir definieren nun $\alpha := \xi - \mathbb{E}(\xi)$ und $\beta := v - \mathbb{E}(v)$. Dann besagt die Cauchy-Schwarz Ungleichung gerade, dass

$$\mathbb{E}((\xi - \mathbb{E}(\xi))(v - \mathbb{E}(v)))^2 \leq \mathbb{E}((\xi - \mathbb{E}(\xi))^2) \mathbb{E}((v - \mathbb{E}(v))^2). \quad (16)$$

Also gilt

$$C(\xi, v)^2 \leq V(\xi)V(v) \Leftrightarrow \frac{C(\xi, v)^2}{V(\xi)V(v)} \leq 1. \quad (17)$$

Weiterhin folgt aus der Definition der Korrelation dann sofort, dass auch

$$\rho(\xi, v)^2 \leq 1. \quad (18)$$

Dann gilt aber auch

$$|\rho(\xi, v)| \leq 1 \Leftrightarrow -1 \leq \rho(\xi, v) \leq 1, \quad (19)$$

denn

(1) für $\rho(\xi, v) \geq 0$ gilt $\rho(\xi, v)^2 \leq 1 \Rightarrow \sqrt{\rho(\xi, v)^2} \leq \sqrt{1} \Rightarrow \rho(\xi, v) \leq 1 \Rightarrow |\rho(\xi, v)| \leq 1$ und

(2) für $\rho(\xi, v) < 0$ gilt $\rho(\xi, v)^2 \leq 1 \Rightarrow \sqrt{\rho(\xi, v)^2} \leq \sqrt{1} \Rightarrow -\rho(\xi, v) \leq 1 \Rightarrow |\rho(\xi, v)| \leq 1$

□

Theorem (Jensensche Ungleichung)

ξ sei eine Zufallsvariable und $g : \mathbb{R} \rightarrow \mathbb{R}$ eine konvexe Funktion, d.h.

$$g(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda g(x_1) + (1 - \lambda)g(x_2) \quad (20)$$

für alle $x_1, x_2 \in \mathbb{R}, \lambda \in [0, 1]$. Dann gilt

$$\mathbb{E}(g(\xi)) \geq g(\mathbb{E}(\xi)). \quad (21)$$

Analog sei $g : \mathbb{R} \rightarrow \mathbb{R}$ eine konkave Funktion, d.h.

$$g(\lambda x_1 + (1 - \lambda)x_2) \geq \lambda g(x_1) + (1 - \lambda)g(x_2) \quad (22)$$

für alle $x_1, x_2 \in \mathbb{R}, \lambda \in [0, 1]$. Dann gilt

$$\mathbb{E}(g(\xi)) \leq g(\mathbb{E}(\xi)). \quad (23)$$

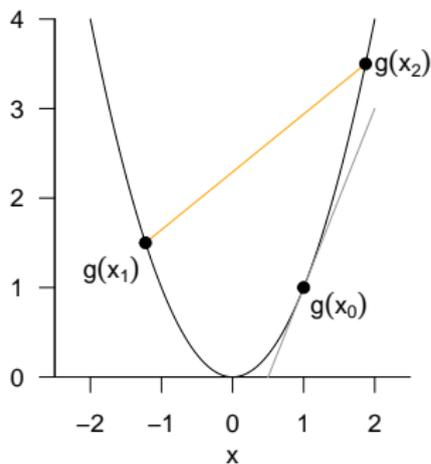
Bemerkungen

- Bei konvexem g liegt der Funktionsgraph unter der Geraden von $g(x_1)$ zu $g(x_2)$.
- Bei konkavem g liegt der Funktionsgraph über der Geraden von $g(x_1)$ zu $g(x_2)$.
- Der Logarithmus ist eine konkave Funktion, also gilt $\mathbb{E}(\ln \xi) \leq \ln \mathbb{E}(\xi)$.

Erwartungswertungleichungen

Visualisierung einer konvexen Funktion

$$- g(x) := x^2$$



$$- \lambda g(x_1) + (1 - \lambda)g(x_2) \text{ f\"ur } x_1 := -\sqrt{1.5}, x_2 := \sqrt{3.5}, \lambda \in [0, 1]$$

$$- t(x) := g(x_0) + g'(x_0)(x - x_0) \text{ f\"ur } x_0 := 1$$

Erwartungswertungleichungen

Beweis

Es sei g eine konvexe Funktion. Dann gilt für die Tangente t von g in $x_0 \in \mathbb{R}$ für alle $x \in \mathbb{R}$, dass

$$g(x) \geq t(x) := g(x_0) + g'(x_0)(x - x_0) \quad (24)$$

Wir setzen nun $x := \xi$ und $x_0 := \mathbb{E}(\xi)$. Dann gilt mit obiger Ungleichung, dass

$$g(\xi) \geq g(\mathbb{E}(\xi)) + g'(\mathbb{E}(\xi))(\xi - \mathbb{E}(\xi)) \quad (25)$$

Erwartungswertbildung ergibt dann

$$\begin{aligned} \mathbb{E}(g(\xi)) &\geq \mathbb{E}(g(\mathbb{E}(\xi))) + \mathbb{E}(g'(\mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))) \\ &\Leftrightarrow \mathbb{E}(g(\xi)) \geq g(\mathbb{E}(\xi)) + g'(\mathbb{E}(\xi))\mathbb{E}((\xi - \mathbb{E}(\xi))) \\ &\Leftrightarrow \mathbb{E}(g(\xi)) \geq g(\mathbb{E}(\xi)) + g'(\mathbb{E}(\xi))(\mathbb{E}(\xi) - \mathbb{E}(\xi)) \\ &\Leftrightarrow \mathbb{E}(g(\xi)) \geq g(\mathbb{E}(\xi)). \end{aligned} \quad (26)$$

Sei nun g eine konkave Funktion. Dann ist $-g$ eine konvexe Funktion. Mit der Jensenschen Ungleichung für konvexe Funktionen folgt dann die Jensensche Ungleichung für konkave Funktionen aus

$$\begin{aligned} \mathbb{E}(-g(\xi)) &\geq -g(\mathbb{E}(\xi)) \\ \Leftrightarrow -\mathbb{E}(g(\xi)) &\geq -g(\mathbb{E}(\xi)) \\ \Leftrightarrow \mathbb{E}(g(\xi)) &\leq g(\mathbb{E}(\xi)). \end{aligned} \quad (27)$$

□

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 dem Schwachen Gesetz.

Theorem (Schwaches Gesetz der Großen Zahl)

ξ_1, \dots, ξ_n seien u.i.v. Zufallsvariablen mit Erwartungswert $\mathbb{E}(\xi_i) =: \mu$ für alle $i = 1, \dots, n$ und

$$\bar{\xi}_n := \frac{1}{n} \sum_{i=1}^n \xi_i \quad (28)$$

sei das Stichprobenmittel der $\xi_i, i = 1, \dots, n$. Dann gilt für ein beliebig kleines $\epsilon > 0$, dass

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\bar{\xi}_n - \mu| \geq \epsilon) = 0. \quad (29)$$

Bemerkungen

- Die Wahrscheinlichkeit, dass das Stichprobenmittel nahe am Erwartungswert der zugrundeliegenden Verteilungen der ξ_i liegt, nähert sich 1, wenn $n \rightarrow \infty$; die Wahrscheinlichkeit, dass das Stichprobenmittel weit entfernt vom Erwartungswert der zugrundeliegenden Verteilungen der ξ_i liegt, nähert sich 0, wenn $n \rightarrow \infty$.
- Die hier betrachtete Folgenkonvergenzart ist die sogenannte *Konvergenz in Wahrscheinlichkeit*.

Gesetze der Großen Zahl

Beweis

Mit der Linearkombinationseigenschaft des Erwartungswerts gilt zunächst, dass

$$\mu = \mathbb{E}(\xi_i) = \frac{1}{n} n \mathbb{E}(\xi_i) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(\xi_i) = \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n \xi_i\right) = \mathbb{E}(\bar{\xi}_n). \quad (30)$$

Der Erwartungswert der i ten Stichprobenvariable ξ_i stimmt also mit dem Erwartungswert des Stichprobenmittels $\bar{\xi}_n$ überein. Weiterhin halten wir fest, dass mit der Linearkombinationseigenschaft bei Unabhängigkeit der Varianz gilt, dass

$$\mathbb{V}(\bar{\xi}_n) = \mathbb{V}\left(\frac{1}{n} \sum_{i=1}^n \xi_i\right) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(\xi_i) = \frac{1}{n^2} n \mathbb{V}(\xi_i) = \frac{\mathbb{V}(\xi_i)}{n}. \quad (31)$$

für $i = 1, \dots, n$. Die Varianz des Stichprobenmittels ergibt sich also durch Teilen der Varianz der i ten Stichprobe durch n . Mit der Chebyshev Ungleichung gilt dann aber

$$\mathbb{P}(|\bar{\xi}_n - \mathbb{E}(\bar{\xi}_n)| \geq \epsilon) \leq \frac{\mathbb{V}(\bar{\xi}_n)}{\epsilon^2} = \frac{\mathbb{V}(\xi_i)}{n\epsilon^2}. \quad (32)$$

Dann aber gilt für beliebige $\mathbb{V}(\xi_i) \geq 0$ und $\epsilon > 0$ und mit der Nichtnegativität der Wahrscheinlichkeit

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\bar{\xi}_n - \mathbb{E}(\bar{\xi}_n)| \geq \epsilon) \leq \lim_{n \rightarrow \infty} \frac{\mathbb{V}(\xi_i)}{n\epsilon^2} \Leftrightarrow \lim_{n \rightarrow \infty} \mathbb{P}(|\bar{\xi}_n - \mu| \geq \epsilon) \leq 0 \Leftrightarrow \lim_{n \rightarrow \infty} \mathbb{P}(|\bar{\xi}_n - \mu| \geq \epsilon) = 0 \quad (33)$$

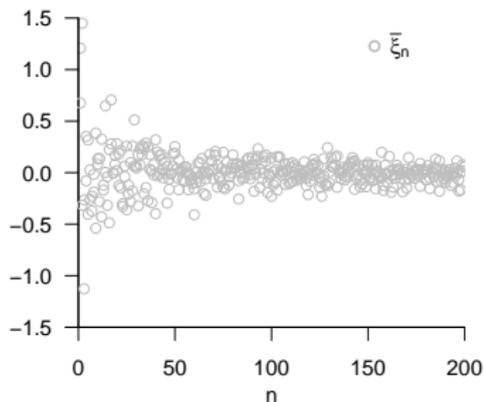
und es ist alles gezeigt. \square

Gesetze der Großen Zahl

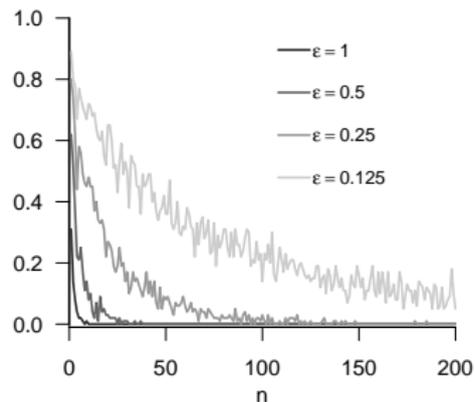
Beispiel ($\xi_1, \dots, \xi_n \sim N(0, 1)$)

- Die linke Abbildung zeigt Realisationen von $\bar{\xi}_n$ als Funktion von n .
- Die rechte Abbildung zeigt Schätzungen von $\mathbb{P}(|\bar{\xi}_n - \mu| \geq \epsilon)$ als Funktionen von n und ϵ .

Stichprobenmittelrealisationen



Wahrscheinlichkeitschätzungen



Wahrscheinlichkeitsungleichungen

Erwartungswertungleichungen

Gesetze der Großen Zahl

Zentrale Grenzwertsätze

Selbstkontrollfragen

Überblick

- Die Zentralen Grenzwertsätze besagen, dass die Summe von unabhängigen Zufallsvariablen mit Erwartungswert 0 asymptotisch, d.h. für unendlich viele Zufallsvariablen, normalverteilt mit Erwartungswertparameter 0 ist.
- Modelliert man einen Datenpunkt v also als Summe eines deterministischen Einflusses μ und der Summe

$$\varepsilon := \sum_{i=1}^n \xi_i \quad (34)$$

einer Vielzahl von unabhängigen Zufallsvariablen $\xi_i, i = 1, \dots, n$, welche unbekannte Störeinflüsse beschreiben, so ist für großes n die Annahme

$$v = \mu + \varepsilon \text{ mit } \varepsilon \sim N(0, \sigma^2) \quad (35)$$

also mathematisch gerechtfertigt.

- Wie wir später sehen werden, liegt die Annahme in Gleichung (35) vielen probabilistischen 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 aber auf die Angabe von Beweisen.
- Zur mathematik-geschichtlichen Genese der Zentralen Grenzwertsätze siehe z.B. Fischer (2011).

Theorem (Zentraler Grenzwertsatz nach Lindenberg und Lévy)

ξ_1, \dots, ξ_n seien unabhängig und identisch verteilte Zufallsvariablen mit

$$\mathbb{E}(\xi_i) := \mu \text{ und } \mathbb{V}(\xi_i) := \sigma^2 > 0 \text{ für alle } i = 1, \dots, n. \quad (36)$$

Weiterhin sei P_{ζ_n} die KVF der Zufallsvariable

$$\zeta_n := \sqrt{n} \frac{\bar{\xi}_n - \mu}{\sigma}. \quad (37)$$

Dann gilt für alle $z \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} P_{\zeta_n}(z) = \Phi(z), \quad (38)$$

wobei Φ die KVF der Standardnormalverteilung bezeichnet.

Bemerkung

- Die hier betrachtete Folgenkonvergenzart ist die sogenannte *Konvergenz in Verteilung*.
- Wir zeigen später, dass damit für $n \rightarrow \infty$ asymptotisch auch gilt, dass

$$\sum_{i=1}^n \xi_i \sim N(n\mu, n\sigma^2) \text{ und } \bar{\xi}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right). \quad (39)$$

Zentrale Grenzwertsätze

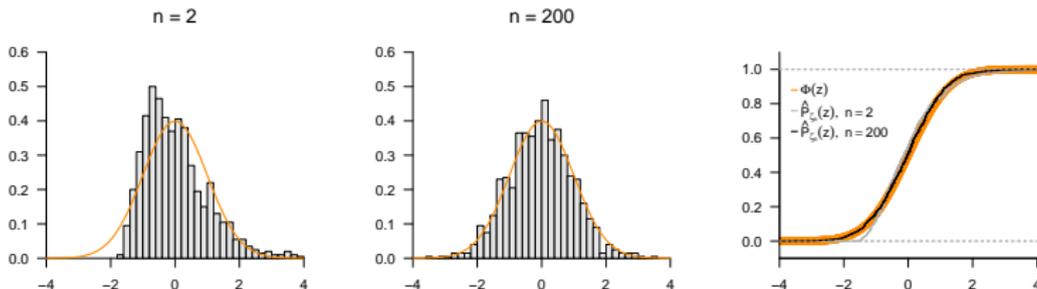
Beispiel $(\xi_1, \dots, \xi_n \sim \chi^2(k))$

- Wir halten ohne Beweis fest, dass $\mathbb{E}(\xi_i) = k$ und $\mathbb{V}(\xi_i) = 2k$.
- Wir betrachten das Szenario $\xi_i \sim \chi^2(3)$ für $i = 1, \dots, n$.
- Die linken Abbildungen zeigen Histogrammschätzer der Wahrscheinlichkeitsdichte von

$$\zeta_n := \sqrt{n} \frac{\bar{\xi}_n - \mu}{\sigma} \quad (40)$$

basierend auf 1000 Realisationen von ζ_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)

ξ_1, \dots, ξ_n seien unabhängige aber nicht notwendigerweise identisch verteilten Zufallsvariablen mit

$$\mathbb{E}(\xi_i) := \mu_i \text{ und } \mathbb{V}(\xi_i) := \sigma_i^2 > 0 \text{ für alle } i = 1, \dots, n \quad (41)$$

und

$$\mathbb{E}(|\xi_i - \mu_i|^3) < \infty \text{ und } \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \mathbb{E}(|\xi_i - \mu_i|^3)}{(\sum_{i=1}^n \sigma_i^2)^{3/2}} = 0. \quad (42)$$

Weiterhin sei P_{ζ_n} die KVF der Zufallsvariable

$$\zeta_n := \frac{\sum_{i=1}^n \xi_i - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}}. \quad (43)$$

Dann gilt für alle $z \in \mathbb{R}$, dass

$$\lim_{n \rightarrow \infty} P_{\zeta_n}(z) = \Phi(z), \quad (44)$$

wobei Φ die KVF der Standardnormalverteilung bezeichnet.

Bemerkungen

- Die hier betrachtete Folgenkonvergenzart ist die sogenannte *Konvergenz in Verteilung*.
- Wir zeigen später, dass dann auch gilt, dass $\sum_{i=1}^n \xi_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$.

Wahrscheinlichkeitsungleichungen

Erwartungswertungleichungen

Gesetze der Großen Zahl

Zentrale Grenzwertsätze

Selbstkontrollfragen

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. Geben Sie das Schwache Gesetz der Großen Zahl wieder.
6. Erläutern Sie den Zentralen Grenzwertsatz nach Lindenberg und Lévy.
7. Erläutern Sie den Zentralen Grenzwertsatz nach Liapunov.
8. Warum sind die Zentralen Grenzwertsätze für die probabilistische Modellbildung wichtig?

Selbstkontrollfragen - Antworten

1. Siehe Theorem (Markov Ungleichung).
2. Siehe Theorem (Chebyshev Ungleichung).
3. Siehe Theorem (Cauchy-Schwarz Ungleichung).
4. Siehe Theorem (Korrelationsungleichung).
5. Siehe Theorem (Schwaches Gesetz der Großen Zahl).
6. Der Zentrale Grenzwertsatz nach Lindenberg und Lévy besagt, dass die standardisierte Summe unabhängig und identisch verteilter Zufallsvariablen asymptotisch normalverteilt ist.
7. Der Zentrale Grenzwertsatz nach Liapunov besagt, dass die standardisierte Summe unabhängig und nicht notwendig identisch verteilter Zufallsvariablen asymptotisch normalverteilt ist.
8. Die Zentralen Grenzwertsätze begründen die häufige Annahme normalverteilter “Stör-”, “Fehler-”, “Abweichungs-” oder “Unsicherheitsvariablen” in probabilistischen Modellen.

- DeGroot, Morris H., and Mark J. Schervish. 2012. *Probability and Statistics*. 4th ed. Boston: Addison-Wesley.
- Fischer, Hans. 2011. *A History of the Central Limit Theorem*. New York, NY: Springer New York. <https://doi.org/10.1007/978-0-387-87857-7>.



Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2023/24

Prof. Dr. Dirk Ostwald

(7) Transformationen der Normalverteilung

Vorbemerkungen

Transformationstheoreme

Standardtransformationen

Selbstkontrollfragen

Vorbemerkungen

Transformationstheoreme

Standardtransformationen

Selbstkontrollfragen

Realisierungen von Zufallsvariablen

Der einzelne Wert, den eine Zufallsvariable bei jedem Durchgang eines Zufallsvorgangs 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 $\xi_i \sim N(\mu, \sigma^2)$, $i = 1, \dots, n$ durch `rnorm(n,mu,sigma)` gegeben ist.

```
rnorm(1,0,1)           # \xi_i \sim N(0,1)
```

```
[1] -0.915625
```

```
rnorm(1,10,1)         # \xi_i \sim N(10,1)
```

```
[1] 9.099382
```

```
rnorm(3,5,sqrt(2))    # \xi_i \sim N(5,2), i = 1,2,3 (u.i.v.)
```

```
[1] 2.580926 4.424935 3.292730
```

```
rnorm(1e1,5,sqrt(2))  # \xi_i \sim N(5,2), i = 1,\dots,10 (u.i.v.)
```

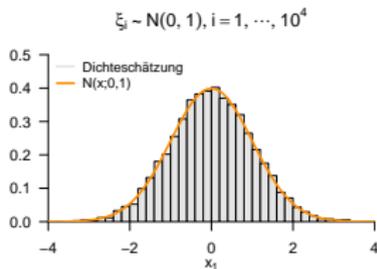
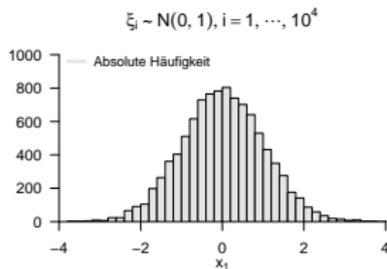
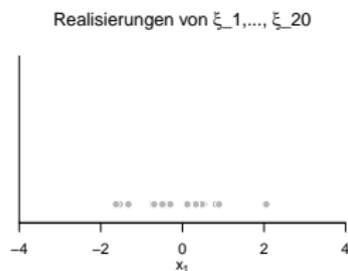
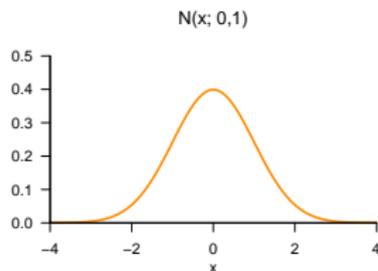
```
[1] 3.607981 5.815495 6.229038 2.896216 5.565196 6.618827 2.494256 3.603984
```

```
[9] 4.800696 5.106866
```

Vorbemerkungen

Realisierungen von Zufallsvariablen

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.



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 ξ normalverteilt ist, wie ist dann eine Zufallsvariable v , die sich durch Transformation von ξ 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 Inferenz 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. $\xi \sim N(0, 1)$ und ihre Transformation $v := \xi^2$ und sind

$$x_1 = 0.10, x_2 = -0.20, x_3 = 0.80 \tag{1}$$

drei u.i.v. Realisierungen von ξ , so entspricht dies den u.i.v. Realisierungen

$$y_1 = x_1^2 = 0.01, y_2 = x_2^2 = 0.04, y_3 = x_3^2 = 0.64 \tag{2}$$

von v . In diesem Beispiel fällt auf, dass v keine negativen Werte annimmt, die Verteilung von v ordnet negativen Werten daher Wahrscheinlichkeitsdichten von 0 zu.

Simulation der Transformation normalverteilter Zufallsvariablen in R

```
# Simulationsspezifikation
n      = 1e4                # Anzahl von u.i.v Realisierungen (ZVen)
mu     = 1                  # Erwartungswertparameter von \xi
sigsqr = 2                  # Varianzparameter von \xi

# Quadrieren einer Zufallsvariable
x      = rnorm(n, mu, sqrt(sigsqr)) # Realisierungen x_i, i = 1,...,n von \xi
y      = x^2                 # Realisierungen y_i = x_i^2 von \ups

# Ausgabe der ersten acht Werte
print(x[1:8], digits = 2)
```

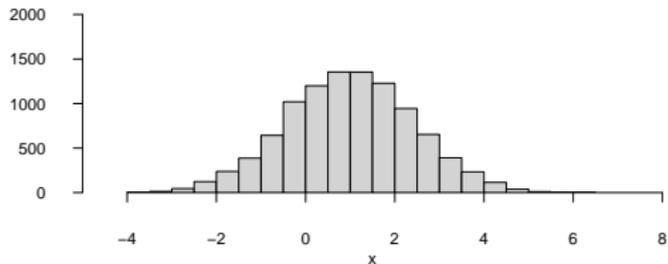
```
[1] -0.173  0.109  1.108  2.652  1.947  0.382  2.903  0.049
```

```
print(y[1:8], digits = 2)
```

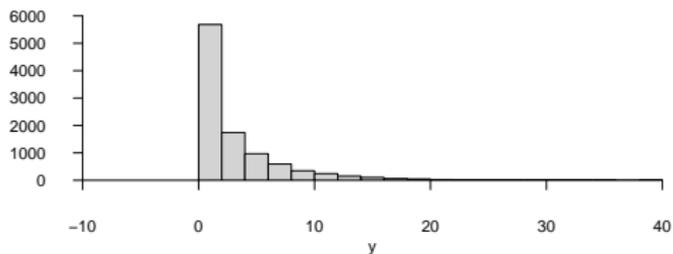
```
[1] 0.0301 0.0120 1.2285 7.0314 3.7896 0.1460 8.4266 0.0024
```

Simulation der Transformation normalverteilter Zufallsvariablen in R

Histogramm von unabhängigen Realisierungen von ξ



Histogramm von unabhängigen Realisierungen von ν



Theorem (Transformation eines Zufallsvektors)

$\xi : \Omega \rightarrow \mathcal{X}$ sei ein Zufallsvektor und $f : \mathcal{X} \rightarrow \mathbb{R}^m$ sei eine multivariate vektorwertige Funktion. Dann ist

$$v : \Omega \rightarrow \mathbb{R}, \omega \mapsto v(\omega) := (f \circ \xi)(\omega) := f(\xi(\omega)) \quad (3)$$

ein Zufallsvektor.

Bemerkungen

- Das Theorem formalisiert die oben etablierte Intuition, dass die Anwendung einer (deterministischen) Funktion auf eine zufällige Größe im Allgemeinen wieder eine zufällige Größe ergibt. Wir verzichten auf einen Beweis.
- In einem Beweis müsste die Messbarkeit von v als Folge der Messbarkeit von ξ nachgewiesen werden.
- Im Folgenden ist oft $\mathcal{X} := \mathbb{R}$ und $f : \mathbb{R} \rightarrow \mathbb{R}$.
- Wir schreiben in diesem Fall in der Regel einfach $v := f(\xi)$ und nennen v die *transformierte Zufallsvariable*.

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 ξ eine Zufallsvariable mit WDF p_ξ und $v := f(\xi)$ die durch f transformierte Zufallsvariable ist, dann gilt für die WDF von v die folgende Formel: $p_v := \{\text{Formel}\}$ ".

Im Abschnitt *Standardtransformationen* diskutieren wir sechs Standardtransformationen normalverteilter Zufallsvariablen, die in der frequentistischen Inferenz und damit im weiteren Verlauf des Kurses zentrale Rollen spielen. Diese Aussagen sind von der allgemeinen Form "Wenn $\xi_i, i = 1, \dots, n$ unabhängig und identisch normalverteilte Zufallsvariablen sind und $v := f(\xi_1, \dots, \xi_n)$ eine Transformation dieser Zufallsvariablen ist, dann ist die WDF von v durch die Formel $p_v := \{\text{Formel}\}$ gegeben und man nennt die Verteilung von v *Verteilungsname*".

Die Aussagen im Abschnitt Standardtransformationen sind für die Frequentistische Inferenz zentral, weil

- (1) die Zentralen Grenzwertsätze die Annahme additiv unabhängig normalverteilter Störvariablen rechtfertigt,
- (2) wir sehen werden, dass Schätzer und Statistiken Transformationen von Zufallsvariablen sind, und
- (3) Parameterschätzergütekriterien, Konfidenzintervalle und Hypothesentests durch die Verteilungen der ihnen jeweils zugrundeliegenden Schätzer und Statistiken charakterisiert und begründet werden.

Ausblick

Das probabilistische Standardmodell von n Datenpunkten hat die Form

$$v_i := \mu_i + \varepsilon_i, i = 1, \dots, n. \quad (4)$$

Die Zufallsvariable v_i dient dabei als das Modell des i ten Datenpunktes $y_i \in \mathbb{R}$, d.h. y_i wird als Realisierung von v_i modelliert. Die Normalverteilung $v_i \sim N(\mu_i, \varepsilon)$ der Zufallsvariable v_i ergibt sich dabei wie wir später sehen werden aus der linear-affinen Transformation der Zufallsvariable ε_i unter der Abbildung $f(\varepsilon_i) := \mu_i + \varepsilon_i$

$\mu_i \in \mathbb{R}$ repräsentiert den deterministischen Aspekt des Datenpunktmodells und liefert die theoretische Erklärung für den Wert von v_i . $\varepsilon_i \sim N(0, \sigma^2)$ dagegen repräsentiert den stochastischen Aspekt des Datenpunktmodells und liefert im Sinne der Zentralen Grenzwertsätze die theoretische Erklärung für die Differenz von μ_i und v_i als Resultat der Addition vieler weiterer Einflüsse in der Generation von v_i über μ_i hinaus.

Statistiken und Schätzer, also Funktionen von $v_i, i = 1, \dots, n$, entsprechen damit im probabilistischen Standardmodell Transformationen von normalverteilten Zufallsvariablen.

Vorbemerkungen

Transformationstheoreme

Standardtransformationen

Selbstkontrollfragen

Überblick

Das *univariate WDF Transformationstheorem bei bijektiven Abbildungen* liefert eine Formel zur Berechnung der WDF p_v von $v := f(\xi)$, wenn ξ eine Zufallsvariable mit WDF p_ξ ist und f eine bijektive Funktion ist.

Das *univariate WDF Transformationstheorem bei linear affiner Abbildung* gibt eine Formel zur Berechnung der WDF p_v von $v := f(\xi)$ an, wenn ξ eine Zufallsvariable mit WDF p_ξ ist und f eine linear-affine Funktion ist.

Das *univariate WDF Transformationstheorem bei stückweisen bijektiven Abbildungen* gibt eine Formel zur Berechnung der WDF p_v von $v := f(\xi)$ an, wenn ξ eine Zufallsvariable mit WDF p_ξ ist und f zumindest in Teilen bijektiv ist.

Das *multivariate WDF Transformationstheorem bei bijektiven Abbildungen* liefert eine Formel zur Berechnung der WDF p_v von $v := f(\xi)$, wenn ξ ein Zufallsvektor mit WDF p_ξ ist und f eine bijektive multivariate vektorwertige Funktion ist.

Das *Faltungstheorem* liefert eine Formel zur Berechnung der WDF p_v von $v := \xi_1 + \xi_2$, wenn ξ_1 und ξ_2 zwei Zufallsvariablen mit WDFen p_{ξ_1} und p_{ξ_2} sind.

Theorem (Univariate WDF Transformation bei bijektiven Abbildungen)

ξ sei eine Zufallsvariable mit WDF p_ξ für die $\mathbb{P}(]a, b[) = 1$ gilt, wobei a und/oder b entweder endlich oder unendlich seien. Weiterhin sei

$$v := f(\xi) \tag{5}$$

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 v gegeben durch

$$p_v : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, y \mapsto p_v(y) := \begin{cases} \frac{1}{|f'(f^{-1}(y))|} p_\xi(f^{-1}(y)) & \text{für } y \in f(]a, b[) \\ 0 & \text{für } y \in \mathbb{R} \setminus f(]a, b[). \end{cases} \tag{6}$$

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

$$P_v(y) = \mathbb{P}(v \leq y) = \mathbb{P}(f(\xi) \leq y) = \mathbb{P}(f^{-1}(f(\xi)) \leq f^{-1}(y)) = \mathbb{P}(\xi \leq f^{-1}(y)) = P_\xi(f^{-1}(y)).$$

P_v ist also differenzierbar an allen Stellen y , an denen sowohl f^{-1} als auch P_ξ 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_v sich ergibt wie folgt:

$$p_v(y) = \frac{d}{dy} P_v(y) = \frac{d}{dy} P_\xi(f^{-1}(y)) = p_\xi(f^{-1}(y)) \frac{d}{dy} f^{-1}(y) = \frac{1}{f'(f^{-1}(y))} p_\xi(f^{-1}(y)),$$

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

$$P_v(y) = \mathbb{P}(f(\xi) \leq y) = \mathbb{P}(f^{-1}(f(\xi)) \geq f^{-1}(y)) = \mathbb{P}(\xi \geq f^{-1}(y)) = 1 - P_\xi(f^{-1}(y)),$$

Mit der Kettenregel und dem Satz von der Umkehrabbildung folgt dann

$$p_v(y) = \frac{d}{dy} (1 - P_v(y)) = -\frac{d}{dy} P_\xi(f^{-1}(y)) = -p_\xi(f^{-1}(y)) \frac{d}{dy} f^{-1}(y) = -\frac{1}{f'(f^{-1}(y))} p_\xi(f^{-1}(y)).$$

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-affine Abbildung)

ξ sei eine Zufallsvariable mit WDF p_ξ und es sei

$$v = f(\xi) \text{ mit } f(\xi) := a\xi + b \text{ für } a \neq 0. \quad (7)$$

Dann ist die WDF von v gegeben durch

$$p_v : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, y \mapsto p_v(y) := \frac{1}{|a|} p_\xi \left(\frac{y-b}{a} \right). \quad (8)$$

Bemerkung

- Das Theorem folgt direkt WDF Transformationstheorem bei bijektiven Abbildungen.
- Die Z -Transformation ist ein wichtiger Anwendungsfall.

Transformationstheoreme

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

$$\begin{aligned} p_v : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}, y \mapsto p_v(y) &= \frac{1}{|f'(f^{-1}(y))|} p_\xi(f^{-1}(y)) \\ &= \frac{1}{|a|} p_\xi\left(\frac{y-b}{a}\right). \end{aligned} \quad (12)$$

□

Theorem (WDF Transformation bei stückweise bijektiven Abbildungen)

ξ sei eine Zufallsvariable mit Ergebnisraum \mathcal{X} und WDF p_ξ . Weiterhin sei

$$v = f(\xi), \quad (13)$$

wobei f so beschaffen sei, dass der Ergebnisraum von ξ 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 v 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 v durch

$$p_v : \mathcal{Y} \rightarrow \mathbb{R}_{\geq 0}, y \mapsto p_v(y) := \sum_{i=1}^k 1_{\mathcal{Y}_i}(y) \frac{1}{|f_i'(f_i^{-1}(y))|} p_\xi(f_i^{-1}(y)). \quad (14)$$

gegeben.

Bemerkungen

- Wir verzichten auf einen Beweis.
- Die χ^2 -Transformation ist ein wichtiger Anwendungsfall.

Theorem (Multivariate WDF Transformation bei bijektiven Abbildungen)

ξ sei ein n -dimensionaler Zufallsvektor mit Ergebnisraum \mathbb{R}^n und WDF p_ξ . Weiterhin sei

$$v := f(\xi), \quad (15)$$

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 v durch

$$p_v : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}, y \mapsto p_v(y) := \begin{cases} \frac{1}{|J^f(f^{-1}(y))|} p_\xi(f^{-1}(y)) & \text{for } y \in f(\mathbb{R}^n) \\ 0 & \text{for } y \in \mathbb{R}^n \setminus f(\mathbb{R}^n) \end{cases} \quad (17)$$

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)

ξ_1 und ξ_2 seien zwei kontinuierliche unabhängige Zufallsvariablen mit WDF p_{ξ_1} und p_{ξ_2} , respektive. $v := \xi_1 + \xi_2$ sei die Summe von ξ_1 und ξ_2 . Dann ergibt sich eine WDF der Verteilung von v als

$$p_v(y) = \int_{-\infty}^{\infty} p_{\xi_1}(y - x_2)p_{\xi_2}(x_2) dx_2 = \int_{-\infty}^{\infty} p_{\xi_1}(x_1)p_{\xi_2}(y - x_1) dx_1 \quad (18)$$

Die Formel für die WDF p_v heißt *Faltung* oder *Konvolution* von p_{ξ_1} und p_{ξ_2} .

Bemerkung

- Die Summen- und Mittelwerttransformation sind wichtige Anwendungsfälle.

Transformationstheoreme

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

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}^2, z \mapsto f(z) := \begin{pmatrix} z_1 - x_2 \\ z_2 \end{pmatrix} \quad (20)$$

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

$$Jf(x) = \begin{pmatrix} \frac{\partial}{\partial x_1} f_1(x) & \frac{\partial}{\partial x_2} f_1(x) \\ \frac{\partial}{\partial x_1} f_2(x) & \frac{\partial}{\partial x_2} f_2(x) \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x_1} (x_1 + x_2) & \frac{\partial}{\partial x_2} (x_1 + x_2) \\ \frac{\partial}{\partial x_1} x_2 & \frac{\partial}{\partial x_2} x_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad (22)$$

und die Jacobideterminante damit zu $|Jf(x)| = 1$.

Transformationstheoreme

Beweis (fortgeführt)

Wir halten weiterhin fest, dass die Unabhängigkeit von ξ_1 und ξ_2 impliziert, dass

$$p_{\xi_1, \xi_2}(x_1, x_2) = p_{\xi_1}(x_1)p_{\xi_2}(x_2) \quad (23)$$

impliziert. Einsetzen und Integration hinsichtlich x_2 ergibt dann ergibt dann für $z \in f(\mathbb{R}^2)$

$$\begin{aligned} p_{\zeta}(z) &= \frac{1}{|Jf(f^{-1}(z))|} p_{\xi}(f^{-1}(z)) \\ &= \frac{1}{1} p_{\xi_1, \xi_2}(z_1 - x_2, x_2) \\ &= p_{\xi_1}(z_1 - x_2)p_{\xi_2}(x_2) \end{aligned} \quad (24)$$

Integration über x_2 ergibt dann eine WDF für die marginale Verteilung von ζ_1

$$p_{\zeta_1}(z_1) = \int_{-\infty}^{\infty} p_{\xi_1}(z_1 - x_2)p_{\xi_2}(x_2) dx_2 \quad (25)$$

Mit $\zeta_1 = \xi_1 + \xi_2 = v$ ergibt sich dann die erste Form des Faltungstheorems zu

$$p_v(y) = \int_{-\infty}^{\infty} p_{\xi_1}(y - x_2)p_{\xi_2}(x_2) dx_2. \quad (26)$$

□

Vorbemerkungen

Transformationstheoreme

Standardtransformationen

Selbstkontrollfragen

Überblick

Das *Summentransformationstheorem* besagt, dass die Summe unabhängig normalverteilter Zufallsvariablen wiederum normalverteilt ist und gibt die Parameter dieser Verteilung an.

Das *Mittelwertstransformationstheorem* 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 transformiert.

Das χ^2 -*Transformationstheorem* besagt, dass die Summe quadrierter unabhängiger 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.

Theorem (Summe unabhängig normalverteilter Zufallsvariablen)

Für $i = 1, \dots, n$ seien $\xi_i \sim N(\mu_i, \sigma_i^2)$ unabhängige normalverteilte Zufallsvariablen. Dann gilt für die Summe $v := \sum_{i=1}^n \xi_i$, dass

$$v \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right) \quad (27)$$

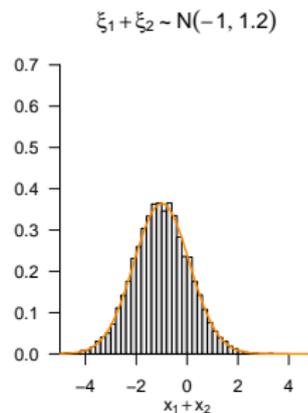
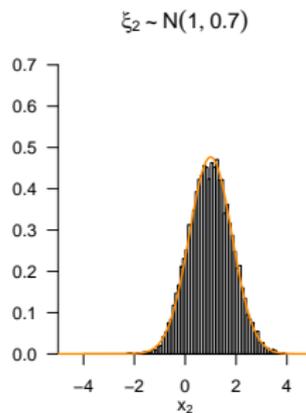
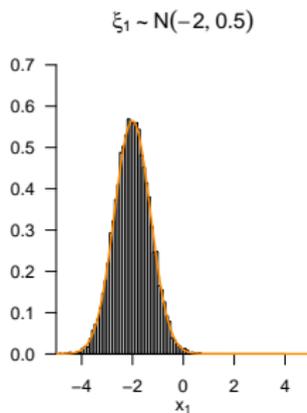
Für unabhängige und identisch normalverteilte Zufallsvariablen $\xi_i \sim N(\mu, \sigma^2)$ gilt folglich

$$v \sim N(n\mu, n\sigma^2). \quad (28)$$

Bemerkungen

- Wir verzichten auf einen Beweis
- Die Mittelwerttransformation ist ein wichtiger Anwendungsfall.
- Die Generalisierung der zentralen Grenzwertsätze sind wichtige Anwendungsfälle.

Standardtransformationen



Theorem (Stichprobenmittel von u.i.v. normalverteilten Zufallsvariablen)

Für $i = 1, \dots, n$ seien $\xi_i \sim N(\mu, \sigma^2)$ unabhängig und identisch normalverteilte Zufallsvariablen. Dann gilt für das Stichprobenmittel $\bar{\xi}_n := \frac{1}{n} \sum_{i=1}^n \xi_i$, dass

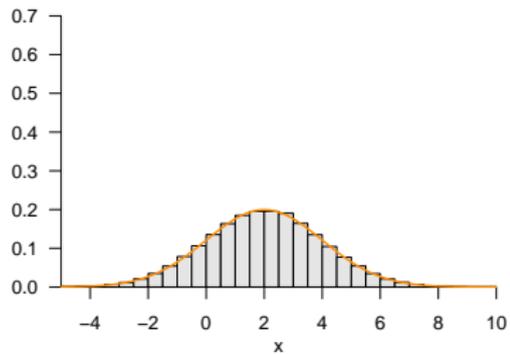
$$\bar{\xi}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right). \quad (29)$$

Bemerkung

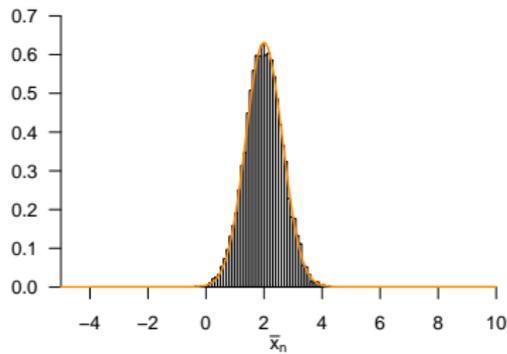
- Die Analyse von Erwartungswertschätzern ist ein wichtiger Anwendungsfall.
- Die Generalisierung der zentralen Grenzwertsätze sind wichtige Anwendungsfälle.

Standardtransformationen

$$\xi_i \sim N(2, 4), \quad i = 1, \dots, 10$$



$$\bar{\xi}_n \sim N(2, 4/10)$$



Beweis

Wir halten zunächst fest, dass mit dem Theorem zur Summe von unabhängig normalverteilten Zufallsvariablen gilt, dass $\bar{\xi}_n = \frac{1}{n} v$ mit $v := \sum_{i=1}^n \xi_i \sim N(n\mu, n\sigma^2)$. Einsetzen in das univariate WDF Transformationstheorem für lineare Funktionen ergibt dann

$$\begin{aligned} p_{\bar{\xi}_n}(\bar{x}_n) &= \frac{1}{|1/n|} N(n\bar{x}_n; n\mu, n\sigma^2) \\ &= \frac{n}{\sqrt{2\pi n\sigma^2}} \exp\left(-\frac{1}{2n\sigma^2} (n\bar{x}_n - n\mu)^2\right) \\ &= \frac{n}{\sqrt{2\pi n\sigma^2}} \exp\left(-\frac{1}{2n\sigma^2} (n\bar{x}_n - n\mu)^2\right) \\ &= nn^{-\frac{1}{2}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(n\bar{x}_n)^2}{2n\sigma^2} + \frac{2(n\bar{x}_n)(n\mu)}{2n\sigma^2} - \frac{(n\mu)^2}{2n\sigma^2}\right) \\ &= \sqrt{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{n\bar{x}_n^2}{2\sigma^2} + \frac{2n\bar{x}_n\mu}{2\sigma^2} - \frac{n\mu^2}{2\sigma^2}\right) \\ &= \frac{1}{1/\sqrt{n}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\bar{x}_n^2}{2(\sigma^2/n)} + \frac{2\bar{x}_n\mu}{2(\sigma^2/n)} - \frac{\mu^2}{2(\sigma^2/n)}\right) \\ &= \frac{1}{\sqrt{2\pi(\sigma^2/n)}} \exp\left(-\frac{1}{2(\sigma^2/n)} (\bar{x}_n - \mu)^2\right) \\ &= N(\bar{x}_n; \mu, \sigma^2/n) \end{aligned} \tag{30}$$

Definition (z -Zufallsvariable)

Z sei eine Zufallsvariable mit Ergebnisraum \mathbb{R} und WDF

$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, z \mapsto p(z) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right). \quad (31)$$

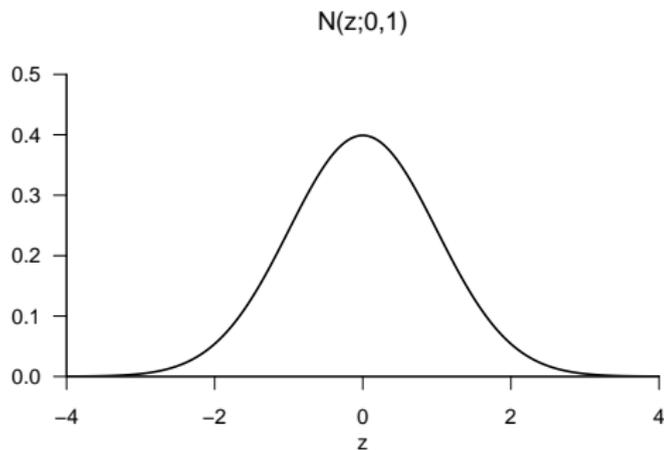
Dann sagen wir, dass Z einer z -Verteilung (oder Standardnormalverteilung) unterliegt und nennen Z eine z -Zufallsvariable. Wir kürzen dies mit $Z \sim N(0, 1)$ ab. Die WDF einer z -Zufallsvariable bezeichnen wir mit

$$N(z; 0, 1) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right). \quad (32)$$

Bemerkung

- Eine z -Zufallsvariable ist eine normalverteilte Zufallsvariable mit $\mu := 0$ und $\sigma^2 := 1$.

Wahrscheinlichkeitsdichtefunktion einer z -Zufallsvariable



Theorem (Z -Transformation)

Es sei $v \sim N(\mu, \sigma^2)$ eine normalverteilte Zufallsvariable. Dann ist die Zufallsvariable

$$Z := \frac{v - \mu}{\sigma} \quad (33)$$

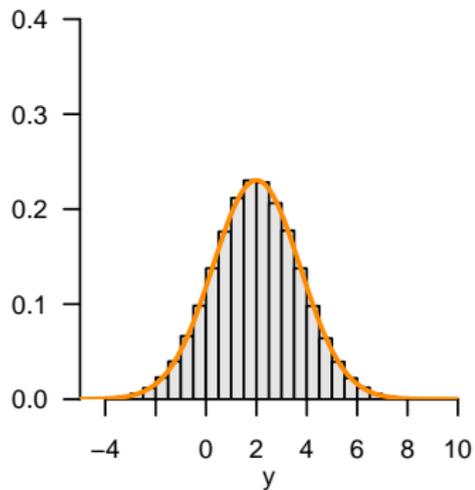
eine Z -verteilte Zufallsvariable, es gilt also $Z \sim N(0, 1)$.

Bemerkungen

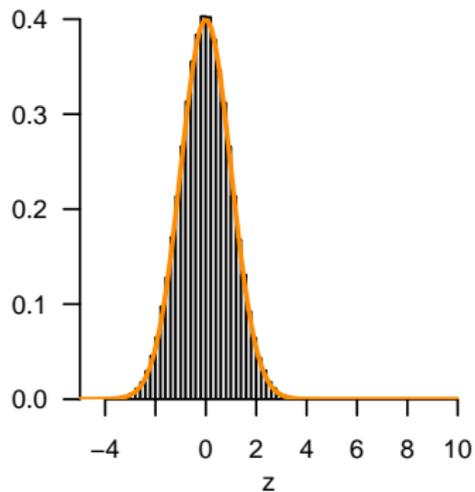
- Z wird hier als $(v - \mu)/\sigma$ definiert. Dass ein solches Z aber eine z -Zufallsvariable ist, muss bewiesen werden und ergibt sich nicht einfach durch die Wahl des Bezeichners für $(v - \mu)/\sigma$, welcher hier zufällig auch Z lautet. In analoger Form gilt diese Bemerkung auch für alle weiteren betrachteten Transformationen.
- Die Z -Konfidenzintervallstatistik und die Z -Teststatistik sind wichtige Anwendungsfälle.

Standardtransformationen

$$v \sim N(2, 3)$$



$$Z = \frac{v - \mu}{\sigma} \sim N(0, 1)$$



Standardtransformationen

Beweis

Wir nutzen das univariate WDF Transformationstheorem für linear-affine Funktionen. Dazu halten wir zunächst fest, dass die Z -Transformation einer Funktion der Form

$$\zeta : \mathbb{R} \rightarrow \mathbb{R}, y \mapsto \zeta(y) := \frac{y - \mu}{\sigma} =: z \quad (34)$$

entspricht. Wir stellen weiterhin fest, dass die Umkehrfunktion von ζ durch

$$\zeta^{-1} : \mathbb{R} \rightarrow \mathbb{R}, z \mapsto \zeta^{-1}(z) := \sigma z + \mu \quad (35)$$

gegeben ist, da für alle $z \in \mathbb{R}$ mit $z = \frac{y - \mu}{\sigma}$ gilt, dass

$$\zeta^{-1}(z) = \zeta^{-1}\left(\frac{y - \mu}{\sigma}\right) = \frac{\sigma(y - \mu)}{\sigma} + \mu = y - \mu + \mu = y. \quad (36)$$

Schließlich stellen wir fest, dass für die Ableitung ζ' von ζ gilt, dass

$$\zeta'(y) = \frac{d}{dy} \left(\frac{y - \mu}{\sigma} \right) = \frac{d}{dy} \left(\frac{y}{\sigma} - \frac{\mu}{\sigma} \right) = \frac{1}{\sigma}. \quad (37)$$

Beweis (fortgeführt)

Einsetzen in das univariate WDF Transformationstheorem für lineare Funktionen ergibt dann

$$\begin{aligned} p_{\zeta}(z) &= \frac{1}{|1/\sigma|} N(\sigma z + \mu; \mu, \sigma^2) \\ &= \frac{1}{1/\sqrt{\sigma^2}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (\sigma z + \mu - \mu)^2\right) \\ &= \frac{\sqrt{\sigma^2}}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \sigma^2 z^2\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2\right) \\ &= N(z; 0, 1) \end{aligned} \tag{38}$$

also, dass $Z \sim N(0, 1)$. Z ist also eine z -Zufallsvariable.

Definition (χ^2 -Zufallsvariable)

U sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}_{>0}$ und WDF

$$p : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}, u \mapsto p(u) := \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} u^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}u\right), \quad (39)$$

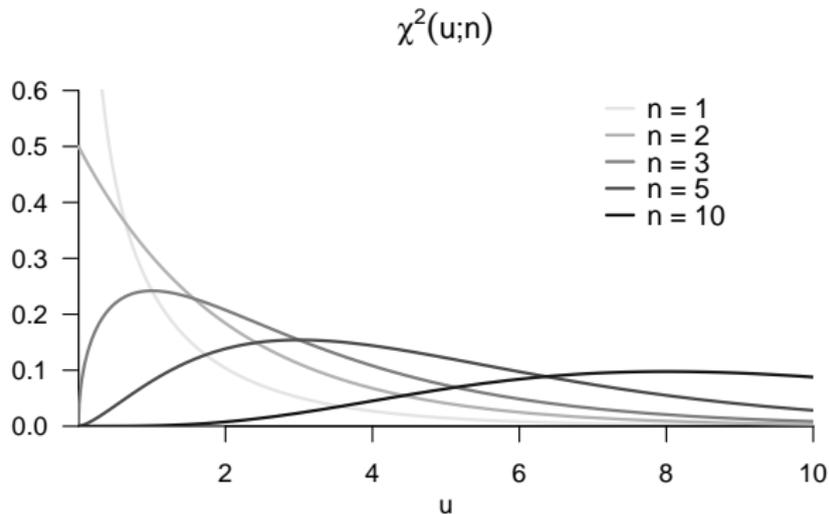
wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass U einer χ^2 -Verteilung mit Freiheitsgradparameter n unterliegt und nennen U eine χ^2 -Zufallsvariable mit Freiheitsgradparameter n . Wir kürzen dies mit $U \sim \chi^2(n)$ ab. Die WDF einer χ^2 -Zufallsvariable bezeichnen wir mit

$$\chi^2(u; n) := \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} u^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}u\right). \quad (40)$$

Bemerkung

- Die WDF der χ^2 -Verteilung entspricht der WDF $G(u; \frac{n}{2}, 2)$ einer Gammaverteilung.

Wahrscheinlichkeitsdichtefunktionen von ξ^2 -Zufallsvariablen



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

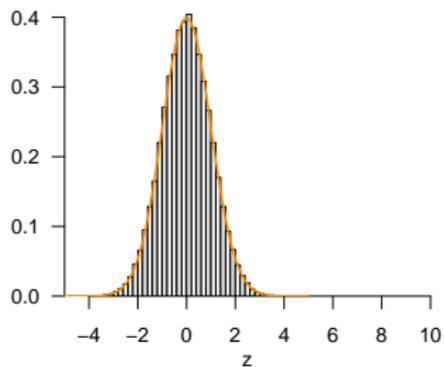
$$U := \sum_{i=1}^n Z_i^2 \quad (41)$$

eine χ^2 -verteilte Zufallsvariable mit Freiheitsgradparameter n , es gilt also $U \sim \chi^2(n)$. Insbesondere gilt für $Z \sim N(0, 1)$ und $U := Z^2$, dass $U \sim \chi^2(1)$.

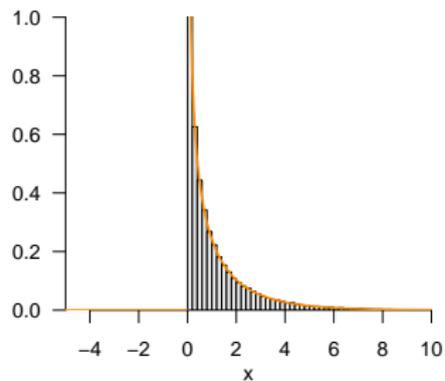
Bemerkungen

- Die U -Konfidenzintervallstatistik ist ein wichtiger Anwendungsfall.
- t - und f -Zufallsvariablen sind wichtige Anwendungsfälle.

$$Z \sim N(0, 1)$$



$$U = Z^2 \sim \chi^2(1)$$



Standardtransformationen

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_ζ durch eine stückweise bijektive Abbildung hervorgeht, gegeben durch

$$p_U(u) = \sum_{i=1}^k 1_{\mathcal{U}_i} \frac{1}{|f'_i(f_i^{-1}(u))|} p_\zeta(f_i^{-1}(u)). \quad (42)$$

Wir definieren

$$\mathcal{U}_1 :=]-\infty, 0[, \mathcal{U}_2 :=]0, \infty[, \text{ und } \mathcal{U}_i := \mathbb{R}_{>0} \text{ für } i = 1, 2, \quad (43)$$

sowie

$$f_i : \mathcal{Z}_i \rightarrow \mathcal{U}_i, x \mapsto f_i(z) := z^2 =: u \text{ für } i = 1, 2. \quad (44)$$

Die Ableitung und die Umkehrfunktion der f_i ergeben sich zu

$$f'_i : \mathcal{Z}_i \rightarrow \mathcal{Z}_i, x \mapsto f'_i(z) = 2z \text{ für } i = 1, 2, \quad (45)$$

und

$$f_1^{-1} : \mathcal{U}_1 \rightarrow \mathcal{U}_1, u \mapsto f_1^{-1}(u) = -\sqrt{u} \text{ und } f_2^{-1} : \mathcal{U}_2 \rightarrow \mathcal{U}_2, u \mapsto f_2^{-1}(u) = \sqrt{u}, \quad (46)$$

respektive.

Beweis (fortgeführt)

Einsetzen in Gleichung (42) ergibt dann

$$\begin{aligned} p_U(u) &= 1_{U_1}(u) \frac{1}{|f_1'(f_1^{-1}(u))|} p_\zeta(f_1^{-1}(u)) + 1_{U_2}(u) \frac{1}{|f_2'(f_2^{-1}(u))|} p_\zeta(f_2^{-1}(u)) \\ &= \frac{1}{|2(-\sqrt{u})|} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(-\sqrt{u})^2\right) + \frac{1}{|2(\sqrt{u})|} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\sqrt{u})^2\right) \\ &= \frac{1}{2\sqrt{u}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u\right) + \frac{1}{2\sqrt{u}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u\right) \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{u}} \exp\left(-\frac{1}{2}u\right). \end{aligned} \tag{47}$$

Andererseits gilt, dass mit $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, die PDF einer χ^2 -Zufallsvariable U mit $n = 1$ durch

$$\frac{1}{\Gamma\left(\frac{1}{2}\right) 2^{\frac{1}{2}}} u^{\frac{1}{2}-1} \exp\left(-\frac{1}{2}u\right) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{u}} \exp\left(-\frac{1}{2}u\right) \tag{48}$$

gegeben ist. Also gilt, dass wenn $Z \sim N(0, 1)$ ist, dann ist $U := Z^2 \sim \chi^2(1)$.

Definition (t -Zufallsvariable)

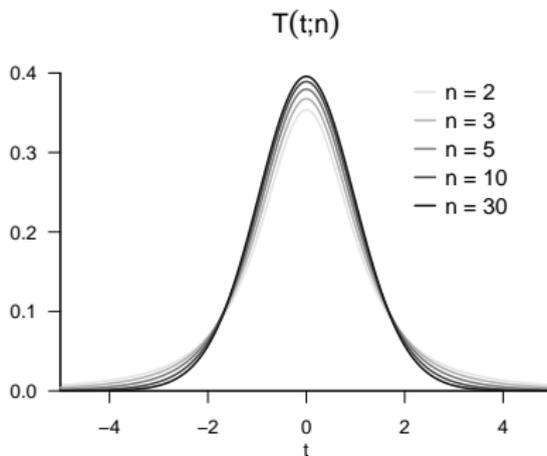
T sei eine Zufallsvariable mit Ergebnisraum \mathbb{R} und WDF

$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, t \mapsto p(t) := \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad (49)$$

wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass T einer t -Verteilung mit Freiheitsgradparameter n unterliegt und nennen T eine t -Zufallsvariable mit Freiheitsgradparameter n . Wir kürzen dies mit $T \sim t(n)$ ab. Die WDF einer t -Zufallsvariable bezeichnen wir mit

$$T(t; n) := \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}. \quad (50)$$

Wahrscheinlichkeitsdichtefunktionen von t -Zufallsvariablen



- 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 Freiheitsgradparameter n , und Z und U seien unabhängige Zufallsvariablen. Dann ist die Zufallsvariable

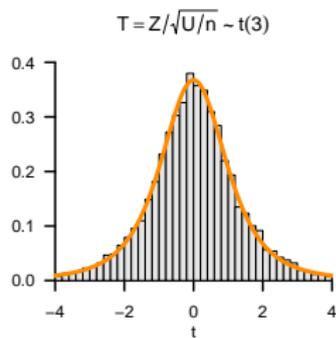
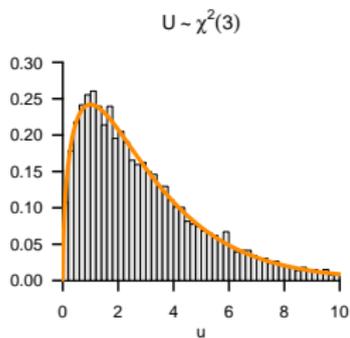
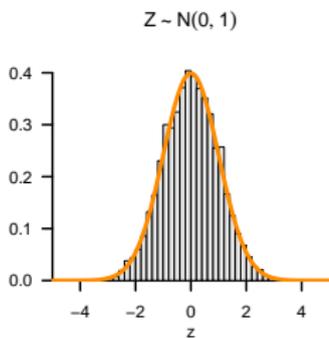
$$T := \frac{Z}{\sqrt{U/n}} \quad (51)$$

eine t -verteilte Zufallsvariable mit Freiheitsgradparameter n , es gilt also $T \sim t(n)$.

Bemerkungen

- Das Theorem geht auf Student (1908) zurück.
- Das Theorem ist das zentrale Resultat der Frequentistischen Statistik.
- Zabell (2008) gibt hierzu einen historischen Überblick.
- Die T -Konfidenzintervallstatistik und die T -Teststatistik sind wichtige Anwendungsfälle.

Standardtransformationen



Standardtransformationen

Beweis

Wir halten zunächst fest, dass die zweidimensionale WDF der gemeinsamen (unabhängigen) Verteilung von Z und U durch

$$p_{Z,U}(z, u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} u^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}u\right). \quad (52)$$

gegeben ist. Wir betrachten dann die multivariate vektorwertige Abbildung

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}^2, (z, u) \mapsto f(z, u) := \left(\frac{z}{\sqrt{u/n}}, u \right) =: (t, w) \quad (53)$$

und benutzen das multivariate WDF Transformationstheorem für bijektive Abbildungen um die WDF von (t, w) herzuleiten. Dazu erinnern wir uns, dass wenn ξ ein n -dimensionaler Zufallsvektor mit WDF p_ξ und $v := f(\xi)$ für eine differenzierbare und bijektive Abbildung $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ist, die WDF des Zufallsvektors v durch

$$p_v : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}, y \mapsto p_v(y) := \frac{1}{|Jf(f^{-1}(y))|} p_\xi(f^{-1}(y)) \quad (54)$$

gegeben ist. Für die im vorliegenden Fall betrachtete Abbildung halten wir zunächst fest, dass

$$f^{-1} : \mathbb{R}^2 \rightarrow \mathbb{R}^2, (t, w) \mapsto f^{-1}(t, w) := (\sqrt{w/nt}, w). \quad (55)$$

T-Transformation

Beweis (fortgeführt)

Dies ergibt sich direkt aus

$$f^{-1}(f(z, u)) = f^{-1}\left(\frac{z}{\sqrt{u/n}}, u\right) = \left(\frac{\sqrt{u/n}z}{\sqrt{u/n}}, u\right) = (z, u) \text{ für alle } (z, u) \in \mathbb{R}^2. \quad (56)$$

Wir halten dann fest, dass die Determinante der Jacobi-Matrix von f an der Stelle (z, u) durch

$$|Jf(z, u)| = \begin{vmatrix} \frac{\partial}{\partial z} \left(\frac{z}{\sqrt{u/n}}\right) & \frac{\partial}{\partial u} \left(\frac{z}{\sqrt{u/n}}\right) \\ \frac{\partial}{\partial z} u & \frac{\partial}{\partial u} u \end{vmatrix} = \left(\frac{v}{n}\right)^{-1/2}, \quad (57)$$

gegeben ist, sodass folgt, dass

$$\frac{1}{|Jf(f^{-1}(z, u))|} = \left(\frac{w}{n}\right)^{1/2}. \quad (58)$$

Einsetzen in Gleichung (54) ergibt dann

$$p_{T,W}(t, w) = \left(\frac{w}{n}\right)^{1/2} p_{Z,V}(\sqrt{w/nt}, w), \quad (59)$$

Standardtransformationenen

Beweis (fortgeführt)

Es folgt also

$$\begin{aligned} p_T(t) &= \int_0^{\infty} p_{T,W}(t, w) dw \\ &= \int_0^{\infty} \left(\frac{w}{n}\right)^{1/2} p_{Z,V}(\sqrt{w/nt}, w) dw \\ &= \int_0^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\sqrt{w/nt})^2\right) \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} w^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}w\right) \left(\frac{w}{n}\right)^{1/2} dw \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}} n^{\frac{1}{2}}} \int_0^{\infty} \exp\left(-\frac{1}{2}\frac{w}{n}t^2\right) w^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}w\right) w^{1/2} dw \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}} n^{\frac{1}{2}}} \int_0^{\infty} \exp\left(-\frac{1}{2}\frac{w}{n}t^2 - \frac{1}{2}w\right) w^{\frac{n}{2}-1} w^{\frac{1}{2}} dw \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}} n^{\frac{1}{2}}} \int_0^{\infty} \exp\left(-\frac{1}{2}\left(\frac{w}{n}t^2 + w\right)\right) w^{\frac{n+1}{2}-1} dw \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}} n^{\frac{1}{2}}} \int_0^{\infty} \exp\left(-\frac{1}{2}\left(1 + \frac{t^2}{n}\right)\right) w^{\frac{n+1}{2}-1} dw \end{aligned} \tag{60}$$

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

$$p_T(t) = \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}} n^{\frac{1}{2}}} \int_0^{\infty} \Gamma\left(w; \frac{n+1}{2}, \frac{2}{1+\frac{t^2}{n}}\right) dw. \quad (61)$$

Schließlich stellen wir fest, dass der Integralterm in obiger Gleichung dem Normalisierungsterm einer Gamma WDF entspricht. Abschließend ergibt sich also

$$p_T(t) = \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}} n^{\frac{1}{2}}} \Gamma\left(\frac{n+1}{2}\right) \left(\frac{2}{1+\frac{t^2}{n}}\right)^{\frac{n+1}{2}}. \quad (62)$$

Die Verteilung von $Z/\sqrt{U/n}$ hat also die WDF einer T -Zufallsvariable.

□

Definition (f -Zufallsvariable)

F sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}_{>0}$ und WDF

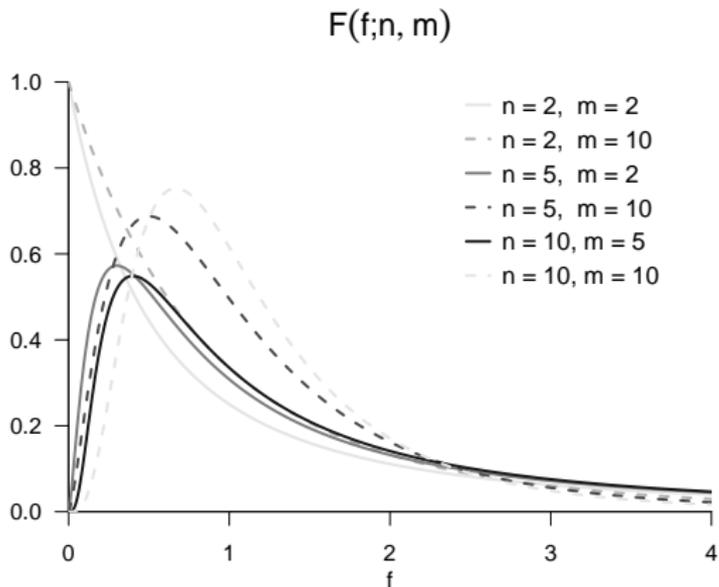
$$p_F : \mathbb{R} \rightarrow \mathbb{R}_{>0}, f \mapsto p_F(f) := m \frac{m}{2} n \frac{n}{2} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{f^{\frac{m}{2}-1}}{\left(1 + \frac{m}{n} f\right)^{\frac{m+n}{2}}}, \quad (63)$$

wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass F einer f -Verteilung mit Freiheitsgradparametern n, 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 bezeichnen wir mit

$$F(f; n, m) := m \frac{m}{2} n \frac{n}{2} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{f^{\frac{m}{2}-1}}{\left(1 + \frac{m}{n} f\right)^{\frac{m+n}{2}}}. \quad (64)$$

Standardtransformationen

Wahrscheinlichkeitsdichtefunktionen von f -Zufallsvariablen



Theorem (F -Transformation)

$V \sim \chi^2(n)$ und $W \sim \chi^2(m)$ seien zwei unabhängige χ^2 -Zufallsvariablen mit Freiheitsgradparametern n und m , respektive. Dann ist die Zufallsvariable

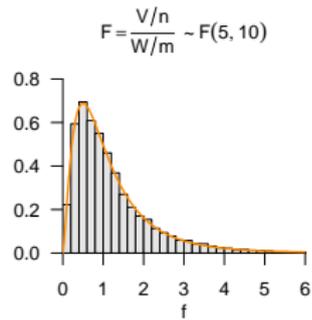
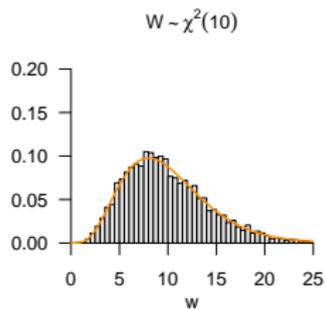
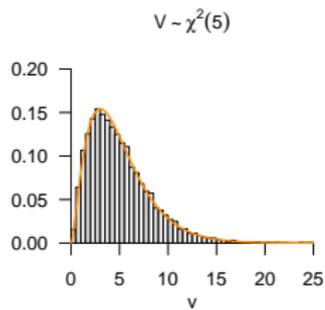
$$F := \frac{V/n}{W/m} \quad (65)$$

eine f -verteilte Zufallsvariable mit Freiheitsgradparametern n, m , es gilt also $F \sim f(n, m)$.

Bemerkungen

- Wir verzichten auf einen Beweis
- Das Theorem kann bewiesen werden, in dem man zunächst ein Transformationstheorem 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.

Standardtransformationen



Vorbemerkungen

Transformationstheoreme

Standardtransformationen

Selbstkontrollfragen

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 Frequentistische Inferenz.
4. Geben Sie das Theorem zur Summe unabhängig normalverteilter Zufallsvariablen wieder.
5. Geben Sie das Theorem zum Stichprobenmittel von u.i.v. normalverteilten Zufallsvariablen 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 ihres Freiheitsgradparameters.
9. Geben Sie das T -Transformationstheorem wieder.
10. Geben Sie das F -Transformationstheorem wieder.

Selbstkontrollfragen - Antworten

1. Mit einer Transformation ist hier die Anwendung einer Funktion auf eine Zufallsvariable oder die Verknüpfung einer Zufallsvariable mit einer anderen Zufallsvariable gemeint.
2. Die zentrale Idee der Transformationstheoreme ist es, generelle Werkzeuge bereitzustellen, die es ermöglichen, basierend auf der WDF einer kontinuierlichen Zufallsvariable und der Form ihrer Transformation die WDF, also die Verteilung, der entsprechend transformierten Zufallsvariable zu berechnen.
3. Die Standardtransformationen sind in der Frequentistischen Inferenz zentral, da viele der gebräuchlichsten Modelle von normalverteilten Fehlervariablen ausgehen, die im Rahmen der Modelle und durch auf ihnen basierenden Statistiken im Sinne der Standardtransformationen transformiert werden. Wissen um die Resultate der Standardtransformationen spiegelt entspricht dann dem Wissen um die Verteilungen von Datenvariablen und Statistiken in der Frequentistischen Inferenz.
4. Siehe Theorem (Summe unabhängig normalverteilter Zufallsvariablen).
5. Siehe Theorem (Stichprobenmittel von u.i.v. normalverteilten Zufallsvariablen).
6. Siehe Theorem (Z -Transformation).
7. Siehe Theorem (χ^2 -Transformation).
8. Unabhängig von ihrem Freiheitsgradparameter ist die WDF der t -Verteilung um 0 symmetrisch und ähnelt der WDF der Standardnormalverteilung. Bei geringem Freiheitsgradparameter besitzt die WDF der t -Verteilung in ihren Ausläufern eine höhere Wahrscheinlichkeitsdichte als die Standardnormalverteilung, für steigenden Freiheitsgradparameter nähert sich die WDF der t -Verteilung der WDF der Standardnormalverteilung an und ist ab etwa $n = 30$ mit ihr im Wesentlichen identisch.
9. Siehe Theorem (T -Transformation).
10. Siehe Theorem (F -Transformation).

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. <https://doi.org/10.1198/016214508000000030>.



Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2023/24

Prof. Dr. Dirk Ostwald

(8) Grundbegriffe Frequentistischer Inferenz

Frequentistische Inferenzmodelle

Statistiken und Schätzer

Standardproblemstellungen und Standardannahme

Selbstkontrollfragen

Frequentistische Inferenzmodelle

Statistiken und Schätzer

Standardproblemstellungen und Standardannahme

Selbstkontrollfragen

Definition (Frequentistisches Inferenzmodell)

Ein *Frequentistisches Inferenzmodell* ist ein Tripel

$$\mathcal{M} := (\mathcal{Y}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\}) \quad (1)$$

bestehend aus einem *Datenraum* \mathcal{Y} , einer σ -Algebra \mathcal{A} auf \mathcal{Y} und einer mindestens zweielementigen Menge $\{\mathbb{P}_\theta | \theta \in \Theta\}$ von Wahrscheinlichkeitsmaßen auf $(\mathcal{Y}, \mathcal{A})$, die durch $\theta \in \Theta$ indiziert sind. Wenn $\Theta \subset \mathbb{R}^k$ ist, heißt ein Frequentistisches Inferenzmodell auch *parametrisches* Frequentistisches Inferenzmodell und Θ heißt *Parameterraum* des Frequentistischen Inferenzmodells.

Bemerkungen

- Für Erwartungswerte und (Ko)Varianzen bezüglich \mathbb{P}_θ schreiben wir $\mathbb{E}_\theta, \mathbb{V}_\theta, \mathbb{C}_\theta$.
- Ein Frequentistisches Inferenzmodell \mathcal{M} heißt ein *diskretes Modell*, wenn \mathcal{Y} endlich oder abzählbar ist und jedes \mathbb{P}_θ eine WMF p_θ besitzt, ein Frequentistisches Inferenzmodell \mathcal{M} heißt ein *stetiges Modell*, wenn $\mathcal{Y} \subseteq \mathbb{R}^m$ ist und jedes \mathbb{P}_θ eine WDF p_θ besitzt.
- Für ein Frequentistisches Inferenzmodell $\mathcal{M}_0 := (\mathcal{Y}_0, \mathcal{A}_0, \{\mathbb{P}_\theta^0 | \theta \in \Theta\})$ heißt das Modell $\mathcal{M} := (\mathcal{Y}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\})$, für das \mathcal{Y} das n -fache kartesische Produkt von \mathcal{Y}_0 mit sich selbst, \mathcal{A} die entsprechende Produkt- σ -Algebra ist, und $\{\mathbb{P}_\theta | \theta \in \Theta\}$ die entsprechende Menge an Produktmaßen ist, das zu \mathcal{M}_0 gehörige *Produktmodell*.
- Wenn für ein Produktmodell die Menge \mathcal{Y}_0 eindimensional ist, also z.B. $\mathcal{Y}_0 = \mathbb{R}$ gilt, spricht man von einem *univariaten Frequentistischen Inferenzmodell*. Wenn für ein Produktmodell die Menge \mathcal{Y}_0 mehrdimensional ist, also z.B. $\mathcal{Y}_0 = \mathbb{R}^m, m > 1$ ist, spricht man von einem *multivariaten Frequentistischen Inferenzmodell*.

Bemerkungen (fortgeführt)

- Der Vorgang der Datenbeobachtung wird durch einen Zufallsvektor v , der Werte in \mathcal{Y} annimmt, beschrieben. Im Kontext von Inferenzmodellen nennt man diesen Zufallsvektor *Daten*, *Beobachtung*, *Messung* oder *Stichprobe*. Eine Realisierung von v , also konkret vorliegende Datenwerte $y \in \mathcal{Y}$, werden *Datensatz*, *Beobachtungswert*, *Messwert* oder *Stichprobenwert* genannt.
- Produktmodelle modellieren die n -fache unabhängige Wiederholung eines Zufallsvorgangs. Der entsprechende Zufallsvektor $v := (v_1, \dots, v_n)$ entspricht dann einer Menge von n u.i.v. Zufallsvariablen.
- Im Gegensatz zum Wahrscheinlichkeitsraummodell betrachtet man bei Frequentistische Inferenzmodellen zwei oder mehr Wahrscheinlichkeitsmaße, die die Verteilung von v mutmaßlich bestimmen. Das jeweils zugrundeliegende Wahrscheinlichkeitsmaß ist mit $\theta \in \Theta$ indiziert,
- In einem konkreten Datenanalyseproblem nimmt man an, dass die beobachteten Werte $y = (y_1, \dots, y_n)$ von $v = (v_1, \dots, v_n)$ durch θ^* generiert wurde, wobei θ^* hier den *wahren, aber unbekanntem, Parameterwert* bezeichnet. Der wahre, aber unbekanntem, Parameterwert θ^* bleibt auch nach der Datenanalyse unbekannt. In der mathematischen Analyse von Inferenzmethoden betrachtet man alle möglichen wahren, aber unbekanntem, Parameterwerte, schreibt also einfach $\{\mathbb{P}_\theta | \theta \in \Theta\}$.

Definition (Normalverteilungsmodell)

Das univariate parametrische Produktmodell

$$\mathcal{M} := (Y, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\}) \quad (2)$$

mit

$$Y := \mathbb{R}^n, \mathcal{A} := \mathcal{B}(\mathbb{R}^n), \theta := (\mu, \sigma^2), \Theta := \mathbb{R} \times \mathbb{R}_{>0}, \quad (3)$$

also

$$\{\mathbb{P}_\theta | \theta \in \Theta\} := \left\{ \prod_{i=1}^n N(\mu, \sigma^2) | (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0} \right\}, \quad (4)$$

und damit

$$v_1, \dots, v_n \sim N(\mu, \sigma^2) \text{ mit } (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0} \quad (5)$$

heißt *Normalverteilungsmodell*.

Bemerkungen

- Das Normalverteilungsmodell ist Grundlage vieler populärer Inferenzverfahren.
- Diese Verfahren werden im Allgemeinen Linearen Modell integrativ betrachtet.
- Das Modell ist grundlegend durch normalverteilte Fehlerterme motiviert.

Definition (Bernoullimodell)

Das univariate parametrische Produktmodell

$$\mathcal{M} := (\mathcal{Y}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\}) \quad (6)$$

mit

$$\mathcal{Y} := \{0, 1\}^n, \mathcal{A} := \mathcal{P}(\{0, 1\}^n), \theta := \mu, \Theta :=]0, 1[\quad (7)$$

also

$$\{\mathbb{P}_\theta | \theta \in \Theta\} := \left\{ \prod_{i=1}^n \text{Bern}(\mu) | \mu \in]0, 1[\right\}, \quad (8)$$

und damit

$$v_1, \dots, v_n \sim \text{Bern}(\mu) \text{ mit } \mu \in]0, 1[\quad (9)$$

heißt *Bernoullimodell*.

Bemerkung

- Das Bernoullimodell spielt in der Anwendung eine eher untergeordnete Rolle.

Frequentistische Inferenzmodelle

Statistiken und Schätzer

Standardproblemstellungen und Standardannahme

Selbstkontrollfragen

Definition (Statistik)

\mathcal{M} sei ein Frequentistisches Inferenzmodell und (Σ, \mathcal{S}) sei ein Messraum. Dann wird eine Zufallsvariable der Form

$$S : \mathcal{Y} \rightarrow \Sigma \tag{10}$$

Statistik genannt.

Bemerkungen

- Daten und Statistiken werden durch Zufallsvariablen modelliert. Statistiken modellieren dabei von Datenwissenschaftler:innen konstruierte Funktionen, die bestenfalls datenbasierte Information liefern, aus der sich Schlüsse über die latenten datengenerierenden Prozesse ziehen lassen.

Beispiele (Statistiken)

\mathcal{M} sei das Normalverteilungsmodell. Dann sind zum Beispiel folgende Zufallsvariablen Statistiken, was wir hier explizit durch die Notation deutlich machen wollen, was aber oft zur Vereinfachung der Notation implizit (aber trotzdem wichtig) bleibt:

- Das *Stichprobenmittel*

$$\bar{y} : \mathbb{R}^n \rightarrow \mathbb{R}, y \mapsto \bar{y}(y) := \frac{1}{n} \sum_{i=1}^n y_i, \quad (11)$$

- Die *Stichprobenvarianz*

$$s^2 : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}, y \mapsto s^2(y) := \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y}(y))^2, \quad (12)$$

- Die *Stichprobenstandardabweichung*

$$s : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}, y \mapsto s(y) := \sqrt{s^2(y)}, \quad (13)$$

Definition (Schätzer)

\mathcal{M} sei ein Frequentistisches Inferenzmodell, (Σ, \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. Dann heißt eine Statistik

$$\hat{\tau} : \mathcal{Y} \rightarrow \Sigma \quad (14)$$

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 ,
 - $\tau(\theta) := \mathbb{E}_\theta(y_1)$ für die Schätzung des Erwartungswert,
 - $\tau(\theta) := \mathbb{V}_\theta(y_1)$ für die Schätzung der Varianz.
- Für $\hat{\tau}$ bei $\tau(\theta) := \theta$ schreibt man üblicherweise $\hat{\theta}$.
- Schätzer nehmen Zahlwerte in Σ an und heißen deshalb auch *Punktschätzer*.
- Nicht jeder Schätzer ist ein guter Schätzer, man definiert deshalb *Schätzgütekriterien*.
- Gütekriterien für Schätzer sind der Inhalt von Vorlesungseinheit (9) Punktschätzung.

Beispiele (Schätzer)

\mathcal{M} sei das Normalverteilungsmodell.

- Dann ist zum Beispiel das Stichprobenmittel $\bar{y} : \mathbb{R}^n \rightarrow \mathbb{R}$ ein Schätzer für

$$\tau : \mathbb{R} \times \mathbb{R}_{>0} \rightarrow \mathbb{R}, (\mu, \sigma^2) \mapsto \tau(\mu, \sigma^2) := \mu. \quad (15)$$

Ebenso ist \bar{y} ein Schätzer für

$$\tau : \mathbb{R} \times \mathbb{R}_{>0} \rightarrow \mathbb{R}, (\mu, \sigma^2) \mapsto \tau(\mu, \sigma^2) := \mathbb{E}_{\mu, \sigma^2}(y_1). \quad (16)$$

- Weiterhin ist die konstante Funktion

$$\hat{\tau} : \mathbb{R}^n \rightarrow \mathbb{R}, y \mapsto \hat{\tau}(y) := 42 \quad (17)$$

ein Schätzer für

$$\tau : \mathbb{R} \times \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}, (\mu, \sigma^2) \mapsto \tau(\mu, \sigma^2) := \sigma^2. \quad (18)$$

Dass eine Funktion $\hat{\tau} : \mathcal{Y} \rightarrow \Sigma$ ein Schätzer ist, heißt nicht, dass sie ein guter Schätzer ist!

Frequentistische Inferenzmodelle

Statistiken und Schätzer

Standardproblemstellungen und Standardannahme

Selbstkontrollfragen

Mithilfe von Modellen behandelt die Frequentistische Inferenz folgende Standardproblemstellungen:

(1) Punktschätzung

Ziel der Punktschätzung ist es, einen möglichst guten Tipp für wahre, aber unbekannte, Parameterwerte oder Funktionen dieser abzugeben, typischerweise mithilfe der Daten.

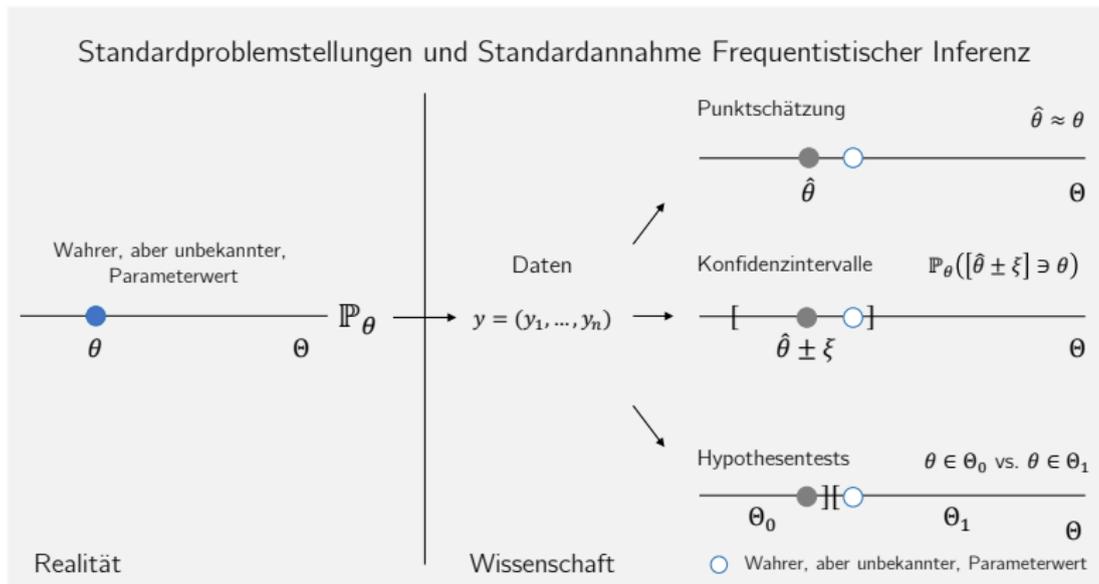
(2) Konfidenzintervalle

Ziel der Konfidenzintervallbestimmung ist es, basierend auf der angenommenen Datenverteilung und den beobachteten Daten durch eine Intervallschätzung einen möglichst sicheren, wenn auch oft unpräzisen, Tipp für wahre, aber unbekannte, Parameterwerte abzugeben.

(3) Hypothesentests

Ziel des Hypothesentestens ist es, basierend auf der angenommenen Verteilung der Daten in einer möglichst zuverlässigen Form zu entscheiden, ob ein wahrer, aber unbekannter, Parameterwert in einer von zwei sich gegenseitig ausschließenden Untermengen des Parameterraumes liegt.

Standardproblemstellungen und Standardannahme



Standardannahme Frequentistischer Inferenz

\mathcal{M} sei ein Frequentistisches Inferenzmodell mit Zufallsvektor v . Es wird angenommen, dass ein Datensatz $y \in \mathbb{R}^n$ eine der möglichen Realisierungen von v ist.

Aus Frequentistischer Sicht kann man eine Studie unendlich oft wiederholen und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. das Stichprobenmittel:

$$\text{Datensatz (1)} : y^{(1)} = \left(y_1^{(1)}, y_2^{(1)}, \dots, y_n^{(1)} \right) \text{ mit } \bar{y}^{(1)} = \frac{1}{n} \sum_{i=1}^n y_i^{(1)}$$

$$\text{Datensatz (2)} : y^{(2)} = \left(y_1^{(2)}, y_2^{(2)}, \dots, y_n^{(2)} \right) \text{ mit } \bar{y}^{(2)} = \frac{1}{n} \sum_{i=1}^n y_i^{(2)}$$

$$\text{Datensatz (3)} : y^{(3)} = \left(y_1^{(3)}, y_2^{(3)}, \dots, y_n^{(3)} \right) \text{ mit } \bar{y}^{(3)} = \frac{1}{n} \sum_{i=1}^n y_i^{(3)}$$

$$\text{Datensatz (4)} : y^{(4)} = \left(y_1^{(4)}, y_2^{(4)}, \dots, y_n^{(4)} \right) \text{ mit } \bar{y}^{(4)} = \frac{1}{n} \sum_{i=1}^n y_i^{(4)}$$

$$\text{Datensatz (5)} : y^{(5)} = \dots$$

Um die Qualität ihrer Methoden zu beurteilen betrachtet die Frequentistische Inferenz deshalb die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken. Was zum Beispiel ist die Verteilung der $\bar{y}^{(1)}, \bar{y}^{(2)}, \bar{y}^{(3)}, \bar{y}^{(4)}, \dots$ also die Verteilung der Zufallsvariable \bar{v}_n ?

Wenn eine Methode im Sinne der Frequentistischen Standardannahme "gut" ist, dann heißt das also, dass sie bei häufiger Anwendung "im Mittel gut" ist. Im Einzelfall, also im Normalfall nur eines vorliegenden Datensatzes, kann sie auch "schlecht" sein.

Standardproblemstellungen und Standardannahme

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Pre-BDI



n = 12

Post-BDI

⇒ Pre-Post BDI Score Reduktion

BDI-II		Fragebogen	
Name		Alter	Erkrankt
		W/M	JA/NEIN
Anleitung: Dieser Fragebogen enthält 23 Gruppen von Aussagen. Bitte lesen Sie jede Aussage sorgfältig von Anfang bis Ende und wählen 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 für Sie am besten passt. Sie können sich für eine Aussage entscheiden, die mehrere Aspekte und für mehrere Aussagen für die Aussage mit der höchsten Zahl an. Achten Sie bitte darauf, dass Sie in jeder Gruppe nicht mehr als eine Aussage ankreuzen, das gilt auch für Gruppen für Verdächtigungen der Selbstwahrnehmung oder Gruppen für Verdächtigungen des Systems.			
1.) Traurigkeit		6.) Berufungsgefühle	
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.
3.) pessimistische		7.) Selbstwahrnehmung	
0 Ich sehe nicht trüblich in die Zukunft.	1 Ich sehe mühsam in die Zukunft als blass.	2 Ich bin müde und anstrengend nicht, dass Dinge glücklicher werden.	3 Ich glaube, dass meine Zukunft hoffnungslos ist und nur noch schlechter wird.
3.) Vorwiegend		8.) Selbstwahrnehmung	
0 Ich fühle mich nicht als Versager.	1 Ich habe häufiger Vorwiegendgefühle.	2 Wenn ich zurückblicke, sehe ich eine Menge Misserfolge.	3 Ich fühle das Gefühl, ein Mensch ein völlig Versager zu sein.
4.) Verlust von Freude		9.) Selbstwahrnehmung	
0 Ich kann die Dinge genauso gut genießen wie früher.	1 Ich kann die Dinge nicht mehr so genießen wie früher.	2 Ich würde es nicht tun.	3 Ich würde mich nicht anstrengen, wenn ich die Möglichkeit dazu hätte.
5.) Schuldgefühle		10.) Meinen	
0 Ich habe keine besonderen Schuldgefühle.	1 Ich habe oft Schuldgefühle wegen Dingen, die ich getan habe oder habe tun sollen.	2 Ich habe die meiste Zeit Schuldgefühle.	3 Ich habe ständig Schuldgefühle.
		11.) Meinen	
		0 Ich denke nicht daran, mir etwas zu tun.	
		1 Ich denke manchmal an Selbstmord, aber ich würde es nicht tun.	
		2 Ich würde mich an Selbstmord überlegen, wenn ich die Gelegenheit dazu hätte.	
		3 Ich würde mich selbst verletzen, wenn ich die Gelegenheit dazu hätte.	
		12.) Meinen	
		0 Ich würde nicht über ein Problem nachdenken.	
		1 Ich würde gerne wissen, was ich kann.	
		2 Ich würde gerne wissen, was ich kann.	
		3 Ich würde gerne wissen, was ich kann.	

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

i	dBDI
1	-1
2	3
3	-2
4	9
5	3
6	-2
7	4
8	5
9	5
10	1
11	9
12	4

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Für die Pre-Post BDI Score Reduktion y_i der i ten von n Patient:innen legen wir das Modell

$$v_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, \dots, n \quad (19)$$

zugrunde. Dabei wird die Pre-Post BDI Reduktion y_i der i ten Patient:in also mithilfe einer über die Gruppe von Patient:innen identischen Pre-Post BDI Score Reduktion $\mu \in \mathbb{R}$ und einer Patient:innen-spezifischen normalverteilten Pre-Post BDI Score Reduktionsabweichung ε_i erklärt.

Wie unten gezeigt ist dieses Modell äquivalent zum oben eingeführten Normalverteilungsmodell

$$v_1, \dots, v_n \sim N(\mu, \sigma^2). \quad (20)$$

Die Standardproblemstellungen der Frequentistischen Inferenz führen in diesem Szenario auf folgende Fragen:

- (1) Was sind sinnvolle Tipps für die wahren, aber unbekanntenen, 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 gilt $\mu \neq 0$?

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Die Äquivalenz beider Modellformen folgt direkt aus der Transformation normalverteilter Zufallsvariablen durch linear-affine Funktionen. Speziell gilt im vorliegenden Fall für $\varepsilon_i \sim N(0, \sigma^2)$ u.i.v. mit $i = 1, \dots, n$, dass

$$v_i = f(\varepsilon_i) \text{ mit } f: \mathbb{R} \rightarrow \mathbb{R}, e_i \mapsto f(e_i) := e_i + \mu. \quad (21)$$

Dann gilt

$$\begin{aligned} p_{v_i}(y_i) &= \frac{1}{|1|} p_{\varepsilon_i} \left(\frac{y_i - \mu}{1} \right) \\ &= N(y_i - \mu; 0, \sigma^2) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{1}{2\sigma^2} (y_i - \mu - 0)^2 \right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{1}{2\sigma^2} (y_i - \mu)^2 \right) \\ &= N(y_i; \mu, \sigma^2), \end{aligned} \quad (22)$$

also $v_i \sim N(\mu, \sigma^2)$ u.i.v. und damit $v_1, \dots, v_n \sim N(\mu, \sigma^2)$.

Frequentistische Inferenzmodelle

Statistiken und Schätzer

Standardproblemstellungen und Standardannahme

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die Definition eines parametrischen Frequentistischen Inferenzmodells wieder.
2. Was ist der Unterschied zwischen univariaten und multivariaten Produktmodell?
3. Geben Sie die Definition des Normalverteilungsmodells wieder.
4. Geben Sie die Definition des Begriffs der Statistik wieder.
5. Geben Sie die Definition des Begriffs des Schätzers wieder.
6. Erläutern Sie die Standardproblemstellungen der Frequentistischen Inferenz.
7. Erläutern Sie die Standardannahme der Frequentistischen Inferenz.

1. Siehe Definition (Frequentistischen Inferenzmodell)
2. Bei einem univariaten Produktmodell ist der Datenraum eindimensional und ein einzelner Datenpunkt ist ein Skalar. Bei einem multivariaten Produktmodell ist der Datenraum mehrdimensional und ein einzelner Datenpunkt ist ein (mehrdimensionaler) Vektor.
3. Siehe Definition (Normalverteilungsmodell).
4. Siehe Definition (Statistik).
5. Siehe Definition (Schätzer).
6. Siehe Folien 15 und 16.
7. Siehe Folie 17.



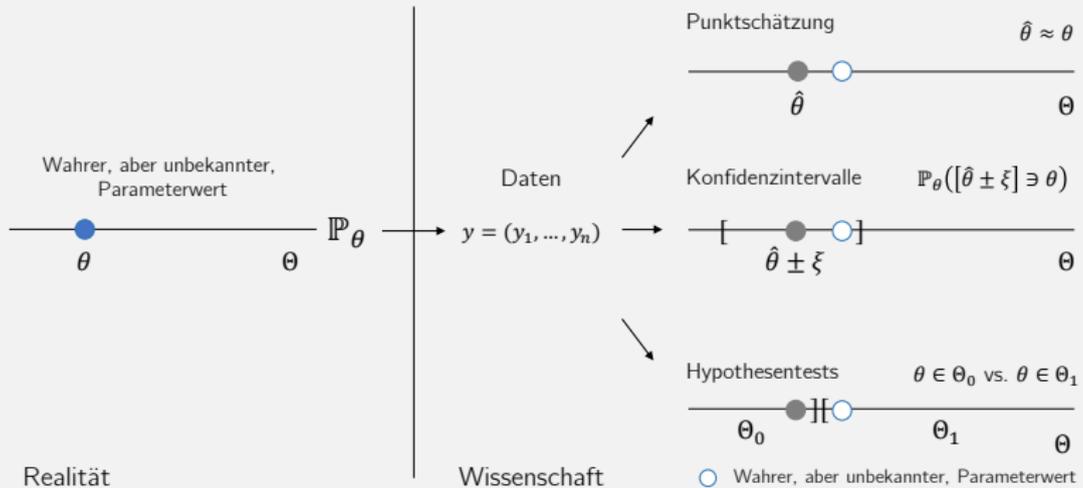
Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2023/24

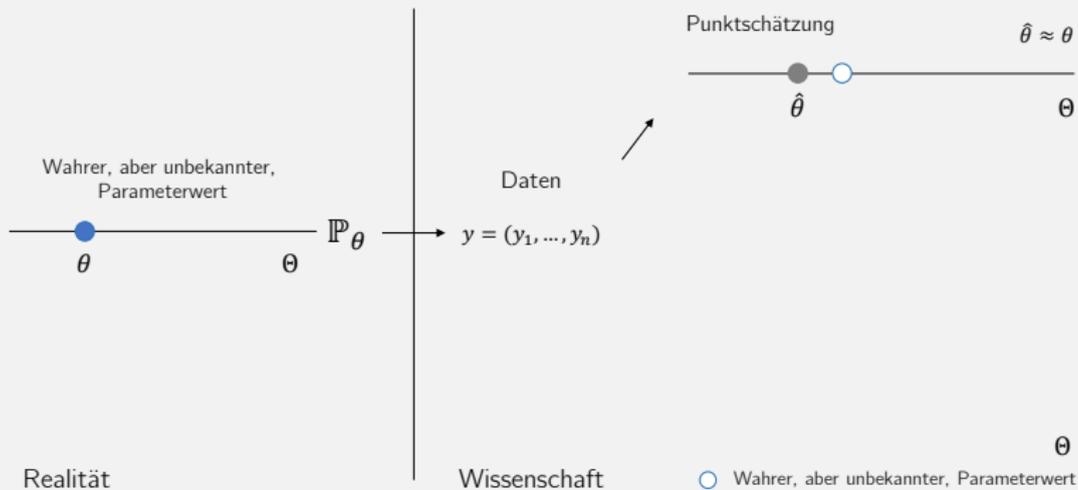
Prof. Dr. Dirk Ostwald

(9) Punktschätzung

Standardproblemstellungen und Standardannahme Frequentistischer Inferenz



Standardproblemstellungen und Standardannahme Frequentistischer Inferenz



Standardannahme Frequentistischer Inferenz

\mathcal{M} sei ein Frequentistisches Inferenzmodell mit Zufallsvektor v . Es wird angenommen, dass ein Datensatz $y \in \mathbb{R}^n$ eine der möglichen Realisierungen von v ist.

Aus Frequentistischer Sicht kann man eine Studie unendlich oft wiederholen und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. das Stichprobenmittel:

$$\text{Datensatz (1)} : y^{(1)} = \left(y_1^{(1)}, y_2^{(1)}, \dots, y_n^{(1)} \right) \text{ mit } \bar{y}^{(1)} = \frac{1}{n} \sum_{i=1}^n y_i^{(1)}$$

$$\text{Datensatz (2)} : y^{(2)} = \left(y_1^{(2)}, y_2^{(2)}, \dots, y_n^{(2)} \right) \text{ mit } \bar{y}^{(2)} = \frac{1}{n} \sum_{i=1}^n y_i^{(2)}$$

$$\text{Datensatz (3)} : y^{(3)} = \left(y_1^{(3)}, y_2^{(3)}, \dots, y_n^{(3)} \right) \text{ mit } \bar{y}^{(3)} = \frac{1}{n} \sum_{i=1}^n y_i^{(3)}$$

$$\text{Datensatz (4)} : y^{(4)} = \left(y_1^{(4)}, y_2^{(4)}, \dots, y_n^{(4)} \right) \text{ mit } \bar{y}^{(4)} = \frac{1}{n} \sum_{i=1}^n y_i^{(4)}$$

$$\text{Datensatz (5)} : y^{(5)} = \dots$$

Um die Qualität ihrer Methoden zu beurteilen betrachtet die Frequentistische Inferenz deshalb die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken. Was zum Beispiel ist die Verteilung der $\bar{y}^{(1)}, \bar{y}^{(2)}, \bar{y}^{(3)}, \bar{y}^{(4)}, \dots$ also die Verteilung der Zufallsvariable \bar{v}_n ?

Wenn eine Methode im Sinne der Frequentistischen Standardannahme "gut" ist, dann heißt das also, dass sie bei häufiger Anwendung "im Mittel gut" ist. Im Einzelfall, also im Normalfall nur eines vorliegenden Datensatzes, kann sie auch "schlecht" sein.

Maximum-Likelihood-Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood-Schätzern

Selbstkontrollfragen

Definition (Parameterpunktschätzer)

$\mathcal{M} := (\mathcal{Y}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\})$ sei ein Frequentistisches Inferenzmodell, (Θ, \mathcal{S}) sei ein Messraum und $\hat{\theta} : \mathcal{Y} \rightarrow \Theta$ sei eine Abbildung. Dann nennen wir $\hat{\theta}$ einen *Parameterpunktschätzer* für θ .

Bemerkungen

- 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 $\hat{\theta}(y)$ unterschieden.

Prinzipien zur Gewinnung von Parameterschätzern

Die Definition eines Parameterschätzers macht keine Aussage darüber, wie man Parameterschätzer findet. Zur Gewinnung von Parameterschätzern in Frequentistischen Inferenzmodellen haben sich deshalb verschiedene Prinzipien etabliert. Populäre Prinzipien zur Gewinnung von Parameterschätzern sind

- Momentenmethode (\approx est. 1890)
- Maximum-Likelihood Methode (\approx est. 1920)
- M-, Z-, W-Schätzung (\approx est. 1960)

Perse 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 Produktmodell mit WMF oder WDF p_θ . Dann ist die *Likelihood-Funktion* definiert als

$$L : \Theta \rightarrow [0, \infty[, \theta \mapsto L(\theta) := \prod_{i=1}^n p_\theta(y_i) \quad (1)$$

und die *Log-Likelihood-Funktion* ist definiert als

$$\ell : \Theta \rightarrow \mathbb{R}, \theta \mapsto \ell(\theta) := \ln L(\theta). \quad (2)$$

Bemerkungen

- L ist eine Funktion des Parameters eines Frequentistischen Inferenzmodells.
- Werte von L sind die gemeinsamen Wahrscheinlichkeitsmassen bzw. -dichten von Datenwerten y_1, \dots, y_n .
- Generell gibt es keinen Grund anzunehmen, dass L 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 Produktmodell mit Parameter $\theta \in \Theta$. Ein *Maximum-Likelihood-Schätzer* von θ ist definiert als

$$\hat{\theta}^{\text{ML}} : \mathcal{Y} \rightarrow \Theta, y \mapsto \hat{\theta}^{\text{ML}}(y) := \arg \max_{\theta \in \Theta} L(\theta) = \arg \max_{\theta \in \Theta} \ell(\theta). \quad (3)$$

Bemerkungen

- $L(\theta) := \prod_{i=1}^n p_{\theta}(y_i)$ hängt von $y := (y_1, \dots, y_n)$ ab, also hängt auch $\hat{\theta}^{\text{ML}}(y)$ von y ab.
- Weil \ln monoton steigend ist, entspricht eine Maximumstelle von ℓ einer Maximumstelle von L .
- Das Arbeiten mit der Log-Likelihood-Funktion ist oft einfacher als mit der Likelihood-Funktion.
- Multiplikation von L mit einer positiven Konstante, die nicht von θ abhängt, verändert einen Maximum-Likelihood-Schätzer nicht, konstante additive Terme in der Log-Likelihood können also vernachlässigt werden.
- Maximum-Likelihood Schätzung ist ein Optimierungsproblem

Vorgehen zur Gewinnung von Maximum-Likelihood-Schätzern

- (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 analytischen Optimierung in klassischen Beispielen
- Methoden der numerischen Optimierung im Anwendungskontext

Theorem (Maximum-Likelihood-Schätzer des Bernoullimodells)

Es sei $v_1, \dots, v_n \sim \text{Bern}(\mu)$ die Stichprobe des Bernoullimodells. Dann ist

$$\hat{\mu}^{\text{ML}} : \{0, 1\}^n \rightarrow [0, 1], y \mapsto \hat{\mu}^{\text{ML}}(y) := \frac{1}{n} \sum_{i=1}^n y_i \quad (4)$$

ein Maximum-Likelihood-Schätzer von μ .

Bemerkung

- $\hat{\mu}^{\text{ML}}$ ist offenbar mit dem Stichprobenmittel identisch.

Maximum-Likelihood-Schätzer

Beweis

Wir formulieren zunächst die Log-Likelihood-Funktion. Für die Likelihood-Funktion gilt

$$L :]0, 1[\rightarrow]0, 1[, \mu \mapsto L(\mu) := \prod_{i=1}^n \mu^{y_i} (1 - \mu)^{1 - y_i} = \mu^{\sum_{i=1}^n y_i} (1 - \mu)^{n - \sum_{i=1}^n y_i}. \quad (5)$$

Logarithmieren ergibt

$$\ell :]0, 1[\rightarrow \mathbb{R}, \mu \mapsto \ell(\mu) = \ln \mu \sum_{i=1}^n y_i + \ln(1 - \mu) \left(n - \sum_{i=1}^n y_i \right). \quad (6)$$

Wir werten dann die Ableitung der Log-Likelihood-Funktion aus. Es gilt

$$\begin{aligned} \frac{d}{d\mu} \ell(\mu) &= \frac{d}{d\mu} \left(\ln \mu \sum_{i=1}^n y_i + \ln(1 - \mu) \left(n - \sum_{i=1}^n y_i \right) \right) \\ &= \frac{d}{d\mu} \ln \mu \sum_{i=1}^n y_i + \frac{d}{d\mu} \ln(1 - \mu) \left(n - \sum_{i=1}^n y_i \right) \\ &= \frac{1}{\mu} \sum_{i=1}^n y_i - \frac{1}{1 - \mu} \left(n - \sum_{i=1}^n y_i \right). \end{aligned} \quad (7)$$

Nullsetzen ergibt dann folgende *Maximum-Likelihood-Gleichung* als notwendige Bedingung für einen Maximum-Likelihood-Schätzer des Bernoullimodelparameters:

$$\frac{1}{\hat{\mu}^{\text{ML}}} \sum_{i=1}^n y_i - \frac{1}{1 - \hat{\mu}^{\text{ML}}} \left(n - \sum_{i=1}^n y_i \right) = 0. \quad (8)$$

Maximum-Likelihood-Schätzer

Beweis (fortgeführt)

Auflösen der Maximum-Likelihood-Gleichung nach $\hat{\mu}^{\text{ML}}$ ergibt dann

$$\begin{aligned} & \frac{1}{\hat{\mu}^{\text{ML}}} \sum_{i=1}^n y_i - \frac{1}{1 - \hat{\mu}^{\text{ML}}} \left(n - \sum_{i=1}^n y_i \right) = 0 \\ \Leftrightarrow \hat{\mu}^{\text{ML}} (1 - \hat{\mu}^{\text{ML}}) & \left(\frac{1}{\hat{\mu}^{\text{ML}}} \sum_{i=1}^n y_i - \frac{1}{1 - \hat{\mu}^{\text{ML}}} \left(n - \sum_{i=1}^n y_i \right) \right) = 0 \\ \Leftrightarrow \sum_{i=1}^n y_i - \hat{\mu}^{\text{ML}} \sum_{i=1}^n y_i - n \hat{\mu}^{\text{ML}} + \hat{\mu}^{\text{ML}} \sum_{i=1}^n y_i & = 0 \tag{9} \\ \Leftrightarrow n \hat{\mu}^{\text{ML}} & = \sum_{i=1}^n y_i \\ \Leftrightarrow \hat{\mu}^{\text{ML}} & = \frac{1}{n} \sum_{i=1}^n y_i. \end{aligned}$$

$\hat{\mu}^{\text{ML}} = \frac{1}{n} \sum_{i=1}^n y_i$ ist also ein potentieller Maximum-Likelihood-Schätzer von μ . Dies kann durch Betrachten der zweiten Ableitung von ℓ verifiziert werden, worauf hier verzichtet werden soll.

□

Theorem (Maximum-Likelihood-Schätzer des Normalverteilungsmodells)

Es sei $v_1, \dots, v_n \sim N(\mu, \sigma^2)$ die Stichprobe des Normalverteilungsmodells. Dann sind

$$\hat{\mu}^{\text{ML}} : \mathbb{R}^n \rightarrow \mathbb{R}, y \mapsto \hat{\mu}^{\text{ML}}(y) := \frac{1}{n} \sum_{i=1}^n y_i \quad (10)$$

und

$$\hat{\sigma}^{2\text{ML}} : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}, y \mapsto \hat{\sigma}_n^{2\text{ML}}(y) := \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu}^{\text{ML}})^2. \quad (11)$$

Maximum-Likelihood-Schätzer von μ und σ^2 .

Bemerkungen

- $\hat{\mu}^{\text{ML}}$ ist identisch mit dem Stichprobenmittel \bar{v} .
- $\hat{\sigma}^{2\text{ML}}$ ist nicht identisch mit der Stichprobenvarianz S^2 .

Maximum-Likelihood-Schätzer

Beweis

Wir formulieren zunächst die Log-Likelihood-Funktion. Für die Likelihood-Funktion ergibt sich

$$\begin{aligned} L : \mathbb{R} \times \mathbb{R}_{>0} &\rightarrow \mathbb{R}_{>0}, (\mu, \sigma^2) \mapsto L(\mu, \sigma^2) := \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right). \end{aligned} \quad (12)$$

Logarithmieren ergibt dann

$$\ell : \mathbb{R} \times \mathbb{R}_{>0} \rightarrow \mathbb{R}, (\mu, \sigma^2) \mapsto \ell(\mu, \sigma^2) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2. \quad (13)$$

Die Bestimmung der Ableitung der Log-Likelihood-Funktion hinsichtlich μ ergibt dann

$$\frac{d}{d\mu} \ell(\mu, \sigma^2) = -\frac{d}{d\mu} \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 = -\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{d}{d\mu} (y_i - \mu)^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) \quad (14)$$

und Bestimmung der Ableitung der Log-Likelihood-Funktion hinsichtlich σ^2 ergibt

$$\frac{d}{d\sigma^2} \ell(\mu, \sigma^2) = -\frac{n}{2} \frac{d}{d\sigma^2} \ln \sigma^2 - \frac{d}{d\sigma^2} \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mu)^2. \quad (15)$$

Das System der Maximum-Likelihood Gleichungen hat also die Form

$$\sum_{i=1}^n (y_i - \hat{\mu}^{\text{ML}}) = 0 \quad \text{und} \quad -\frac{n}{2\hat{\sigma}^2 \text{ML}} + \frac{1}{2\hat{\sigma}^4 \text{ML}} \sum_{i=1}^n (y_i - \mu)^2 = 0 \quad (16)$$

Maximum-Likelihood-Schätzer

Beweis (fortgeführt)

Lösen des Systems der Maximum-Likelihood Gleichungen ergibt dann

$$\sum_{i=1}^n (y_i - \hat{\mu}^{\text{ML}}) = 0 \Leftrightarrow \sum_{i=1}^n y_i = n\hat{\mu}^{\text{ML}} \Leftrightarrow \hat{\mu}^{\text{ML}} = \frac{1}{n} \sum_{i=1}^n y_i. \quad (17)$$

Also ist

$$\hat{\mu}^{\text{ML}} = \frac{1}{n} \sum_{i=1}^n y_i \quad (18)$$

ein potentieller Maximum-Likelihood-Schätzer von μ . Einsetzen in die zweite Maximum-Likelihood Gleichung ergibt dann

$$\begin{aligned} -\frac{n}{2\hat{\sigma}^2\text{ML}} + \frac{1}{2\hat{\sigma}^4\text{ML}} \sum_{i=1}^n (y_i - \hat{\mu}^{\text{ML}})^2 &= 0 \\ \Leftrightarrow -n\hat{\sigma}^2\text{ML} + \sum_{i=1}^n (y_i - \hat{\mu}^{\text{ML}})^2 &= 0 \\ \Leftrightarrow \hat{\sigma}^2\text{ML} &= \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu}^{\text{ML}})^2. \end{aligned} \quad (19)$$

Also ist

$$\hat{\sigma}^2\text{ML} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu}^{\text{ML}})^2 \quad (20)$$

ein potentieller Maximum-Likelihood-Schätzer von σ^2 . Beide potentiellen Maximum-Likelihood-Schätzer können durch Betrachten der zweiten Ableitung von ℓ verifiziert werden, worauf hier verzichtet werden soll.

Maximum-Likelihood-Schätzer

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression



BDI-II Fragebogen

Bitte nicht etwas bis stark

12 2 3 4

Anleitung: Dieser Fragebogen enthält 21 Gruppen von Aussagen. Bitte lesen Sie alle Aussagen sorgfältig durch und wählen Sie sich dann in jeder Gruppe eine Aussage heraus, die am besten beschreibt, wie Sie sich in den letzten zwei Wochen **eigentlich** gefühlt haben. Beachten Sie, die Zeit vor der Aussage an, die Sie sich herausgewählt haben (1, 2 oder 3). Falls in einer Gruppe mehrere Aussagen gleichschwer sind, können Sie die Aussage mit der höchsten Zahl ankreuzen. Achten Sie bitte darauf, dass Sie in jeder Gruppe nicht mehr als eine Aussage ankreuzen, die gilt für Ihre Gruppe in den vergangenen 14 Tagen (einschließlich des Tages).

1.) Traurigkeit 0 Ich bin nicht traurig. 1 Ich bin ein bisschen traurig. 2 Ich bin ziemlich traurig. 3 Ich bin so traurig, mir geht es gar nicht. 4 Ich bin so traurig, dass ich nicht arbeiten kann.	6.) Besorgungsgefühle 0 Ich habe nicht das Gefühl, mir etwas bedroht zu sein. 1 Ich habe das Gefühl, vielleicht bedroht zu werden. 2 Ich empfinde, bedroht zu werden. 3 Ich habe das Gefühl, bedroht zu sein.
2.) Zukunftsangst 0 Ich sehe nicht mehr in die Zukunft. 1 Ich sehe nicht mehr in die Zukunft als sonst. 2 Ich bin müde und anstrengender, dass meine Situation besser wird. 3 Ich glaube, dass meine Zukunft hoffungslos ist und nur noch schlechter wird.	7.) Selbstabwertung 0 Ich habe von mir genauso viel wie immer. 1 Ich habe Vertrauen in mich verloren. 2 Ich bin von mir enttäuscht. 3 Ich fühle mich wertlos.
3.) Vergesslichkeit 0 Ich fühle mich nicht als Vergesslicher. 1 Ich habe häufiger Vergesslichkeitsfälle. 2 Wenn ich zurückblicke, sehe ich eine Menge Verdrängtes. 3 Ich habe das Gefühl, ein Mensch zu vergessen, den ich nicht vergessen sollte.	8.) Selbstverurteilung 0 Ich kritisiere oder tadle mich nicht mehr so sehr. 1 Ich bin mir gegenüber kritischer als sonst. 2 Ich kritisiere mich für all meine Mängel. 3 Ich gebe mir die Schuld für alles Schleine, 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.) Selbstverleugern 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 ein Schuldgefühl wegen Dingen, die ich getan habe oder hätte tun sollen. 2 Ich habe die meisten Zeit Schuldgefühle. 3 Ich habe ständig Schuldgefühle.	10.) Werten 0 Ich werte nicht mehr als ich bin. 1 Ich werte jetzt mehr als ich bin. 2 Ich werte mich geringere Werten. 3 Ich möchte gern mehr sein, aber ich kann nicht.

PERSON: Gruppe: Datum:

© 1997-2002 Pearson Assessment, Inc. Alle Rechte vorbehalten.

⇒ Pre-Post BDI Score Reduktion

Maximum-Likelihood-Schätzer

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

i	BDI.Reduktion
1	-1
2	3
3	-2
4	9
5	3
6	-2
7	4
8	5
9	5
10	1
11	9
12	4

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Für die Pre-Post BDI Score Reduktion v_i der i ten von n Patient:innen legen wir das Modell

$$v_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, \dots, n \quad (21)$$

zugrunde. Dabei wird die Pre-Post BDI Reduktion v_i der i ten Patient:in also mithilfe einer über die Gruppe von Patient:innen identischen Pre-Post BDI Score Reduktion $\mu \in \mathbb{R}$ und einer Patient:innen-spezifischen normalverteilten Pre-Post BDI Score Reduktionsabweichung ε_i erklärt.

Wie gezeigt ist dieses Modell äquivalent zum Normalverteilungsmodell

$$v_1, \dots, v_n \sim N(\mu, \sigma^2). \quad (22)$$

Die Standardproblemstellungen der Frequentistischen Inferenz führen dann auf folgende Fragen:

- (1) Was sind sinnvolle Tipps für die wahren, aber unbekanntenen, Parameterwerte μ und σ^2 ?
- (2) Wie kann im Sinne einer Intervallschätzung eine möglichst sichere Schätzung von μ gelingen?
- (3) Entscheiden wir uns sinnvollerweise für die Hypothese, dass gilt $\mu \neq 0$?

Maximum-Likelihood-Schätzer

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

```
# Einlesen und Auswahl der Daten
fname = "./9_Daten/9_Punktschätzung.csv"
D      = read.table(fname, sep = ",", header = T)
y      = D$BDI.Reduktion

# ML Schätzung des Erwartungswertparameters
mu_hat = mean(y)           # mean(y) berechnet das Stichprobenmittel
print(mu_hat)             # Ausgabe
```

```
[1] 3.166667
```

```
# ML Schätzung des Varianzparameters
n      = length(y)        # Anzahl der Datenpunkte
sigsqr_hat = ((n-1)/n)*var(y) # var(y) berechnet die Stichprobenvarianz
print(sigsqr_hat)        # Ausgabe
```

```
[1] 12.63889
```

Basierend auf dem Prinzip der Maximum-Likelihood Schätzung sind also

$$\hat{\mu}^{\text{ML}} = 3.17, \text{ und } \hat{\sigma}^2{}^{\text{ML}} = 12.6 \quad (23)$$

sinnvolle Tipps für μ und σ^2 basierend auf den vorliegenden 12 Datenpunkten.

Maximum-Likelihood-Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood-Schätzern

Selbstkontrollfragen

Vorbemerkungen zu Frequentistischen Schätzereigenschaften

Wir gehen von einem parametrischem Produktmodell $\mathcal{M} := \{\mathcal{Y}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\}\}$ mit n -dimensionalen Stichprobenraum (z.B. $\mathcal{Y} := \mathbb{R}^n$), d -dimensionalen Parameteraum $\Theta \subset \mathbb{R}^d$ und gegebener WMF oder WDF p_θ für alle $\theta \in \Theta$ aus. $v := (v_1, \dots, v_n)$ bezeichnet die zu \mathcal{M} gehörende Stichprobe unabhängig und identisch verteilter Zufallsvariablen, es gilt also $v_1 \sim \mathbb{P}_\theta$ und $v_i \sim \mathbb{P}_\theta$ für alle $i = 1, \dots, n$.

Für einen Messraum (Σ, \mathcal{S}) sei $\hat{\tau} : \mathcal{Y} \rightarrow \Sigma$ ein Schätzer von $\tau : \Theta \rightarrow \Sigma$. Wir betrachten Erwartungswerts-, Varianz-, und Standardabweichungsschätzer, also Schätzer für

$$\tau : \Theta \rightarrow \Sigma, \theta \mapsto \tau(\theta) \text{ mit } \tau(\theta) := \mathbb{E}_\theta(v_1), \tau(\theta) := \mathbb{V}_\theta(v_1), \text{ und } \tau(\theta) := \mathbb{S}_\theta(v_1) \quad (24)$$

respektive, sowie Parameterschätzer, also Schätzer für

$$\tau : \Theta \rightarrow \Sigma, \tau(\theta) := \theta. \quad (25)$$

In der Folge führen wir *Frequentistische Schätzereigenschaften* ein. Frequentistische Schätzereigenschaften betrachten die Verteilung der Schätzwerte in Abhängigkeit von der Verteilung der Stichprobenwerte. Weil die Stichprobenwerte zufällig sind, sind auch die Schätzwerte zufällig; ein Schätzer $\hat{\tau}$ ist also immer eine Zufallsvariable. Wir unterscheiden zwischen *Schätzereigenschaften bei endlichen Stichproben*, das heißt Eigenschaften von $\hat{\tau}$ für ein fixes $n \in \mathbb{N}$ und *Asymptotischen Schätzereigenschaften*, das heißt Eigenschaften von $\hat{\tau}$ für unendlich groß werdende Stichproben mit $n \rightarrow \infty$.

Schätzereigenschaften bei endlichen Stichproben

Wir betrachten in der Folge zwei Aspekte von Schätzereigenschaften bei endlichen Stichproben:

- (1) Erwartungstreue
- (2) Varianz und Standardfehler

Intuitiv haben diese folgende Bedeutungen: Ein Schätzer $\hat{\tau}$ heißt *erwartungstreu*, wenn sein Erwartungswert dem wahren, aber unbekanntem, Wert $\tau(\theta)$ für alle $\theta \in \Theta$ gleich ist. Die *Varianz* eines Schätzers $\hat{\tau}$ ist die Varianz der Zufallsvariable $\hat{\tau}(v)$. Der *Standardfehler* eines Schätzers $\hat{\tau}$ ist die Standardabweichung der Zufallsvariable $\hat{\tau}(v)$.

Für folgende Schätzereigenschaften bei endlichen Stichproben verweisen wir auf das Vorlesungsskript:

- (1) Mittlerer Quadratischer Fehler
- (2) Cramér-Rao Ungleichung

Intuitiv haben diese folgende Bedeutungen: Der *mittlere quadratische Fehler* von $\hat{\tau}$ ist der Erwartungswert der quadrierten Abweichung von $\hat{\tau}(v)$ von $\tau(\theta)$ über Stichproben vom Umfang n . 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.

Definition (Fehler, Systematischer Fehler, und Erwartungstreue)

v sei die Stichprobe eines Frequentistischen Inferenzmodells und $\hat{\tau}$ sei ein Schätzer für τ .

- Der *Fehler* von $\hat{\tau}$ ist definiert als

$$\hat{\tau}(v) - \tau(\theta). \quad (26)$$

- Der *systematische Fehler (Bias)* von $\hat{\tau}$ ist definiert als

$$B(\hat{\tau}) := \mathbb{E}_{\theta}(\hat{\tau}(v)) - \tau(\theta). \quad (27)$$

- $\hat{\tau}$ heißt *erwartungstreu (unbiased)*, wenn

$$B(\hat{\tau}) = 0 \Leftrightarrow \mathbb{E}_{\theta}(\hat{\tau}(v)) = \tau(\theta) \text{ für alle } \theta \in \Theta \text{ und alle } n \in \mathbb{N}. \quad (28)$$

Andernfalls heißt $\hat{\tau}$ *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 $\mathbb{E}_{\theta}(\hat{\theta}(v)) = \theta$.

Theorem (Erwartungstreue von Stichprobenmittel und -varianz)

$v = (v_1, \dots, v_n)$ sei die Stichprobe eines parametrischen Produktmodells. Dann gelten

- Das Stichprobenmittel

$$\bar{v} := \frac{1}{n} \sum_{i=1}^n v_i \quad (29)$$

ist ein erwartungstreuer Schätzer des Erwartungswerts $\mathbb{E}_\theta(v_1)$.

- Die Stichprobenvarianz

$$S^2 := \frac{1}{n-1} \sum_{i=1}^n (v_i - \bar{v})^2 \quad (30)$$

ist ein erwartungstreuer Schätzer der Varianz $\mathbb{V}_\theta(v_1)$.

Schätzereigenschaften bei endlichen Stichproben

Beweis

Mit der Linearität von Erwartungswerten ergibt sich zunächst ergibt sich dann

$$\mathbb{E}_\theta(\bar{v}) = \mathbb{E}_\theta\left(\frac{1}{n} \sum_{i=1}^n v_i\right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_\theta(v_i) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_\theta(v_1) = \frac{1}{n} n \mathbb{E}_\theta(v_1) = \mathbb{E}_\theta(v_1).$$

Dies zeigt die Erwartungstreue des Stichprobenmittels als Schätzer des Erwartungswertes.

Um die Erwartungstreue der Stichprobenvarianz zu zeigen, halten wir zunächst fest, dass

$$\mathbb{V}_\theta(\bar{v}) = \mathbb{V}_\theta\left(\frac{1}{n} \sum_{i=1}^n v_i\right) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}_\theta(v_i) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}_\theta(v_1) = \frac{1}{n^2} n \mathbb{V}_\theta(v_1) = \frac{\mathbb{V}_\theta(v_1)}{n}.$$

Weiterhin gilt

$$\sum_{i=1}^n (v_i - \bar{v})^2 = \sum_{i=1}^n (v_i - \mathbb{E}_\theta(v_1))^2 - n(\bar{v} - \mathbb{E}_\theta(v_1))^2,$$

weil

$$\begin{aligned}\sum_{i=1}^n (v_i - \bar{v})^2 &= \sum_{i=1}^n (v_i - \mathbb{E}_\theta(v_1) - \bar{v} + \mathbb{E}_\theta(v_1))^2 \\ &= \sum_{i=1}^n ((v_i - \mathbb{E}_\theta(v_1)) - (\bar{v} - \mathbb{E}_\theta(v_1)))^2 \\ &= \sum_{i=1}^n (v_i - \mathbb{E}_\theta(v_1))^2 - 2(\bar{v} - \mathbb{E}_\theta(v_1)) \left(\sum_{i=1}^n (v_i - \mathbb{E}_\theta(v_1)) \right) + \sum_{i=1}^n (\bar{v} - \mathbb{E}_\theta(v_1))^2 \\ &= \sum_{i=1}^n (v_i - \mathbb{E}_\theta(v_1))^2 - 2(\bar{v} - \mathbb{E}_\theta(v_1)) \left(\sum_{i=1}^n v_i - n\mathbb{E}_\theta(v_1) \right) + n(\bar{v} - \mathbb{E}_\theta(v_1))^2 \quad (31) \\ &= \sum_{i=1}^n (v_i - \mathbb{E}_\theta(v_1))^2 - 2(\bar{v} - \mathbb{E}_\theta(v_1)) \left(n \left(\frac{1}{n} \sum_{i=1}^n v_i \right) - n\mathbb{E}_\theta(v_1) \right) + n(\bar{v} - \mathbb{E}_\theta(v_1))^2 \\ &= \sum_{i=1}^n (v_i - \mathbb{E}_\theta(v_1))^2 - 2n(\bar{v} - \mathbb{E}_\theta(v_1))^2 + n(\bar{v} - \mathbb{E}_\theta(v_1))^2 \\ &= \sum_{i=1}^n (v_i - \mathbb{E}_\theta(v_1))^2 - n(\bar{v} - \mathbb{E}_\theta(v_1))^2.\end{aligned}$$

Schätzereigenschaften bei endlichen Stichproben

Es ergibt sich also

$$\begin{aligned}\mathbb{E}_\theta \left((n-1)S^2 \right) &= \mathbb{E}_\theta \left(\sum_{i=1}^n (v_i - \bar{v})^2 \right) \\ &= \mathbb{E}_\theta \left(\sum_{i=1}^n (v_i - \mathbb{E}_\theta(v_1))^2 - n(\bar{v} - \mathbb{E}_\theta(v_1))^2 \right) \\ &= \sum_{i=1}^n \mathbb{E}_\theta \left((v_i - \mathbb{E}_\theta(v_1))^2 \right) - n\mathbb{E}_\theta \left((\bar{v} - \mathbb{E}_\theta(v_1))^2 \right) \\ &= n\mathbb{V}_\theta(v_1) - n\mathbb{V}_\theta(\bar{v}) \\ &= n\mathbb{V}_\theta(v_1) - n \frac{\mathbb{V}_\theta(v_1)}{n} \\ &= n\mathbb{V}_\theta(v_1) - \mathbb{V}_\theta(v_1) \\ &= (n-1)\mathbb{V}_\theta(v_1)\end{aligned}$$

Schließlich ergibt sich

$$\mathbb{E}_\theta(S^2) = \mathbb{E}_\theta \left(\frac{1}{n-1} (n-1)S^2 \right) = \frac{1}{n-1} \mathbb{E}_\theta \left((n-1)S^2 \right) = \frac{1}{n-1} (n-1)\mathbb{V}_\theta(v_1) = \mathbb{V}_\theta(v_1)$$

und damit die Erwartungstreue der Stichprobenvarianz als Schätzer der Varianz. □

Theorem (Verzerrtheit der Stichprobenstandardabweichung)

$v := (v_1, \dots, v_n)$ sei die Stichprobe eines parametrischen Produktmodells. Dann ist die Stichprobenstandardabweichung

$$S := \sqrt{S^2} \quad (32)$$

ein verzerrter Schätzer der Standardabweichung $\mathbb{S}_\theta(v_1)$.

Beweis

Wir halten zunächst fest, dass $\sqrt{\cdot}$ eine strikt konkave Funktion und $\sigma^2 > 0$ ist. Dann aber gilt mit der Jensenschen Ungleichung $\mathbb{E}(f(\xi)) < f(\mathbb{E}(\xi))$ für strikt konkave Funktionen, dass

$$\mathbb{E}_\theta(S) = \mathbb{E}_\theta(\sqrt{S^2}) < \sqrt{\mathbb{E}_\theta(S^2)} = \sqrt{\mathbb{V}_\theta(v_1)} = \mathbb{S}_\theta(v_1). \quad (33)$$

□

Bemerkung

- Nichtlineare Transformationen von erwartungstreuen Schätzern liefern oft verzerrte Schätzer.

Schätzereigenschaften bei endlichen Stichproben

Simulation von $v_1, \dots, v_{12} \sim N(\mu, \sigma^2)$ 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
nsim    = 1e5               # Anzahl der Simulationen
y_bar   = rep(NaN,nsim)     # Stichprobenmittelarray
s_sqr   = rep(NaN,nsim)     # Stichprobenvarianzarray
s       = rep(NaN,nsim)     # Stichprobenstandardabweichungarray

# Simulationsiterationen
for(sim in 1:nsim){

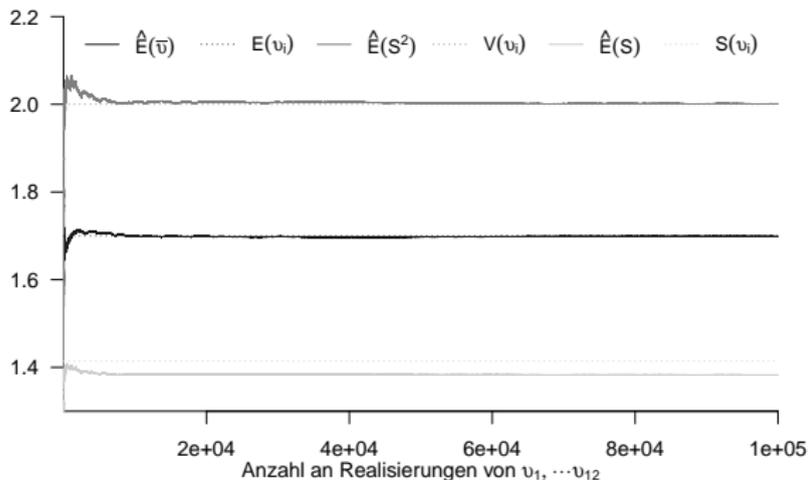
  # Stichprobenrealisation von \ups_1, \dots, \ups_{12}
  y      = rnorm(n,mu,sqrt(sigsqr))

  # Erwartungswert-, Varianz-, Standardabweichungsschaetzer
  y_bar[sim] = mean(y)      # Stichprobenmittel
  s_sqr[sim] = var(y)       # Stichprobenvarianz
  s[sim]     = sd(y)        # Stichprobenstandardabweichung
}

# Erwartungswertschaetzung
E_hat_y_bar = cumsum(y_bar)/(1:nsim) # \mathbb{E}(\bar{\ups}) Schaetzungen
E_hat_s_sqr = cumsum(s_sqr)/(1:nsim) # \mathbb{E}(S^2) Schaetzungen
E_hat_s     = cumsum(s) / (1:nsim)  # \mathbb{E}(S) Schaetzungen
```

Schätzereigenschaften bei endlichen Stichproben

Simulation von $v_1, \dots, v_{12} \sim N(\mu, \sigma^2)$ mit $n = 12$, $\mu = 1.7$, $\sigma^2 = 2$, $\sigma \approx 1.41$



Definition (Varianz und Standardfehler)

$v := (v_1, \dots, v_n)$ die Stichprobe eines frequentistischen Inferenzmodells und $\hat{\tau}$ sei ein Schätzer von τ .

- Die *Varianz* von $\hat{\tau}$ ist definiert als

$$\mathbb{V}_\theta(\hat{\tau}) := \mathbb{E}_\theta((\hat{\tau}(v) - \mathbb{E}_\theta(\hat{\tau}(v)))^2). \quad (34)$$

- Der *Standardfehler* von $\hat{\tau}$ ist definiert als

$$\text{SE}(\hat{\tau}) := \sqrt{\mathbb{V}_\theta(\hat{\tau})} \quad (35)$$

Bemerkungen

- Die Varianz eines Schätzers $\hat{\tau}$ ist die Varianz der Zufallsvariable $\hat{\tau}(v)$.
- Der Standardfehler eines Schätzers $\hat{\tau}$ ist die Standardabweichung von $\hat{\tau}(v)$.

Theorem (Standardfehler des Stichprobenmittels)

$v := (v_1, \dots, v_n)$ sei die Stichprobe eines parametrischen Produktmodells. Dann ist der *Standardfehler des Stichprobenmittels* gegeben durch

$$\text{SE}(\bar{v}) = \frac{\mathbb{S}_\theta(v_1)}{\sqrt{n}}. \quad (36)$$

Beweis

Per definitionem und mit $\mathbb{V}_\theta(\bar{v}) = \mathbb{V}_\theta(v_1)/n$, ergibt sich

$$\text{SE}(\bar{v}) = \sqrt{\mathbb{V}_\theta(\bar{v})} = \sqrt{\frac{\mathbb{V}_\theta(v_1)}{n}} = \frac{\mathbb{S}_\theta(v_1)}{\sqrt{n}}. \quad (37)$$

Bemerkungen

- Der Standardfehler des Mittelwerts beschreibt die Variabilität des Stichprobenmittels.
- Da $\mathbb{S}_\theta(v_1)$ unbekannt ist, ist auch $\text{SE}(\bar{v})$ unbekannt.
- Ein verzerrter Schätzer für den Standardfehler des Stichprobenmittels ist gegeben durch $\hat{\text{SE}}(\bar{v}) = \frac{s}{\sqrt{n}}$.

Beispiel (Standardfehler des Bernoulli Parameter Maximum-Likelihood Schätzes)

Es sei $v := (v_1, \dots, v_n)$ die Stichprobe eines Bernoulli Modells und $\hat{\mu}^{\text{ML}}$ der Maximum-Likelihood Schätzer für μ . Dann ist

$$\text{SE}(\hat{\mu}^{\text{ML}}) = \sqrt{\frac{\mu(1-\mu)}{n}}. \quad (38)$$

Beweis

Es gilt

$$\text{SE}(\hat{\mu}^{\text{ML}}) = \sqrt{\mathbb{V}_{\mu}(\hat{\mu}^{\text{ML}})} = \sqrt{\mathbb{V}_{\mu}\left(\frac{1}{n} \sum_{i=1}^n v_i\right)} = \sqrt{\frac{1}{n^2} \sum_{i=1}^n \mathbb{V}_{\mu}(v_i)} = \sqrt{\frac{n\mu(1-\mu)}{n^2}} = \sqrt{\frac{\mu(1-\mu)}{n}}, \quad (39)$$

wobei die dritte Gleichung mit der Unabhängigkeit der $v_i, i = 1, \dots, n$ und die vierte Gleichung mit der Varianz $\mathbb{V}_{\mu}(v_1) = \mathbb{V}_{\mu}(v_i) = \mu(1-\mu), i = 1, \dots, n$ der Bernoulli Stichprobenvariablen folgt.

Bemerkung

- Ein Schätzer für den Standardfehler $\text{SE}(\hat{\mu}^{\text{ML}})$ ist $\hat{\text{SE}}(\hat{\mu}^{\text{ML}}) = \sqrt{\frac{\hat{\mu}^{\text{ML}}(1-\hat{\mu}^{\text{ML}})}{n}}$

Schätzereigenschaften bei endlichen Stichproben

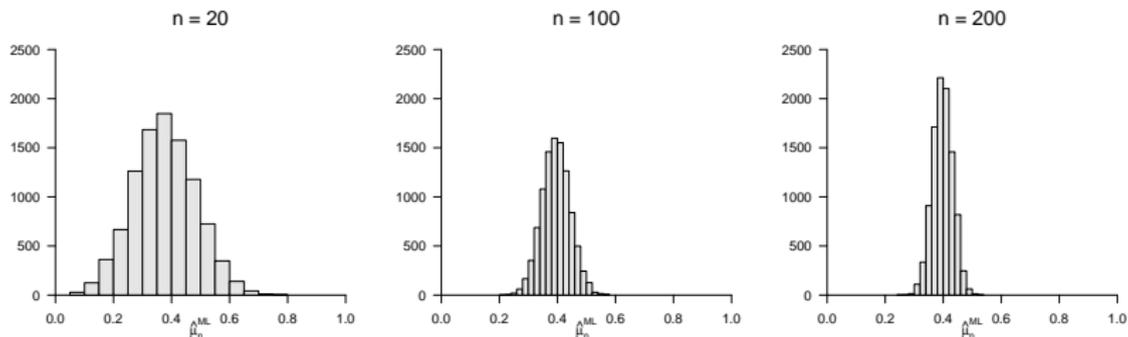
Simulation von $v_1, \dots, v_n \sim \text{Bern}(\mu)$ mit $\mu = 0.4$

```
# Modellformulierung
set.seed(0)
mu      = 0.4
n_all   = c(20,100,200)
ns      = 1e4
mu_hat_ML = matrix(rep(NaN, length(n_all)*ns),
                   nrow = length(n_all))

# Zufallszahlengeneratorzustand
# wahrer, aber unbekannter, Parameterwert
# Stichprobenumfänge
# Anzahl der Simulationen
# Maximum-Likelihood Schätzearray

# Stichprobengroesseniterationen
for(i in seq_along(n_all)){
  # Simulationsiterationen
  for(s in 1:ns){
    y      = rbinom(n_all[i],1,mu)
    mu_hat_ML[i,s] = mean(y)
  }
}
```

Die Varianz bzw. der Standardfehler von $\hat{\mu}^{\text{ML}}$ hängen von n ab.



Maximum-Likelihood-Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood-Schätzern

Selbstkontrollfragen

Asymptotische Schätzereigenschaften

Vorbemerkungen zu Asymptotischen Schätzereigenschaften

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 Stichprobenumfänge zu approximieren.

Moderne Stichproben sind üblicherweise groß.

Die Methoden der AS sind also praktisch einsetzbar und gerechtfertigt.

Vaart (1998) gibt eine ausführliche Einführung in die AS.

Asymptotische Schätzereigenschaften

Wir betrachten im Folgenden drei asymptotische Schätzereigenschaften:

- (1) Asymptotische Erwartungstreue
- (2) Konsistenz
- (3) Asymptotische Normalverteilung

Um zu betonen, dass in diesem Abschnitt die Eigenschaften eines Schätzers $\hat{\tau}$ vom Stichprobenumfang abhängen, schreiben wir im Folgenden $\hat{\tau}_n$. Intuitiv haben obige asymptotische Schätzereigenschaften folgende Bedeutungen: Ein Schätzer $\hat{\tau}_n$ für τ heißt *asymptotisch erwartungstreu*, wenn der Erwartungswert von $\hat{\tau}_n$ für große Stichprobenumfänge $n \rightarrow \infty$ gleich dem wahren, aber unbekanntem, Wert $\tau(\theta)$ ist. Ein Schätzer $\hat{\tau}_n$ für τ heißt *konsistent*, wenn für große Stichprobenumfänge $n \rightarrow \infty$ die Wahrscheinlichkeit dafür, dass $\hat{\tau}_n$ vom wahren, aber unbekanntem, Wert $\tau(\theta)$ abweicht, beliebig klein wird. Ein Schätzer $\hat{\tau}_n$ für τ heißt *asymptotisch normalverteilt*, wenn für große Stichprobenumfänge $n \rightarrow \infty$, die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung gegeben ist.

Für folgende asymptotische Schätzereigenschaften verweisen wir auf das Vorlesungsskript:

- (1) Asymptotische Effizienz

Intuitiv hat diese die folgende Bedeutung: Ein Schätzer $\hat{\tau}_n$ für τ heißt *asymptotisch effizient*, wenn für große Stichprobenumfänge $n \rightarrow \infty$ 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)

v sei die Stichprobe eines parametrischen Produktmodells und $\hat{\tau}_n$ sei ein Schätzer für τ . $\hat{\tau}_n$ heißt *asymptotisch erwartungstreu*, wenn

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\theta}(\hat{\tau}_n(v)) = \tau(\theta) \text{ für alle } \theta \in \Theta. \quad (40)$$

Bemerkungen

- Asymptotisch erwartungstreu Schätzer sind für "unendlich große" Stichproben erwartungstreu.
- Erwartungstreu Schätzer sind immer auch asymptotisch erwartungstreu.

Theorem (Asymp. Erwartungstreue des Varianzparameterschätzers)

$v := (v_1, \dots, v_n)$ sei die Stichprobe eines Normalverteilungsmodells mit Varianzparameter σ^2 . Dann ist der Maximum-Likelihood-Schätzer von σ^2 ,

$$\hat{\sigma}_n^{2\text{ML}} := \frac{1}{n} \sum_{i=1}^n (v_i - \bar{v}_n)^2 \quad (41)$$

nicht erwartungstreu, aber asymptotisch erwartungstreu.

Beweis

Mit der Erwartungstreue der Stichprobenvarianz ergibt sich

$$\mathbb{E}_{\mu, \sigma^2} (\hat{\sigma}_n^{2\text{ML}}) = \mathbb{E}_{\mu, \sigma^2} \left(\frac{1}{n} \sum_{i=1}^n (v_i - \bar{v}_n)^2 \right) = \frac{1}{n} \mathbb{E}_{\mu, \sigma^2} \left(\sum_{i=1}^n (v_i - \bar{v}_n)^2 \right) = \frac{n-1}{n} \sigma^2$$

Also gilt $\mathbb{E}_{\mu, \sigma^2} (\hat{\sigma}_n^{2\text{ML}}) \neq \sigma^2$. $\hat{\sigma}_n^{2\text{ML}}$ ist also ein verzerrter Schätzer von σ^2 . Allerdings gilt $(n-1)/n \rightarrow 1$ für $n \rightarrow \infty$, so dass

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mu, \sigma^2} (\hat{\sigma}_n^{2\text{ML}}) = \lim_{n \rightarrow \infty} \frac{n-1}{n} \sigma^2 = \sigma^2 \lim_{n \rightarrow \infty} \frac{n-1}{n} = \sigma^2. \quad (42)$$

Asymptotische Schätzereigenschaften

Simulation von $v_1, \dots, v_n \sim N(\mu, \sigma^2)$ 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(
  rep(NaN, length(n)*ns), # \hat{\sigma^2}^{ML} Array
  ncol = length(n))

# Stichprobengroesseniterationen
for(i in seq_along(n)){

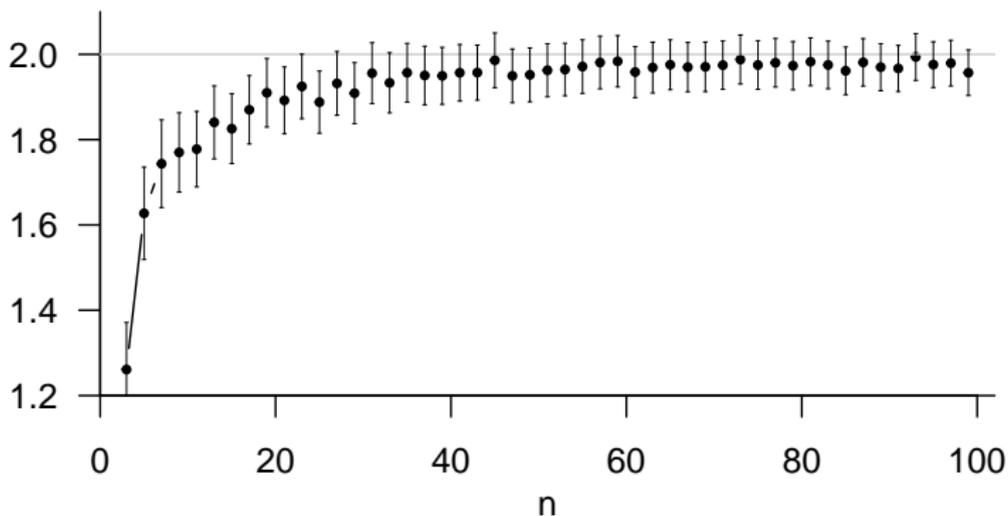
  # Simulationsiterationen
  for(s in 1:ns){

    # Stichprobenrealisation
    y = rnorm(n[i], mu, sqrt(sigsqr))

    # \hat{\sigma^2}^{ML}
    sigsqr_ml[s,i] = ((n[i]-1)/n[i])*var(y)
  }
}
E_sigsqr_ml = colMeans(sigsqr_ml) # Erwartungswertschaetzung
```

Asymptotische Schätzereigenschaften

Simulation von $v_1, \dots, v_n \sim N(\mu, \sigma^2)$ mit $\mu = 1, \sigma^2 = 2$



Definition (Konsistenz)

v sei die Stichprobe eines parametrischen Produktmodells 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 beliebig kleine $\epsilon > 0$ und jedes $\theta \in \Theta$ gilt, dass

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\theta} (|\hat{\tau}_n(v) - \tau(\theta)| \geq \epsilon) = 0. \quad (43)$$

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 $\hat{\tau}_n(v)$ beliebig nah bei $\tau(\theta)$ liegt, groß.
- Für $n \rightarrow \infty$ wird die Wahrscheinlichkeit, dass $\hat{\tau}_n(v)$ von $\tau(\theta)$ abweicht, klein.
- Diese Eigenschaften gelten für alle möglichen wahren, aber unbekanntenen, Parameterwerte.

Asymptotische Schätzeigenschaften

Simulation der Konsistenz von \bar{v}_n bei $v_1, \dots, v_n \sim N(\mu, \sigma^2)$ 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      = 1000             # Anzahl Simulationen
E       = array(rep(NA,n*ne*ns),dim = c(nn,ne,ns)) # Ereignisindikatorarray

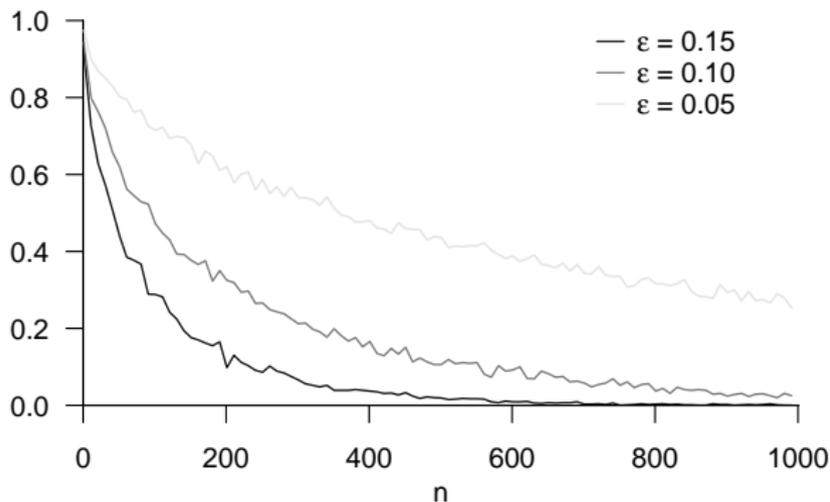
# Simulation
for(e in seq_along(eps)){ # \epsilon Iterationen
  for(i in seq_along(n)){ # n Iterationen
    for(s in 1:ns){ # Simulationsiterationen

      # Stichprobenrealisationen
      y = rnorm(n[i], mu, sqrt(sigsqr))
      if(abs(mean(y) - mu) >= eps[e]){ # |y_bar - \mu| \ge \epsilon
        E[i,e,s] = 1
      } else { # |y_bar - \mu| < \epsilon
        E[i,e,s] = 0
      }
    }
  }
}

# Schaetzung von \mathbb{P}(|\hat{\tau}_n(\ups) - \tau(\theta)| \ge \epsilon)
P_hat = apply(E, c(1,2), mean)
```

Asymptotische Schätzereigenschaften

Simulation der Konsistenz von \bar{v}_n bei $v_1, \dots, v_n \sim N(\mu, \sigma^2)$ mit $\mu = 1, \sigma^2 = 2$



Definition (Asymptotische Normalität)

v sei die Stichprobe eines parametrischen Produktmodells 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

$$\hat{\theta}_n \stackrel{a}{\sim} N(\mu, \sigma^2). \quad (44)$$

Bemerkung

- Konvergenz in Verteilung heißt $\lim_{n \rightarrow \infty} P_n(\hat{\theta}_n) = P(\tilde{\theta})$.

Asymptotische Schätzereigenschaften

Simulation der asymptotische Normalverteilung des ML Bernoulliparameterschätzers

```
# 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_y = seq(0,1,len = mu_hat_ML_r)          # ML Schaetzerraum
mu_hat_ML_p = matrix(rep(NA, length(n_all)*mu_hat_ML_r),
  nrow = length(n_all))                            # ML WDF Array

# Stichprobengroesseniterationen
for(i in seq_along(n_all)){

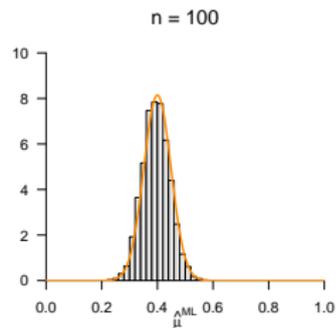
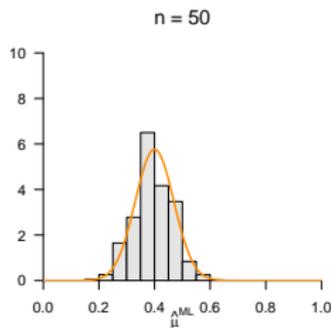
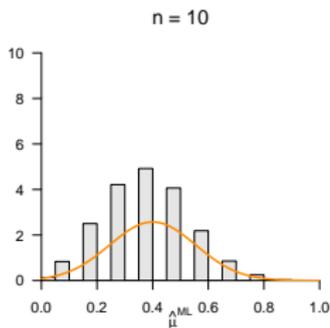
  # Simulationsiterationen
  for(s in 1:ns){
    y           = rbinom(n_all[i],1,mu)           # Stichprobenrealisation
    mu_hat_ML[i,s] = mean(y)                     # ML Schaetzer
  }

  # WDF der asymptotischen Verteilung
  mu_hat_ML_p[i,] = dnorm(mu_hat_ML_y, mu, sqrt(mu*(1-mu)/n_all[i]))
}
```

Asymptotische Schätzereigenschaften

Simulation der asymptotischen Normalverteilung des ML Bernoulli parameterschätzers

— Histogramm — $N(\hat{\mu}^{\text{ML}}; \mu, J_n^{-1}(\mu))$



Maximum-Likelihood-Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood-Schätzern

Selbstkontrollfragen

Theorem (Eigenschaften von Maximum-Likelihood-Schätzern)

Gegeben sei in Frequentistisches Inferenzmodell mit wahren, aber unbekanntem, Parameter θ und $\hat{\theta}_n^{\text{ML}}$ sei ein Maximum-Likelihood-Schätzer für θ . Dann gilt, dass $\hat{\theta}_n^{\text{ML}}$

- (1) nicht notwendigerweise erwartungstreu, aber
- (2) asymptotisch erwartungstreu,
- (3) konsistent und
- (4) asymptotisch normalverteilt ist.

Bemerkungen

- Maximum-Likelihood-Schätzer sind überdies asymptotisch effizient.
- Für einen Beweis verweisen wir auf Held and Sabanés Bové (2014), Abschnitt 3.4.

Maximum-Likelihood-Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood-Schätzern

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die Definition des Begriffs eines Parameterpunktschätzers wieder.
2. Erläutern Sie den Begriff des Parameterpunktschätzers.
3. Geben Sie Definition der Begriffe der Likelihood-Funktion und der Log-Likelihood-Funktion wieder.
4. Geben Sie Definition des Begriffs des Maximum-Likelihood Schätzes wieder.
5. Erläutern Sie das Vorgehen zur Gewinnung von Maximum-Likelihood-Schätzern.
6. Geben Sie das Theorem zum Maximum-Likelihood-Schätzer des Bernoullimodellparameters wieder.
7. Geben Sie das Theorem zu den Maximum-Likelihood-Schätzern der Normalverteilungsmodellparameter wieder.
8. Geben Sie die Definition des Begriffs der Erwartungstreue eines Schätzers wieder.
9. Erläutern Sie den Begriff der Erwartungstreue eines Schätzers.
10. Geben Sie Definition der Begriffe der Varianz und des Standardfehlers eines Schätzers wieder.
11. Erläutern Sie den Begriff der asymptotischen Erwartungstreue eines Schätzers.
12. Erläutern Sie den Begriff der Konsistenz eines Schätzers.
13. Erläutern Sie den Begriff der asymptotischen Normalität eines Schätzers.
14. Geben Sie das Theorem zu den Eigenschaften von Maximum-Likelihood-Schätzern wieder.

1. Siehe Definition (Parameterpunktschätzer).
2. Ein Parameterpunktschätzer gibt basierend auf einer Stichprobe einen skalaren Tipp für den entsprechenden wahren, aber unbekanntem, Parameter an.
3. Siehe Definition (Likelihood-Funktion und Log-Likelihood-Funktion).
4. Siehe Definition (Maximum-Likelihood-Schätzer)
5. Um einen Maximum-Likelihood-Schätzer zu gewinnen, formuliert man zunächst die Log-Likelihood-Funktion und bestimmt dann die Nullstellen ihrer Ableitung als potentielle Maximumstellen. Dazu nutzt man in klassischen Beispielen meist die analytische, in der Anwendung meist die numerische Optimierung.
6. Siehe Theorem (Maximum-Likelihood-Schätzer des Bernoulli-Modells).
7. Siehe Theorem (Maximum-Likelihood-Schätzer des Normalverteilungsmodells).
8. Siehe Definition (Fehler, Systematischer Fehler, und Erwartungstreue).
9. Ein Schätzer heißt erwartungstreu, wenn sein Erwartungswert mit dem von ihm geschätzten wahren, aber unbekanntem, Wert identisch ist.
10. Siehe Definition (Varianz und Standardfehler).
11. Ein Schätzer heißt asymptotisch erwartungstreu, wenn sein Erwartungswert für gegen unendlich gehende Stichprobenumfänge mit dem von ihm geschätzten wahren, aber unbekanntem, Wert identisch ist.
12. Ein Schätzer heißt konsistent, wenn für große Stichprobenumfänge die Wahrscheinlichkeit dafür, dass der Schätzerwert vom wahren, aber unbekanntem, Wert abweicht, beliebig klein wird.
13. Ein Schätzer heißt asymptotisch normalverteilt, die Verteilung des Schätzers für große Stichprobenumfänge durch eine Normalverteilung gegeben ist.
14. Siehe Theorem (Eigenschaften von Maximum-Likelihood-Schätzern).

- Held, Leonhard, and Daniel Sabanés Bové. 2014. *Applied Statistical Inference*. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-37887-4>.
- Vaart, A. W. van der. 1998. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge, UK ; New York, NY, USA: Cambridge University Press.



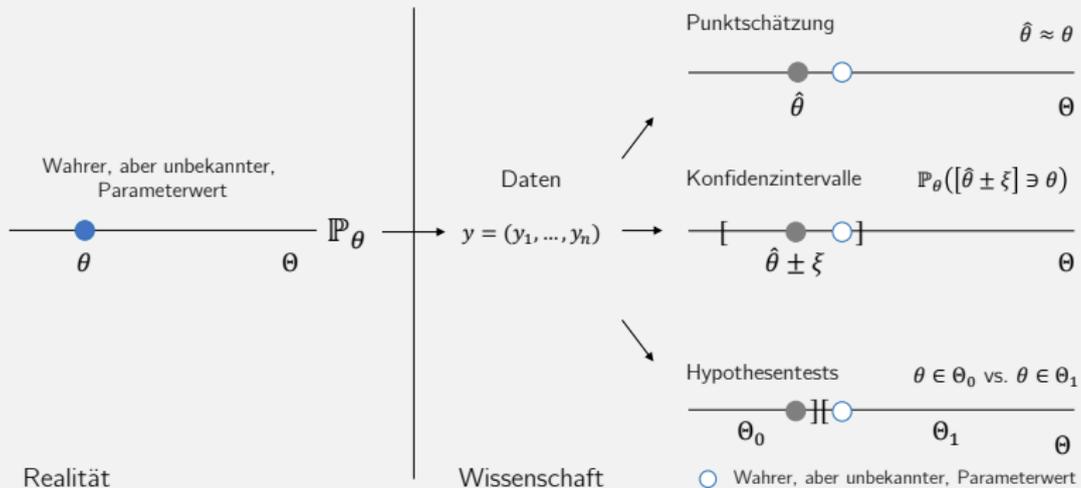
Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2023/24

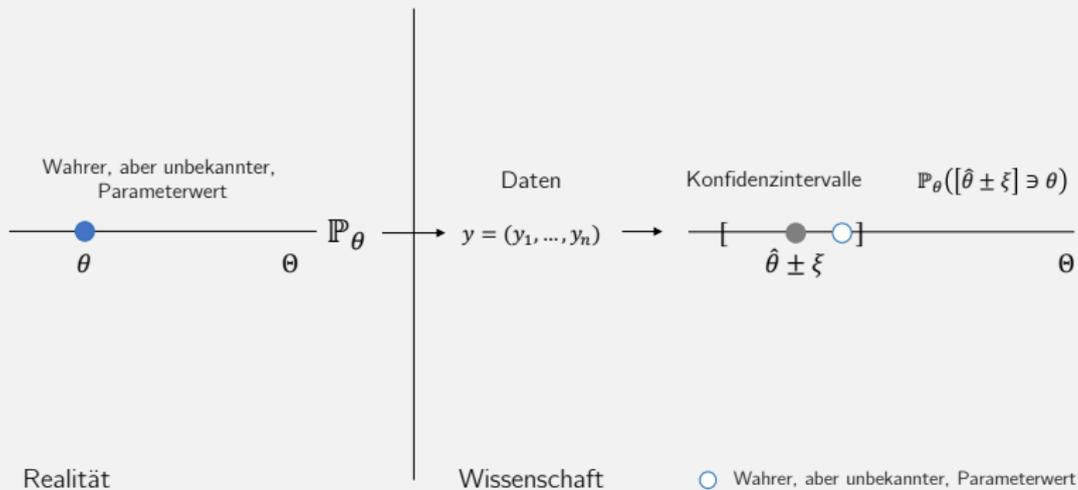
Prof. Dr. Dirk Ostwald

(10) Konfidenzintervalle

Standardproblemstellungen und Standardannahme Frequentistischer Inferenz



Standardproblemstellungen und Standardannahme Frequentistischer Inferenz



Standardannahme Frequentistischer Inferenz

\mathcal{M} sei ein Frequentistisches Inferenzmodell mit Zufallsvektor v . Es wird angenommen, dass ein Datensatz $y \in \mathbb{R}^n$ eine der möglichen Realisierungen von v ist.

Aus Frequentistischer Sicht kann man eine Studie unendlich oft wiederholen und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. das Stichprobenmittel:

$$\text{Datensatz (1)} : y^{(1)} = \left(y_1^{(1)}, y_2^{(1)}, \dots, y_n^{(1)} \right) \text{ mit } \bar{y}^{(1)} = \frac{1}{n} \sum_{i=1}^n y_i^{(1)}$$

$$\text{Datensatz (2)} : y^{(2)} = \left(y_1^{(2)}, y_2^{(2)}, \dots, y_n^{(2)} \right) \text{ mit } \bar{y}^{(2)} = \frac{1}{n} \sum_{i=1}^n y_i^{(2)}$$

$$\text{Datensatz (3)} : y^{(3)} = \left(y_1^{(3)}, y_2^{(3)}, \dots, y_n^{(3)} \right) \text{ mit } \bar{y}^{(3)} = \frac{1}{n} \sum_{i=1}^n y_i^{(3)}$$

$$\text{Datensatz (4)} : y^{(4)} = \left(y_1^{(4)}, y_2^{(4)}, \dots, y_n^{(4)} \right) \text{ mit } \bar{y}^{(4)} = \frac{1}{n} \sum_{i=1}^n y_i^{(4)}$$

$$\text{Datensatz (5)} : y^{(5)} = \dots$$

Um die Qualität ihrer Methoden zu beurteilen betrachtet die Frequentistische Inferenz deshalb die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken. Was zum Beispiel ist die Verteilung der $\bar{y}^{(1)}, \bar{y}^{(2)}, \bar{y}^{(3)}, \bar{y}^{(4)}, \dots$ also die Verteilung der Zufallsvariable \bar{v}_n ?

Wenn eine Methode im Sinne der Frequentistischen Standardannahme "gut" ist, dann heißt das also, dass sie bei häufiger Anwendung "im Mittel gut" ist. Im Einzelfall, also im Normalfall nur eines vorliegenden Datensatzes, kann sie auch "schlecht" sein.

Definition

Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

Anwendungsbeispiel

Selbstkontrollfragen

Definition

Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

Anwendungsbeispiel

Selbstkontrollfragen

Definition (δ -Konfidenzintervall)

Es sei v die Stichprobe eines Frequentistischen Inferenzmodells mit wahren, aber unbekanntem Parameter, $\theta \in \Theta$, es sei $\delta \in]0, 1[$ und es seien $G_u(v)$ und $G_o(v)$ zwei Statistiken. Dann heißt ein Intervall der Form

$$\kappa(v) := [G_u(v), G_o(v)], \quad (1)$$

so dass

$$\mathbb{P}_\theta(\kappa(v) \ni \theta) = \mathbb{P}_\theta(G_u(v) \leq \theta \leq G_o(v)) = \delta \text{ für alle } \theta \in \Theta \text{ gilt} \quad (2)$$

ein δ -Konfidenzintervall für θ . δ ist die Überdeckungswahrscheinlichkeit von $\kappa(v)$ für θ und wird *Konfidenzlevel* genannt. Die Statistiken $G_u(v)$ und $G_o(v)$ heißen die *unteren* und *oberen Grenzen* des Konfidenzintervalls, respektive.

Bemerkungen

- θ ist fest, nicht zufällig, und unbekannt.
- κ ist ein zufälliges Intervall, weil $G_u(v)$ und $G_o(v)$ Zufallsvariablen sind.
- $\kappa \ni \theta$ bedeutet $\theta \in \kappa$, aber κ ist zufällig und steht deshalb vorn (cf. $\mathbb{P}(\xi = x)$).
- 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 Zufallsvorgang unter gleichen Umständen häufig wiederholt, so überdeckt das zugehörige δ -Konfidenzintervall den wahren, aber unbekanntem Parameterwert im langfristigen Mittel in $\delta \cdot 100\%$ der Fälle. Technischer ausgedrückt, für unabhängig und identisch realisierte Stichproben einer Verteilung mit wahren, aber unbekanntem, Parameter θ überdeckt im langfristigen Mittel ein entsprechendes δ -Konfidenzintervall θ in $\delta \cdot 100\%$ aller Fälle.
- (2) Gegeben sei eine Menge von Zufallsvorgängen mit wahren, aber unbekanntem, Parametern $\theta_1, \theta_2, \dots$ und realisierte δ -Konfidenzintervalle für eben jene Menge von wahren, aber unbekanntem Parametern $\theta_1, \theta_2, \dots$. Dann überdecken im langfristigen Mittel $\delta \cdot 100\%$ der Konfidenzintervalle den wahren, aber unbekanntem, Wert θ_i für $i = 1, 2, \dots$. Technischer ausgedrückt, für unabhängig realisierte Stichproben von Verteilungen mit wahren, aber unbekanntem, Parametern $\theta_1, \theta_2, \dots$ überdecken im langfristigen Mittel entsprechende δ -Konfidenzintervalle θ_i für $i = 1, 2, \dots$ in $\delta \cdot 100\%$ aller Fälle.

Wir demonstrieren im Folgenden beide Interpretationen mithilfe von Simulationen.

Allgemeine Konstruktion von Konfidenzintervallen

- (1) Definition des Frequentistischen Inferenzmodells
- (2) Definition der Konfidenzintervallstatistik
- (3) Analyse der Verteilung der Konfidenzintervallstatistik
- (4) Etablierung der Konfidenzbedingung

Beispiel

- (1) Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

Definition

Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

Anwendungsbeispiel

Selbstkontrollfragen

(1) Definition des Frequentistischen Inferenzmodells

$v := (v_1, \dots, v_n)$ sei die Stichprobe eines Normalverteilungsmodells mit unbekanntem Parameter μ und $\sigma^2 > 0$. Wir entwickeln ein δ -Konfidenzintervall für den Erwartungswertparameter μ .

(2) Definition der Statistik

Wir betrachten die T -Konfidenzintervallstatistik

$$T := \frac{\sqrt{n}}{S}(\bar{v} - \mu) \text{ mit } \bar{v} := \frac{1}{n} \sum_{i=1}^n v_i \text{ und } S := \sqrt{\frac{1}{n-1} \sum_{i=1}^n (v_i - \bar{v})^2}. \quad (3)$$

(3) Analyse der Verteilung der Konfidenzintervallstatistik

Für die T -Konfidenzintervallstatistik gilt $T \sim t(n-1)$, die T -Konfidenzintervallstatistik ist also eine t -verteilte Zufallsvariable mit Freiheitsgradparameter $n-1$. Wir verzichten auf einen Beweis. Die T -Konfidenzintervallstatistik ist eine Funktion der Stichprobe v_1, \dots, v_n (via \bar{v} und S), während ihre Verteilung weder von μ noch von σ^2 abhängt. Wir bezeichnen die WDF einer t -verteilten Zufallsvariable mit t , die KVF einer t -verteilten Zufallsvariable mit Ψ und die inverse KVF einer t -verteilten Zufallsvariable mit Ψ^{-1} .

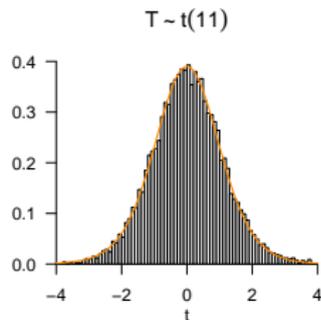
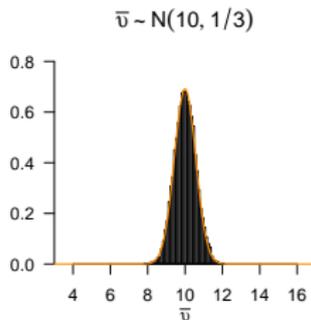
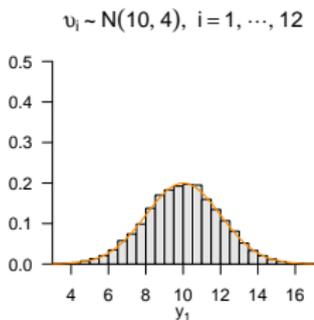
(3) Analyse der Verteilung der Konfidenzintervallstatistik (fortgeführt)

```
# Modellformulierung
mu      = 10                                # w.a.u. Erwartungswertparameter
sigsqr  = 4                                # wahrer bekannter Varianzparameter
n       = 12                                # Stichprobenumfang
ns      = 1e4                               # Anzahl Stichprobenrealisierungen
res     = 1e3                               # Ausgangsraumaufloesung

# analytische Definitionen und Resultate
y_1     = seq(3,17,len = res)              # y_1 Raum
t       = seq(-4,4,len = res)              # t Raum
p_y_1   = dnorm(y_1,mu,sqrt(sigsqr))       # y_1 WDF
p_y_bar = dnorm(y_1,mu,sqrt(sigsqr/n))     # y_bar WDF
p_t     = dt(t,n-1)                       # t WDF

# Simulation
y_i     = rep(NaN,ns)                      # y_1 Array
y_bar   = rep(NaN,ns)                     # \bar{y} Array
Tee     = rep(NaN,ns)                     # T Array
for(s in 1:ns){
  y      = rnorm(n,mu,sqrt(sigsqr))        # Simulationsiterationen
  y_i[s] = y[1]                            # Stichprobenrealisierung
  y_bar[s] = mean(y)                       # \ups_i
  Tee[s]  = sqrt(n)*((y_bar[s] - mu)/sqrt(var(y))) # Stichprobenmittelrealisierung
}                                           # T-Statistik Realisierung
```

(3) Analyse der Verteilung der Konfidenzintervallstatistik (fortgeführt)



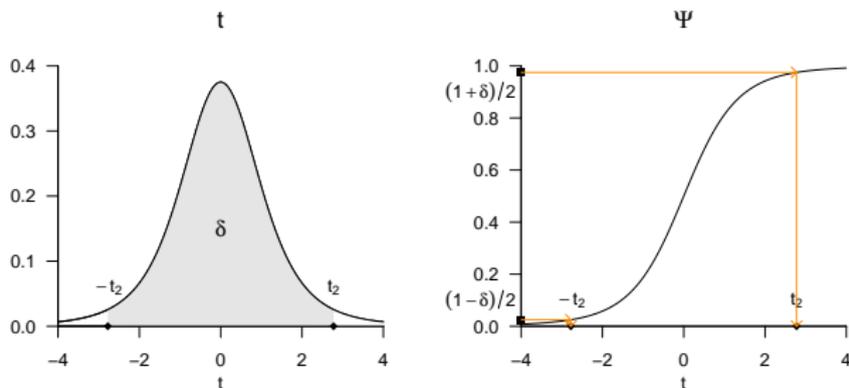
Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

(4) Etablierung der Konfidenzbedingung

Für $\delta \in]0, 1[$ seien

$$t_1 := \Psi^{-1}\left(\frac{1-\delta}{2}; n-1\right) \text{ und } t_2 := \Psi^{-1}\left(\frac{1+\delta}{2}; n-1\right) \quad (4)$$

Es gilt dann $(1+\delta)/2 - (1-\delta)/2 = \delta$ und 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 hier also per Definition $\mathbb{P}(-t_2 \leq T \leq t_2) = \delta$.



(4) Etablierung der Konfidenzbedingung (fortgeführt)

Mit der Definition von t_2 wie oben folgt dann aber

$$\begin{aligned}\delta &= \mathbb{P}(-t_2 \leq T \leq t_2) \\ &= \mathbb{P}\left(-t_2 \leq \frac{\sqrt{n}}{S}(\bar{v} - \mu) \leq t_2\right) \\ &= \mathbb{P}\left(-\frac{S}{\sqrt{n}}t_2 \leq \bar{v} - \mu \leq \frac{S}{\sqrt{n}}t_2\right) \\ &= \mathbb{P}\left(-\bar{v} - \frac{S}{\sqrt{n}}t_2 \leq -\mu \leq -\bar{v} + \frac{S}{\sqrt{n}}t_2\right) \\ &= \mathbb{P}\left(\bar{v} + \frac{S}{\sqrt{n}}t_2 \geq \mu \geq \bar{v} - \frac{S}{\sqrt{n}}t_2\right) \\ &= \mathbb{P}\left(\bar{v} - \frac{S}{\sqrt{n}}t_2 \leq \mu \leq \bar{v} + \frac{S}{\sqrt{n}}t_2\right) \\ &= \mathbb{P}\left(\left[\bar{v} - \frac{S}{\sqrt{n}}t_2, \bar{v} + \frac{S}{\sqrt{n}}t_2\right] \ni \mu\right).\end{aligned}\tag{5}$$

Theorem (Konfidenzintervall für den Erwartungswertparameter)

Gegeben sei das Normalverteilungsmodell

$$v_1, \dots, v_n \sim N(\mu, \sigma^2) \quad (6)$$

mit wahren, aber unbekanntem, Parametern μ und σ^2 , es sei $\delta \in]0, 1[$ und es sei

$$t_\delta := \Psi^{-1} \left(\frac{1 + \delta}{2}; n - 1 \right). \quad (7)$$

mit der inversen KVF Ψ^{-1} einer t -verteilten Zufallsvariable. Dann gilt für das Intervall

$$\kappa(v) := \left[\bar{v} - \frac{S}{\sqrt{n}} t_\delta, \bar{v} + \frac{S}{\sqrt{n}} t_\delta \right], \quad (8)$$

mit dem Stichprobenmittel und der Stichprobenstandardabweichung

$$\bar{v} := \frac{1}{n} \sum_{i=1}^n v_i \quad \text{und} \quad S := \sqrt{\frac{1}{n-1} \sum_{i=1}^n (v_i - \bar{v})^2}, \quad (9)$$

respektive, dass

$$\mathbb{P}_\mu(\kappa(v) \ni \mu) = \delta. \quad (10)$$

Bemerkung

- Der Beweis dieses Theorems ist durch obige Konstruktion gegeben.
- κ ist ein zufälliges Intervall, weil \bar{v} und S Zufallsvariablen sind.

Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

Simulation der ersten Interpretation eines Konfidenzintervalls

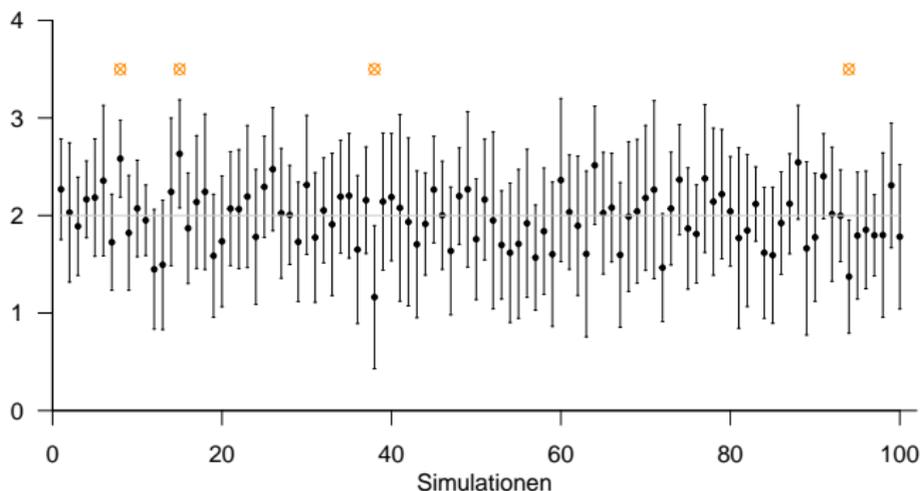
```
# Modellformulierung
set.seed(1) # random number generator seed
mu = 2 # w.a.u. Erwartungswertparameter
sigsqr = 1 # w.a.u. Varianzparameter
sigma = sqrt(sigsqr) # w.a.u. St.Abweichungsparameter
n = 12 # Stichprobenumfang
delta = 0.95 # Konfidenzbedingung
t_delta = qt((1+delta)/2,n-1) # \Psi^{-1}((\delta + 1)/2, n-1)

# Simulation
ns = 1e2 # Anzahl Simulationen
y_bar = rep(NA,n) # Stichprobenmittelarray
S = rep(NA,n) # St.Abweichungsarray
kappa = matrix(rep(NA,2*ns), ncol = 2) # Konfidenzintervallarray
for(i in 1:ns){
  y = rnorm(n,mu,sigma) # Stichprobenrealisierung
  y_bar[i] = mean(y) # Stichprobenmittel
  S[i] = sd(y) # Stichprobenstandardabweichung
  kappa[i,1] = y_bar[i] - (S[i]/sqrt(n))*t_delta # untere Konfidenzintervallgrenze
  kappa[i,2] = y_bar[i] + (S[i]/sqrt(n))*t_delta # obere Konfidenzintervallgrenze
}
```

Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

Simulation der ersten Interpretation eines Konfidenzintervalls

$$\mu = 2 \quad \sigma^2 = 2, \quad n = 12, \quad \delta = 0.95$$



Simulation der zweiten Interpretation eines Konfidenzintervalls

```
# Anzahl Simulationen mit \theta_1, \theta_2, ...
set.seed(2) # random number generator seed
ns = 1e2 # Anzahl Simulationen

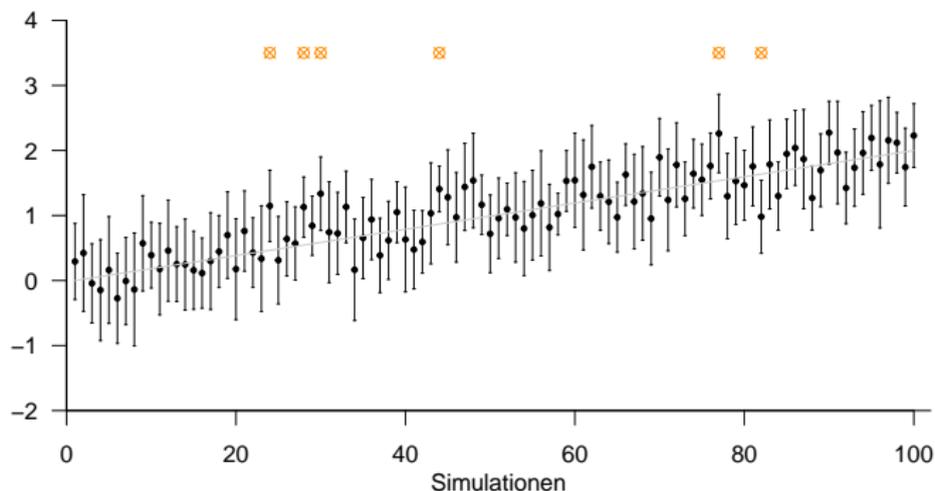
# Modellformulierung
mu = 2 * seq(0,1,len = ns) # w.a.u. Erwartungswertparameter
sigsqr = 1 # w.a.u. Varianzparameter
sigma = sqrt(sigsqr) # w.a.u. St.Abweichungsparameter
n = 12 # Stichprobenumfang
delta = 0.95 # Konfidenzbedingung
t_delta = qt((1+delta)/2,n-1) # \Psi^{-1}((\delta + 1)/2, n-1)

# Simulation
y_bar = rep(NaN,ns) # Stichprobenmittelarray
S = rep(NaN,ns) # St.Abweichungsarray
kappa = matrix(rep(NaN,2*ns), ncol = 2) # Konfidenzintervallarray
for(i in 1:ns){
  y = rnorm(n,mu[i],sigma) # Stichprobenrealisierung
  y_bar[i] = mean(y) # Stichprobenmittel
  S[i] = sd(y) # Stichprobenstandardabweichung
  kappa[i,1] = y_bar[i] - (S[i]/sqrt(n))*t_delta # untere Konfidenzintervallgrenze
  kappa[i,2] = y_bar[i] + (S[i]/sqrt(n))*t_delta # obere Konfidenzintervallgrenze
}
```

Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

Simulation der zweiten Interpretation eines Konfidenzintervalls

$$\mu = 2 \quad \sigma^2 = 2, \quad n = 12, \quad \delta = 0.95$$



Definition

Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

Anwendungsbeispiel

Selbstkontrollfragen

Definition

Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

Anwendungsbeispiel

Selbstkontrollfragen

Anwendungsbeispiel

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression



BDI-II Fragebogen

Name: _____ Alter: _____ Geschlecht: _____ Datum: _____

Anleitung: Dieser Fragebogen enthält 21 Gruppen von Aussagen. Bitte lesen Sie jede dieser Gruppen von Aussagen sorgfältig durch und wählen Sie sich dann in jeder Gruppe **einer Aussage** heraus, die am besten beschreibt, wie Sie sich in den letzten zwei Wochen **überwiegend** gefühlt haben. **Wichtig!** Beachten Sie, die Zeit refers die Aussage an, die für sich herausgehoben haben (1, 2 oder 3). Falls in einer Gruppe mehrere Aussagen gleichberechtigt auf Sie zuzutreffen, können Sie die Aussage wählen, die für Sie den besten Fall an. Achten Sie bitte darauf, dass Sie in jeder Gruppe nicht mehr als eine Aussage auswählen, die gilt noch für Gruppen in verschiedenen der Subskalienscores oder Gruppen in Dimensionen des Äquivalenz.

<p>1.) Traurigkeit</p> <ul style="list-style-type: none">0 Ich bin nicht traurig.1 Ich bin ein bisschen traurig.2 Ich bin so traurig, wie gewöhnlich, dass ich mir nicht darüber Gedanken mache.	<p>6.) Besorgungsgefühle</p> <ul style="list-style-type: none">0 Ich habe nicht das Gefühl, für etwas besorgt zu sein.1 Ich habe das Gefühl, vielleicht besorgt zu sein.2 Ich bin besorgt, besorgt zu sein.3 Ich habe das Gefühl, besorgt zu sein.
<p>2.) Zukunftsangst</p> <ul style="list-style-type: none">0 Ich sehe nicht mehr in die Zukunft, ich sehe nur auf das, was ich gerade mache.1 Ich bin etwas nervös und ängstlich, dass meine Situation besser wird.2 Ich bin etwas nervös, dass meine Zukunft hoffnungslos ist und nur noch schlechter wird.	<p>7.) Selbstabwertung</p> <ul style="list-style-type: none">0 Ich habe mir nie gemacht und wie immer.1 Ich habe mir manchmal ein bisschen weniger angetan.2 Ich bin mir ein bisschen weniger wertig.3 Ich habe mir viel angetan.
<p>3.) Verantwortungsgefühle</p> <ul style="list-style-type: none">0 Ich fühle mich nicht als Versager.1 Ich habe manchmal Verantwortungsgefühle.2 Wenn ich zurückblicke, sehe ich eine Menge Versäumnisse.3 Ich habe das Gefühl, als Mensch ein völliger Versager zu sein.	<p>8.) Selbstverurteilung</p> <ul style="list-style-type: none">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 Schlechte, was passiert.
<p>4.) Verlust von Freude</p> <ul style="list-style-type: none">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.	<p>9.) Selbstverleugern</p> <ul style="list-style-type: none">0 Ich denke nicht daran, mir etwas anzutun.1 Ich denke manchmal an Selbstmord, aber ich würde es nicht tun.2 Ich würde mich am liebsten umbringen, wenn ich die Gelegenheit dazu hätte.3 Ich würde mich umbringen, wenn ich könnte.
<p>5.) Schuldgefühle</p> <ul style="list-style-type: none">0 Ich habe keine besonderen Schuldgefühle.1 Ich habe ein bisschen Schuldgefühle wegen Dingen, die ich getan habe oder hätte tun sollen.2 Ich habe die meisten Zeit Schuldgefühle.3 Ich habe ständig Schuldgefühle.	<p>10.) Wut</p> <ul style="list-style-type: none">0 Ich empfinde nicht mehr als Ärger.1 Ich empfinde manchmal Ärger.2 Ich empfinde oft Ärger.3 Ich empfinde sehr oft Ärger, aber ich kann nicht.

PROBENSKIZZE © Copyright 2001 by Ed. Test Corporation

⇒ Pre-Post BDI Score Reduktion

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

i	BDI.Reduktion
1	-1
2	3
3	-2
4	9
5	3
6	-2
7	4
8	5
9	5
10	1
11	9
12	4

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Für die Pre-Post BDI Score Reduktion v_i der i ten von n Patient:innen legen wir das Modell

$$v_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, \dots, n \quad (11)$$

zugrunde. Dabei wird die Pre-Post BDI Reduktion v_i der i ten Patient:in also mithilfe einer über die Gruppe von Patient:innen identischen Pre-Post BDI Score Reduktion $\mu \in \mathbb{R}$ und einer Patient:innen-spezifischen normalverteilten Pre-Post BDI Score Reduktionsabweichung ε_i erklärt

Wie gezeigt ist dieses Modell äquivalent zum Normalverteilungsmodell

$$v_1, \dots, v_n \sim N(\mu, \sigma^2). \quad (12)$$

Die Standardproblemstellungen der Frequentistischen Inferenz führen dann auf folgende Fragen:

- (1) Was sind sinnvolle Tipps für die wahren, aber unbekanntenen, Parameterwerte μ und σ^2 ?
- (2) Wie kann im Sinne einer Intervallschätzung eine möglichst sichere Schätzung von μ gelingen?
- (3) Entscheiden wir uns sinnvollerweise für die Hypothese, dass gilt $\mu \neq 0$?

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Wir haben in (10) Parameterschätzung gesehen, dass unverzerzte Schätzer für den Erwartungswertparameter μ und den Varianzparameter σ^2 durch das Stichprobenmittel und die Stichprobenvarianz gegeben sind.

```
mu_hat      = mean(y)      # Stichprobenmittel als Erwartungswertparameterschätzer
sigsqr_hat  = var(y)       # Stichprobenvarianz als Varianzparameterschätzer
```

Es sind also $\hat{\mu} = 3.17$ und $\hat{\sigma}^2 = 13.8$ sinnvolle Tipps für μ und σ^2 basierend auf den vorliegenden 12 Datenpunkten.

Um eine möglichst sichere Schätzung für μ zu erlangen, geht die Frequentistische Inferenz zur Intervallschätzung mithilfe von δ -Konfidenzintervallen über, für die die assoziierte Unsicherheit dann ein prädefiniertes Level von $1 - \delta$ hat, im langfristigen Mittel überdeckt ein angegebenes δ -Konfidenzintervall den wahren, aber unbekanntem Parameterwert in (nur) $1 - \delta$ 100 von 100 Fällen nicht. Für ein großes δ wie $\delta = 0.95$ ist die mit dieser Intervallschätzung assoziierte Sicherheit also eher hoch, die assoziierte Unsicherheit eher gering.

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

```
# Konfidenzintervall für den Erwartungswertparameter
delta   = 0.95                # Konfidenzlevel
n       = length(y)          # Anzahl Datenpunkte
t_delta = qt((1+delta)/2,n-1) # \psi^{-1}((\delta + 1)/2, n-1)
y_bar   = mean(y)            # Stichprobenmittel
s       = sd(y)              # Stichprobenstandardabweichung
mu_hat  = y_bar              # Erwartungswertparameterschätzer
mu_hat_u = y_bar - (s/sqrt(n))*t_delta # untere Konfidenzintervallgrenze
mu_hat_o = y_bar + (s/sqrt(n))*t_delta # obere Konfidenzintervallgrenze
```

Das 0.95-Konfidenzintervall für den Erwartungswertparameter ist $[0.80, 5.52]$. Im langfristigen Mittel überdeckt so berechnetes Konfidenzintervall den wahren, aber unbekanntem, Erwartungswertparameter in 95 von 100 Fällen.

Definition

Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

Anwendungsbeispiel

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die Definition des Begriffs eines δ -Konfidenzintervalls wieder.
2. Erläutern Sie die zwei Interpretationen eines δ -Konfidenzintervalls.
3. Erläutern Sie die typischen Schritte zur Konstruktion eines δ -Konfidenzintervalls.
4. Geben Sie das Theorem zum δ -Konfidenzintervall für den Erwartungswert der Normalverteilung wieder.

1. Siehe Definition (δ -Konfidenzintervall).
2. Siehe Folie 9, *Zwei Interpretationen von δ -Konfidenzintervallen*.
3. Siehe Folie 10, *Allgemeine Konstruktion von Konfidenzintervallen*.
4. Siehe Theorem (Konfidenzintervall für den Erwartungswertparameter).



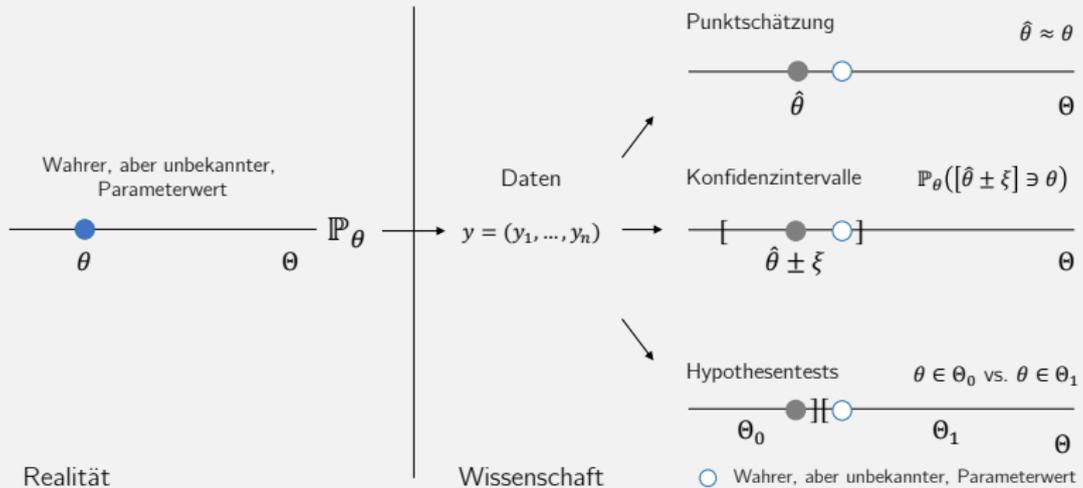
Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2023/24

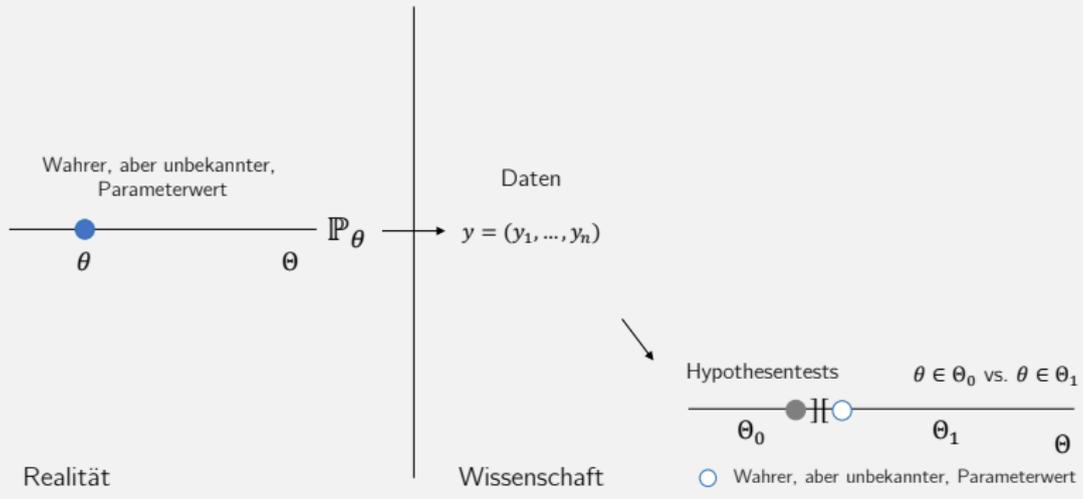
Prof. Dr. Dirk Ostwald

(11) Hypothesentests

Standardproblemstellungen und Standardannahme Frequentistischer Inferenz



Standardproblemstellungen und Standardannahme Frequentistischer Inferenz



Standardannahme Frequentistischer Inferenz

\mathcal{M} sei ein Frequentistisches Inferenzmodell mit Zufallsvektor v . Es wird angenommen, dass ein Datensatz $y \in \mathbb{R}^n$ eine der möglichen Realisierungen von v ist.

Aus Frequentistischer Sicht kann man eine Studie unendlich oft wiederholen und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. das Stichprobenmittel:

$$\text{Datensatz (1)} : y^{(1)} = \left(y_1^{(1)}, y_2^{(1)}, \dots, y_n^{(1)} \right) \text{ mit } \bar{y}^{(1)} = \frac{1}{n} \sum_{i=1}^n y_i^{(1)}$$

$$\text{Datensatz (2)} : y^{(2)} = \left(y_1^{(2)}, y_2^{(2)}, \dots, y_n^{(2)} \right) \text{ mit } \bar{y}^{(2)} = \frac{1}{n} \sum_{i=1}^n y_i^{(2)}$$

$$\text{Datensatz (3)} : y^{(3)} = \left(y_1^{(3)}, y_2^{(3)}, \dots, y_n^{(3)} \right) \text{ mit } \bar{y}^{(3)} = \frac{1}{n} \sum_{i=1}^n y_i^{(3)}$$

$$\text{Datensatz (4)} : y^{(4)} = \left(y_1^{(4)}, y_2^{(4)}, \dots, y_n^{(4)} \right) \text{ mit } \bar{y}^{(4)} = \frac{1}{n} \sum_{i=1}^n y_i^{(4)}$$

$$\text{Datensatz (5)} : y^{(5)} = \dots$$

Um die Qualität ihrer Methoden zu beurteilen betrachtet die Frequentistische Inferenz deshalb die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken. Was zum Beispiel ist die Verteilung der $\bar{y}^{(1)}, \bar{y}^{(2)}, \bar{y}^{(3)}, \bar{y}^{(4)}, \dots$ also die Verteilung der Zufallsvariable \bar{v}_n ?

Wenn eine Methode im Sinne der Frequentistischen Standardannahme "gut" ist, dann heißt das also, dass sie bei häufiger Anwendung "im Mittel gut" ist. Im Einzelfall, also im Normalfall nur eines vorliegenden Datensatzes, kann sie auch "schlecht" sein.

Grundlegende Logik Frequentistischer Hypothesentests

Man hat einen Datensatz y_1, \dots, y_n vorliegen und nimmt an, dass es sich dabei um die Realisation einer Stichprobe handelt, zum Beispiel von $v_1, \dots, v_n \sim N(\mu, \sigma^2)$. Man berechnet basierend auf dem Datensatz eine *Teststatistik*, zum Beispiel das anhand der Stichprobenstandardabweichung s und dem Stichprobenumfang n normalisierte Stichprobenmittel $\bar{y}, \sqrt{n} \frac{\bar{y}}{s}$.

Man fragt sich, wie wahrscheinlich es wäre, den beobachteten oder einen extremeren Wert der Teststatistik unter der Annahme eines *Nullmodells* zu observieren. Dabei meint man mit *Nullmodell* intuitiv ein Wahrscheinlichkeitsverteilungsmodell, bei dem kein "interessanter Effekt" vorliegt, also zum Beispiel $\mu = 0$ gilt. Die Wahrscheinlichkeit ist dabei natürlich frequentistisch zu verstehen, also als idealisierte relative Häufigkeit, über viele viele Stichprobenrealisationen.

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 in einem konkret vorliegenden Fall nun wirklich das Nullmodell oder ein anderes Modell den Datensatz generiert hat. Man kann aber durch die entsprechende Konstruktion der Prozedur sicherstellen, dass man im langfristigen Mittel sinnvolle Entscheidungen trifft, wenn die Annahmen der Prozedur zutreffen und man die Prozedur sehr oft wiederholt.

Testhypothesen und Tests

Testgütekriterien und Testkonstruktion

Einstichproben-T-Test

Anwendungsbeispiel

Konfidenzintervalle und Hypothesentests

Selbstkontrollfragen

Testhypothesen und Tests

Testgütekriterien und Testkonstruktion

Einstichproben-T-Test

Anwendungsbeispiel

Konfidenzintervalle und Hypothesentests

Selbstkontrollfragen

Definition (Testhypothesen und Testszenario)

Gegeben sei ein Frequentistisches Inferenzmodell mit Stichprobe v , Ergebnisraum \mathcal{Y} und Parameterraum Θ . Weiterhin sei $\{\Theta_0, \Theta_1\}$ eine Partition des Parameterraums, so dass

$$\Theta = \Theta_0 \cup \Theta_1 \text{ und } \Theta_0 \cap \Theta_1 = \emptyset. \quad (1)$$

Dann ist eine *Testhypothese* eine Aussage über den wahren, aber unbekanntem, Parameterwert θ in Hinblick auf die Untermengen Θ_0 und Θ_1 des Parameterraums. Speziell werden die Aussagen

- $\theta \in \Theta_0$ als *Nullhypothese* und
- $\theta \in \Theta_1$ als *Alternativhypothese*

bezeichnet. Der Einfachheit halber bezeichnet man auch Θ_0 und Θ_1 direkt als Nullhypothese und Alternativhypothese, respektive. Die Einheit aus Frequentistischem Inferenzmodell und Testhypothesen wird als *Testszenario* bezeichnet.

Definition (Einfache und zusammengesetzte Testhypothesen)

Für die Testhypothesen Θ_i mit $i = 0, 1$ 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 Wahrscheinlichkeitsverteilung von v genau festgelegt.
- Bei einer zusammengesetzten Hypothese ist nur die Verteilungsklasse von v festgelegt.

Definition (Einseitige und zweiseitige Testhypothesen)

Gegeben sei ein Testszenario mit eindimensionalem Parameteraum $\Theta := \mathbb{R}$ und es sei $\theta_0 \in \Theta$.

- Dann werden zusammengesetzte Nullhypothesen der Form $\Theta_0 :=]-\infty, \theta_0]$ oder $\Theta_0 := [\theta_0, \infty[$ *einseitige Nullhypothesen* genannt und auch in der Form $H_0 : \theta \leq \theta_0$ bzw. $H_0 : \theta \geq \theta_0$ geschrieben. Die entsprechenden Alternativhypothesen haben dabei die Form $\Theta_1 :=]\theta_0, \infty[$ bzw. $\Theta_1 :=]-\infty, \theta_0[$, auch geschrieben als $H_1 : \theta > \theta_0$ bzw. $H_1 : \theta < \theta_0$.
- Bei einer einfachen Nullhypothese der Form $\Theta_0 := \{\theta_0\}$, auch geschrieben als $H_0 : \theta = \theta_0$, wird die Alternativhypothese $\Theta_1 := \Theta \setminus \{\theta_0\}$, auch geschrieben als $H_1 : \theta \neq \theta_0$, *zweiseitige Alternativhypothese* genannt.

Definition (Test)

Gegeben sei ein Testszenario. Dann ist ein *Test* eine Abbildung ϕ aus dem Ergebnisraum der Stichprobe \mathcal{Y} in die Menge $\{0, 1\}$, also

$$\phi : \mathcal{Y} \rightarrow \{0, 1\}, y \mapsto \phi(y), \quad (2)$$

wobei

- $\phi(y) = 0$ den Vorgang des Nichtablehnens der Nullhypothese und
- $\phi(y) = 1$ den Vorgang des Ablehnens der Nullhypothese

repräsentieren.

Bemerkung

- Weil y eine Realisation von v ist, ist $\phi(y)$ eine Realisation von $\phi(v)$.

Definition (Standardtest)

Gegeben sei ein Testszenario. Dann ist ein *Standardtest* definiert als die Verkettung einer *Teststatistik*

$$\gamma : \mathcal{Y} \rightarrow \mathbb{R} \quad (3)$$

und einer *Entscheidungsregel*

$$\delta : \mathbb{R} \rightarrow \{0, 1\}. \quad (4)$$

Ein Standardtest kann also geschrieben werden als

$$\phi := \delta \circ \gamma : \mathcal{Y} \rightarrow \{0, 1\}. \quad (5)$$

Bemerkungen

- Weil y eine Realisation von v ist, ist $\gamma(y) \in \mathbb{R}$ eine Realisation von $\gamma(v)$.
- Weil $\gamma(y)$ eine Realisation von $\gamma(v)$ ist, ist $(\delta \circ \gamma)(y)$ eine Realisation von $(\delta \circ \gamma)(v)$.
- Wir betrachten in der Folge nur Standardtests.

Definition (Kritischer Bereich)

Gegeben sei ein Testszenario und ein Test ϕ . Dann heißt die Untermenge K des Ergebnisraums \mathcal{Y} der Stichprobe v , für die der Test den Wert 1 annimmt, *kritischer Bereich* des Tests, formal

$$K := \{y \in \mathcal{Y} | \phi(y) = 1\} \subset \mathcal{Y}. \quad (6)$$

Bemerkungen

- Die Ereignisse $\{\phi(v) = 1\}$ und $\{v \in K\}$ sind äquivalent.
- Die Ereignisse $\{\phi(v) = 1\}$ und $\{v \in K\}$ haben die gleiche Wahrscheinlichkeit.

Definition (Ablehnungsbereich)

Gegeben sei ein Testszenario und ein Standardtest ϕ mit Teststatistik γ . Die Untermenge A des Ergebnisraums der Teststatistik, für die der Test den Wert 1 annimmt, *Ablehnungsbereich des Tests*, formal

$$A := \{\gamma(y) \in \mathbb{R} \mid \phi(y) = 1\} \subset \mathbb{R}. \quad (7)$$

Bemerkungen

- Die Ereignisse $\{\phi(v) = 1\}$ und $\{\gamma(v) \in A\}$ sind äquivalent.
- Die Ereignisse $\{\phi(v) = 1\}$ und $\{\gamma(v) \in A\}$ 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 der Teststatistik mit Ergebnisraum \mathbb{R} abhängt. Speziell ist

- ein *einseitiger kritischer Wert-basierter Test* von der Form

$$\phi : \mathcal{Y} \rightarrow \{0, 1\}, y \mapsto \phi(y) := 1_{\{\gamma(y) \geq k\}} = \begin{cases} 1 & \gamma(y) \geq k \\ 0 & \gamma(y) < k \end{cases} \quad (8)$$

- ein *zweiseitiger kritischer Wert-basierter Test* von der Form

$$\phi : \mathcal{Y} \rightarrow \{0, 1\}, y \mapsto \phi(y) := 1_{\{|\gamma(y)| \geq k\}} = \begin{cases} 1 & |\gamma(y)| \geq k \\ 0 & |\gamma(y)| < k \end{cases} \quad (9)$$

Bemerkung

- Wir betrachten in der Folge nur kritischer Wert-basierte Tests.

Testhypothesen und Tests

Testgütekriterien und Testkonstruktion

Einstichproben-T-Test

Anwendungsbeispiel

Konfidenzintervalle und Hypothesentests

Selbstkontrollfragen

Definition (Richtige Testentscheidungen und Testfehler)

Gegeben seien ein Testscenario und ein Test.

- Dann gibt es mit dem Nichtablehnen der Nullhypothese $\phi(y) = 0$, wenn die Nullhypothese $\theta \in \Theta_0$ zutrifft und dem Ablehnen der Nullhypothese $\phi(y) = 1$, wenn die Alternativhypothese $\theta \in \Theta_1$ zutrifft zwei Formen der *richtigen Testentscheidung*.
- Ebenso gibt es zwei Arten von *Testfehlern*: Das Ablehnen der Nullhypothese $\phi(y) = 1$, wenn die Nullhypothese $\theta \in \Theta_0$ zutrifft, heißt *Typ I Fehler* und das Nichtablehnen der Nullhypothese, wenn die Alternativhypothese $\theta \in \Theta_1$ zutrifft, heißt *Typ II Fehler*.

Bemerkung

		Testentscheidung	
		$\phi(v) = 0$	$\phi(v) = 1$
Wahr, aber unbekannter, Parameter	$\theta \in \Theta_0$	Richtige Entscheidung	Typ I Fehler
	$\theta \in \Theta_1$	Typ II Fehler	Richtige Entscheidung

Definition (Testgütefunktion)

Gegeben sei ein Testszenario und ein Test ϕ . Das ist die *Testgütefunktion* von ϕ definiert als

$$q_\phi : \Theta \rightarrow [0, 1], \theta \mapsto q_\phi(\theta) := \mathbb{P}_\theta(\phi(v) = 1). \quad (10)$$

Für $\theta \in \Theta_1$ heißt q_ϕ auch *Powerfunktion* oder *Trennschärfefunktion*.

Bemerkungen

- \mathbb{P}_θ bezeichnet die Verteilung von ϕ unter der Annahme $v \sim \mathbb{P}_\theta$.
- Es gilt $\mathbb{P}_\theta(\phi(v) = 1) = \mathbb{P}_\theta(v \in K) = \mathbb{P}_\theta(\gamma \in A)$
- Für jedes $\theta \in \Theta$ liefert q_ϕ die Wahrscheinlichkeit, dass die Nullhypothese 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

$$q_\phi(\theta) = \mathbb{P}_\theta(\phi(v) = 1) = 0 \text{ für } \theta \in \Theta_0 \text{ und } q_\phi(\theta) = \mathbb{P}_\theta(\phi(v) = 1) = 1 \text{ für } \theta \in \Theta_1. \quad (11)$$

- Die Testentscheidung eines solchen ϕ wäre mit Wahrscheinlichkeit 1 richtig.

Intuition zur Testkonstruktion

Im Idealfall hätte man einen Test ϕ mit

$$q_\phi(\theta) = \mathbb{P}_\theta(\phi(v) = 1) = 0 \text{ für } \theta \in \Theta_0 \text{ und } q_\phi(\theta) = \mathbb{P}_\theta(\phi(v) = 1) = 1 \text{ für } \theta \in \Theta_1. \quad (12)$$

⇒ 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 $\phi_a(y) := 0$ für alle $y \in \mathcal{Y}$, 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 $\phi_b(y) := 1$ für alle $y \in \mathcal{Y}$, 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 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

$$q_\phi(\theta) \leq \alpha_0 \text{ für alle } \theta \in \Theta_0. \quad (13)$$

Eine konventionelle Wahl für sein solches α_0 ist zum Beispiel $\alpha_0 := 0.05$.

Unter allen Tests und statistischen Modellen, die Ungleichung (13) 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 (manchmal auch als *nominales Niveau* bezeichnet) und des Testumfangs α (manchmal auch als *effektives Niveau* bezeichnet).

Definition (Level- α_0 -Test, Signifikanzlevel α_0 , Testumfang α)

Gegeben seien ein Testszenario, ein Test ϕ , seine Testgütefunktion q_ϕ und ein $\alpha_0 \in [0, 1]$. ϕ heißt ein *Level- α_0 -Test*, wenn gilt, dass

$$q_\phi(\theta) \leq \alpha_0 \text{ für alle } \theta \in \Theta_0. \quad (14)$$

Wenn ϕ ein Level- α_0 -Test ist, nennt man den Wert α_0 auch das *Signifikanzlevel* des Tests. Weiterhin heißt die Zahl

$$\alpha := \max_{\theta \in \Theta_0} q_\phi(\theta) \in [0, 1] \quad (15)$$

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 $\alpha = q_\phi(\theta_0) = \mathbb{P}_{\theta_0}(\phi(v) = 1)$.

Typ I Fehlerwahrscheinlichkeit vs. Testumfang vs. Signifikanzlevel

Bei einfacher Θ_0 ist der Testumfang gleich der Wahrscheinlichkeit eines Typ I Fehlers

$$\alpha := \max_{\theta \in \Theta_0} q_\phi(\theta) = \max_{\theta \in \{\theta_0\}} q_\phi(\theta) = q_\phi(\theta_0) = \mathbb{P}_{\theta_0}(\phi(v) = 1). \quad (16)$$

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

$$\alpha := \max_{\theta \in \Theta_0} q_\phi(\theta) = \max_{\theta \in \Theta_0} \mathbb{P}_\theta(\phi(v) = 1). \quad (17)$$

Ein Test hat Signifikanzlevel α_0 , wenn der Testumfang kleiner oder gleich α_0 ist.

$$\alpha = \max_{\theta \in \Theta_0} q_\phi(\theta) \leq \alpha_0 \quad (18)$$

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*.

Ein Test, bei dem das Signifikanzlevel kleiner als dem Testumfang ist, heißt *liberal*.

Motivation des p-Wert Begriffs

Es werde ein zweiseitiger kritischer Wert-basierter Test durchgeführt.

- H_0 wird abgelehnt, wenn $|\gamma(v)| \geq k$.

Nehmen wir an, es werde $\gamma(v) = k + \frac{k}{100}$ beobachtet.

- Das Testergebnis lautet $\phi(v) = 1 \Leftrightarrow$ Ablehnen der Nullhypothese

Nehmen wir an, es werde $\gamma(v) = k + 100k$ beobachtet.

- Das Testergebnis lautet $\phi(v) = 1 \Leftrightarrow$ Ablehnen der Nullhypothese

Der Bericht des Testergebnis allein supprimiert potentiell interessante Information.

\Rightarrow Neben der Testumfangkontrolle durch beispielsweise $\alpha_0 = 0.05$ ist es üblich, alle Werte von α_0 anzugeben, für die ein Level- α_0 -Test zum Ablehnen der Nullhypothese führen würde.

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.

Bemerkung

- Eine Intuition zu dieser Definition ergibt sich im Rahmen des Einstichproben-T-Tests. p-Werte spiegeln dort 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 Nullmodells zu observieren.
- p-Werte sind extrem populär, aber auch sehr umstritten.
- 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 der vielen Möglichkeiten ein Signal-zu-Rauschen Verhältnis zu quantifizieren.
- p-Werte sind eine der vielen Möglichkeiten, Unsicherheit zu quantifizieren.

(vgl. Wasserstein, Schirm, and Lazar (2019))

Kommentar zum Frequentistischen 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. Pratt, Raiffa, and Schlaifer (1995), Puterman (2005), Ostwald, Starke, and Hertwig (2015))

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 Grundlagewissenschaften 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 und sollte deshalb im Rahmen eines wissenschaftlichen Studiums wie der Psychologie intellektuell durchdrungen werden.

Testhypothesen und Tests

Testgütekriterien und Testkonstruktion

Einstichproben-T-Test

Anwendungsbeispiel

Konfidenzintervalle und Hypothesentests

Selbstkontrollfragen

Einstichproben-T-Test

- (1) Anwendungsszenario
- (2) Frequentistisches Inferenzmodell
- (3) Testhypothesen
- (4) Definition der Teststatistik
- (5) Verteilung der Teststatistik
- (6) Testdefinition
- (7) Testgütefunktion
- (8) Testumfangkontrolle
- (9) p -Wert
- (10) Powerfunktion
- (11) Praktische Durchführung

Anwendungsszenario eines Einstichproben-T-Tests

Eine Stichprobe (Gruppe) randomisierter experimenteller Einheiten.

Annahme der unabhängigen und identischen Normalverteilung $N(\mu, \sigma^2)$ der Datenpunkte.

μ und σ^2 unbekannt.

Absicht der Inferenz hinsichtlich einer Nullhypothese und einer Alternativhypothese

Anwendungsbeispiele

Pre-Post-Psychotherapie BDI Differenzanalyse einer Gruppe von Patient:innen

- $\mu \neq \mu_0 := 0 \Rightarrow$ Evidenz für Depressionssymptomatikveränderung

Gruppenanalysen mit Wechsler Adult Intelligence Scale

- $\mu \neq \mu_0 := 100 \Rightarrow$ Evidenz für über- oder unterdurchschnittliche WAIS Performanz

Gruppenanalysen in der funktionellen Kernspintomographie

- $\mu > \mu_0 := 0 \Rightarrow$ Evidenz für regionale Gehirnaktivierung

Hypothesen- und Testszenarien bei Einstichproben-T-Tests

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

Hier betrachtetes Hypothesen- und Testszenario eines Einstichproben-T-Tests

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

Definition (Frequentistisches Inferenzmodell des Einstichproben-T-Tests)

Das Frequentistische Inferenzmodell des Einstichproben-T-Tests ist gegeben durch das Normalverteilungsmodell

$$v_1, \dots, v_n \sim N(\mu, \sigma^2) \text{ mit } (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0} \quad (19)$$

Bemerkung

- Wir erinnern daran, dass aus generativer Sicht das Normalverteilungsmodell dem Modell

$$v_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ für } i = 1, \dots, n \quad (20)$$

entspricht.

Definition (Testhypothesen des Einstichproben-T-Tests)

Gegeben sei das Frequentistische Inferenzmodell des Einstichproben-T-Tests

$$v_1, \dots, v_n \sim N(\mu, \sigma^2) \text{ mit } (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0} \quad (21)$$

und es sei $\Theta := \mathbb{R}$ der Parameterunterraum des Parameters von Interesse μ . Dann sind für den *Nullhypothesenparameterwert* $\mu_0 \in \mathbb{R}$ die *einfache Nullhypothese* und die *zusammengesetzte Alternativhypothese* des Einstichproben-T-Tests gegeben durch

$$\Theta_0 := \{\mu_0\} \Leftrightarrow H_0 : \mu = \mu_0 \text{ und } \Theta_1 := \mathbb{R} \setminus \{\mu_0\} \Leftrightarrow H_1 : \mu \neq \mu_0. \quad (22)$$

Bemerkung

- μ ist der wahre, aber unbekannte, Parameter, μ_0 ist der Nullhypothesenparameter.

Definition (Einstichproben-T-Test-Statistik)

Gegeben sei das Testscenario eines Einstichproben-T-Tests mit Stichprobe v_1, \dots, v_n , Stichprobenmittel \bar{v} , Stichprobenstandardabweichung S und Nullhypotesenparameter μ_0 . Dann ist die *Einstichproben-T-Test-Statistik* definiert als

$$T := \sqrt{n} \frac{\bar{v} - \mu_0}{S}. \quad (23)$$

Bemerkungen

- Wir kürzen den Begriff *Einstichproben-T-Test-Statistik* z.T. auch mit ETT-Statistik ab.
- Im Gegensatz zur T-Konfidenzintervallstatistik muss bei der ETT-Statistik nicht $\mu_0 = \mu$ gelten.
- Intuitiv kann die ETT-Statistik als mit dem Stichprobenumfang (Evidenz) gewichtetes Verhältnis von Signal (systematischer Variabilität) zu Rauschen (unsystematischer Variabilität) verstanden werden:

$$\sqrt{\text{Stichprobenumfang}} \left(\frac{\text{Signal}}{\text{Rauschen}} \right) = \sqrt{n} \frac{\bar{v} - \mu_0}{S} \quad (24)$$

- Die ETT-Statistik ist eine skalare Deskription des Effekt vs. Variabilität Verhältnisses eines Datensatzes.
- In der ETT-Statistik wird die Effektgröße in Einheiten der Stichprobenstandardabweichung gemessen:
 - $T = 1 \Leftrightarrow \sqrt{n}(\bar{v} - \mu_0) = 1S$
 - $T = 2 \Leftrightarrow \sqrt{n}(\bar{v} - \mu_0) = 2S$

Theorem (Verteilung der Einstichproben-T-Test-Statistik)

Gegeben sei das Testscenario eines Einstichproben-T-Tests mit Stichprobe v_1, \dots, v_n , Stichprobenmittel \bar{v} , Stichprobenstandardabweichung S , Nullhypothese parameter μ_0 und Einstichproben-T-Test-Statistik definiert als

$$T := \sqrt{n} \frac{\bar{v} - \mu_0}{S}. \quad (25)$$

Dann ist T eine nichtzentrale t -Zufallsvariable mit Nichtzentralitätsparameter

$$d = \sqrt{n} \frac{\mu - \mu_0}{\sigma} \quad (26)$$

und Freiheitsgradparameter $n - 1$, es gilt also $T \sim t(d, n - 1)$

Bemerkungen

- Wir verzichten auf einen Beweis

Definition (Nichtzentrale t -Zufallsvariable)

T sei eine Zufallsvariable mit Ergebnisraum \mathbb{R} und WDF

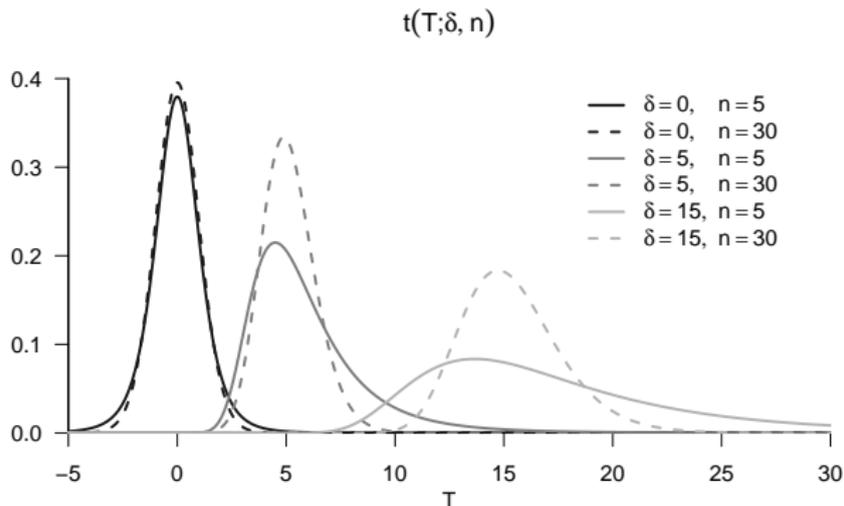
$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, t \mapsto p(t) := \frac{1}{2^{\frac{n-1}{2}} \Gamma\left(\frac{n}{2}\right) (n\pi)^{\frac{1}{2}}} \int_0^\infty \tau^{\frac{n-1}{2}} \exp\left(-\frac{\tau}{2}\right) \exp\left(-\frac{1}{2} \left(t \left(\frac{\tau}{n}\right)^{\frac{1}{2}} - \delta\right)^2\right) d\tau. \quad (27)$$

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 $t(T; 0, n) = t(T; n)$.
- Weiterhin gelten $\Psi(T; 0, n) = \Psi(T; n)$ und $\Psi^{-1}(T; 0, n) = \Psi^{-1}(T; n)$
- Die funktionale Form der WDF findet sich zum Beispiel in Lehmann (1986), Seite 254, Gl. (80).

Wahrscheinlichkeitsdichtefunktionen nichtzentraler t -Verteilungen



Theorem (Nichtzentrale T-Transformation)

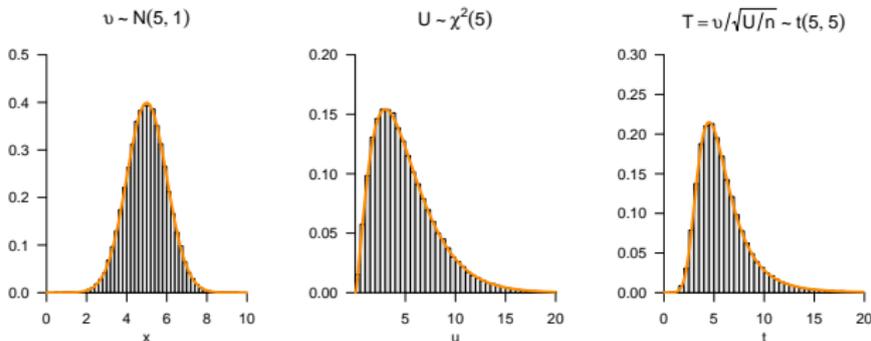
$v \sim N(\mu, 1)$ sei eine normalverteilte Zufallsvariable, $U \sim \chi^2(n)$ sei eine χ^2 Zufallsvariable mit Freiheitsgradparameter n , und v und U seien unabhängige Zufallsvariablen. Dann ist die Zufallsvariable

$$T := \frac{v}{\sqrt{U/n}} \quad (28)$$

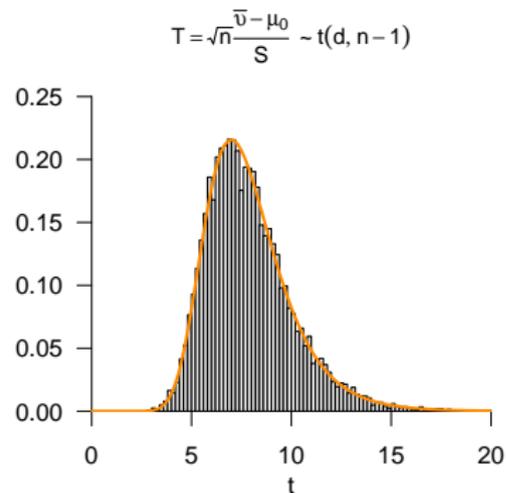
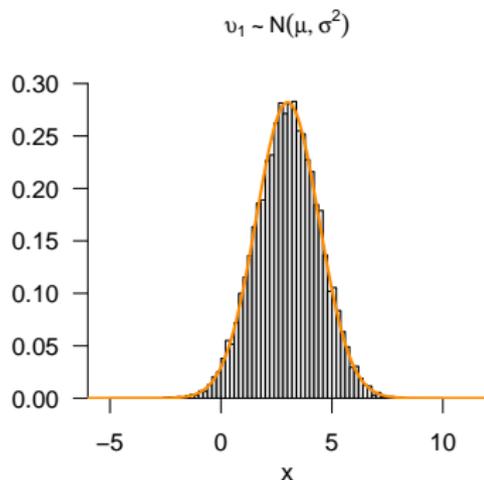
eine nichtzentrale t -Zufallsvariable mit Nichtzentralitätsparameter μ und Freiheitsgradparameter n , also $T \sim t(\mu, n)$.

Bemerkung

- Wir verzichten auf einen Beweis.



Einstichproben-T-Test-Statistik bei $n = 12, \mu = 3, \sigma^2 = 2, \mu_0 = 0$



Definition (Zweiseitiger Einstichproben-T-Test)

Gegeben seien das Frequentistische Inferenzmodell des Einstichproben-T-Tests mit einfacher Nullhypothese und zusammengesetzter Alternativhypothese und T bezeichne die Einstichproben-T-Test-Statistik mit Werten $t \in \mathbb{R}$. Dann ist der *zweiseitige Einstichproben-T-Test mit einfacher Nullhypothese und zusammengesetzter Alternativhypothese* definiert als der zweiseitige kritische Wert-basierte Test

$$\phi : \mathcal{Y} \rightarrow \{0, 1\}, y \mapsto \phi(y) := 1_{\{|t| \geq k\}} = \begin{cases} 1 & |t| \geq k \\ 0 & |t| < k \end{cases}. \quad (29)$$

Theorem (Testgütefunktion des Einstichproben-T-Test)

ϕ sei der zweiseitige Einstichproben-T-Test mit einfacher Nullhypothese und zusammengesetzter Alternativhypothese. Dann ist die Testgütefunktion von ϕ gegeben durch

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \Psi(k; d_\mu, n - 1) + \Psi(-k; d_\mu, n - 1), \quad (30)$$

wobei $\Psi(\cdot; d_\mu, n - 1)$ die KVF der nichtzentralen t -Verteilung mit Nichtzentralitätsparameter

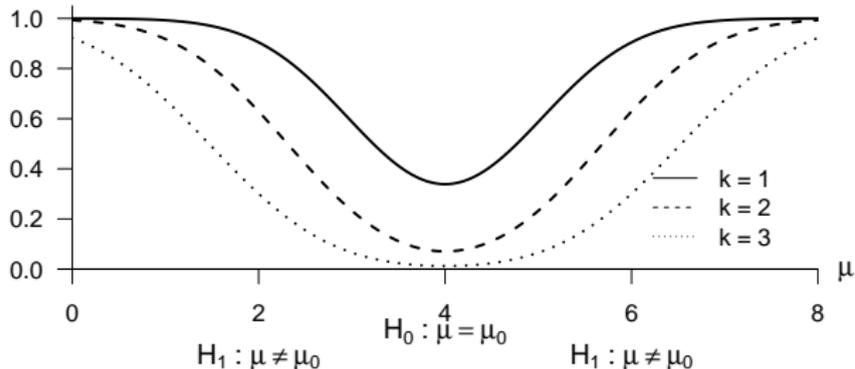
$$d_\mu := \sqrt{n} \frac{\mu - \mu_0}{\sigma} \quad (31)$$

und Freiheitsgradparameter $n - 1$ bezeichnet.

Bemerkungen

- Wir visualisieren die Testgütefunktion unten in Abhängigkeit von k .

$$q_\phi(\mu) = \mathbb{P}_\mu(\phi(v) = 1) \text{ für } \sigma^2 = 9, \mu_0 = 4, n = 12 \text{ und } k = 1, 2, 3$$



Beweis

Die Testgütefunktion des betrachteten Test im vorliegenden Testszenario ist definiert als

$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := \mathbb{P}_\mu(\phi = 1). \quad (32)$$

Da die Wahrscheinlichkeiten für $\phi = 1$ und dafür, dass die zugehörige Teststatistik im Ablehnungsbereich des Tests liegt gleich sind, benötigen wir also zunächst die Verteilung der Teststatistik. Wir haben oben bereits gesehen, dass die Einstichproben-T-Test-Statistik

$$T := \sqrt{n} \frac{\bar{v} - \mu_0}{S} \quad (33)$$

unter der Annahme $v_1, \dots, v_n \sim N(\mu, \sigma^2)$ anhand einer nichtzentralen t -Verteilung $t(d_\mu, n - 1)$ mit Nichtzentralitätsparameter

$$d_\mu := \sqrt{n} \frac{\mu - \mu_0}{\sigma} \quad (34)$$

verteilt ist. Der Ablehnungsbereich des zweiseitigen Einstichproben-T-Tests ist

$$A =] - \infty, -k] \cup]k, \infty[. \quad (35)$$

Beweis (fortgeführt)

Mit diesem Ablehnungsbereich ergibt sich dann

$$\begin{aligned}q_{\phi}(\mu) &= \mathbb{P}_{\mu}(\phi = 1) \\&= \mathbb{P}_{\mu}(T \in]-\infty, -k] \cup]k, \infty[) \\&= \mathbb{P}_{\mu}(T \in]-\infty, -k]) + \mathbb{P}_{\mu}(T \in [k, \infty[) \\&= \mathbb{P}_{\mu}(T \leq -k) + \mathbb{P}_{\mu}(T \geq k) \\&= \mathbb{P}_{\mu}(T \leq -k) + (1 - \mathbb{P}_{\mu}(T \leq k)) \\&= 1 - \mathbb{P}_{\mu}(T \leq k) + \mathbb{P}_{\mu}(T \leq -k) \\&= 1 - \Psi(k; d_{\mu}, n - 1) + \Psi(-k; d_{\mu}, n - 1),\end{aligned}\tag{36}$$

wobei $\Psi(\cdot; d_{\mu}, n - 1)$ die KVF der nichtzentralen T-Verteilung mit Nichtzentralitätsparameter d_{μ} und Freiheitsgradparameter $n - 1$ bezeichnet.

□

Theorem (Testumfangkontrolle für den Einstichproben-T-Test)

ϕ sei der zweiseitige Einstichproben-T-Test mit einfacher Nullhypothese und zusammengesetzter Alternativhypothese. Dann ist ϕ ein Level- α_0 -Test mit Testumfang α_0 , wenn der kritische Wert definiert ist durch

$$k_{\alpha_0} := \Psi^{-1}\left(1 - \frac{\alpha_0}{2}; n - 1\right), \quad (37)$$

wobei $\Psi^{-1}(\cdot; n - 1)$ die inverse KVF der t -Verteilung mit Freiheitsgradparameter $n - 1$ bezeichnet.

Bemerkung

- In \mathbf{R} kann Ψ^{-1} mit der Funktion `qt()` ausgewertet werden.

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

$$\begin{aligned} q_\phi(\mu_0) &= 1 - \Psi(k; d_{\mu_0}, n-1) + \Psi(-k; d_{\mu_0}, n-1) \\ &= 1 - \Psi(k; 0, n-1) + \Psi(-k; 0, n-1) \\ &= 1 - \Psi(k; n-1) + \Psi(-k; n-1), \end{aligned} \tag{38}$$

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.

Beweis (fortgeführt)

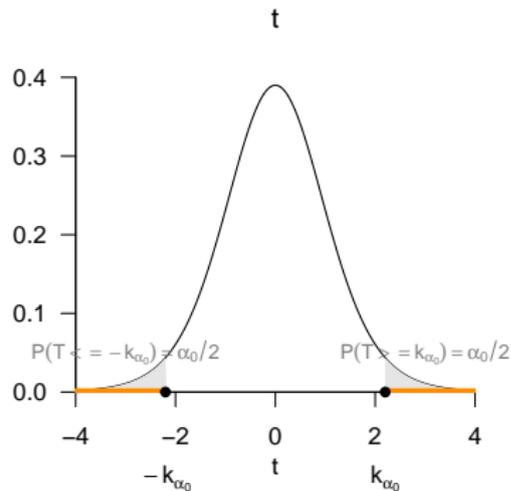
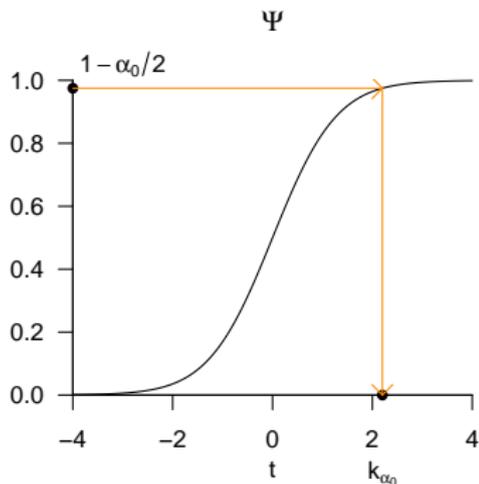
Sei nun also $k := k_{\alpha_0}$. Dann gilt

$$\begin{aligned}q_{\phi}(\mu_0) &= 1 - \Psi(k_{\alpha_0}; n-1) + \Psi(-k_{\alpha_0}; n-1) \\&= 1 - \Psi(k_{\alpha_0}; n-1) + (1 - \Psi(k_{\alpha_0}; n-1)) \\&= 2(1 - \Psi(k_{\alpha_0}; n-1)) \\&= 2\left(1 - \Psi\left(\Psi^{-1}\left(1 - \frac{\alpha_0}{2}, n-1\right), n-1\right)\right) \\&= 2\left(1 - 1 + \frac{\alpha_0}{2}\right) \\&= \alpha_0,\end{aligned}\tag{39}$$

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 | (8) Testumfangkontrolle

Wahl von $k_{\alpha_0} := \Psi^{-1}(1 - \frac{\alpha_0}{2}; n - 1)$ mit $n = 12$, $\alpha_0 := 0.05$ und Ablehnungsbereich



Simulation

```
n          = 12          # Anzahl der Datenpunkte
mu         = 0           # wahrer, aber unbekannter, Erwartungswertparameter
sigsqr    = 2           # wahrer, aber unbekannter, Varianzparameter
mu_0      = 0           # Nullhypothesenparameter, hier \mu = \mu_0
alpha_0   = 0.05        # Signifikanzlevel
k_alpha_0 = qt(1-alpha_0/2,n-1) # Kritischer Wert
set.seed(1)          # Random number generator seed
nsim      = 1e5         # Anzahl Simulationen
phi       = rep(NaN,nsim) # Testentscheidungsarray
for(j in 1:nsim){    # Simulationsiterationen
  y       = rnorm(n,mu,sigsqr) # \ups_i \sim N(\mu,\Sigma), i = 1,...,n
  y_bar   = mean(y)          # Stichprobenmittel
  s       = sd(y)           # Stichprobenstandardabweichung
  Tee     = sqrt(n)*((y_bar - mu_0)/s) # Einstichproben-T-Test-Statistik
  if(abs(Tee) >= k_alpha_0){ # Test  $1_{\{|t| \geq k_{\alpha_0}\}}$ 
    phi[j] = 1              # Ablehnen der Nullhypothese
  } else {
    phi[j] = 0              # Nichablehnen der Nullhypothese
  }
}
cat("Kritischer Wert          =", k_alpha_0,
    "\nGeschätzter Testumfang alpha =", mean(phi))
```

Kritischer Wert = 2.200985

Geschätzter Testumfang alpha = 0.0493

Theorem (p-Wert)

Gegeben sei der zweiseitige Einstichproben-T-Test mit einfacher Nullhypothese und zusammengesetzter Alternativhypothese und t sei ein Wert der Einstichproben-T-Test-Statistik T . Dann gilt

$$\text{p-Wert} = 2(1 - \Psi(|t|; n - 1)) \quad (40)$$

wobei $\Psi(\cdot; n - 1)$ die KVF der t -Verteilung mit Freiheitsgradparameter $n - 1$ bezeichnet.

Bemerkung

- In **R** kann Ψ mit der Funktion `pt()` ausgewertet werden.

Einstichproben-T-Test | (9) p-Wert

Beweis

Per Definition ist der p-Wert das kleinste Signifikanzlevel α_0 bei dem für den betrachteten Test die Nullhypothese basierend auf dem Wert von t abgelehnt werden würde. Im vorliegenden Fall würde die Nullhypothese für jedes α_0 mit

$$|t| \geq \Psi^{-1}\left(1 - \frac{\alpha_0}{2}; n - 1\right) \quad (41)$$

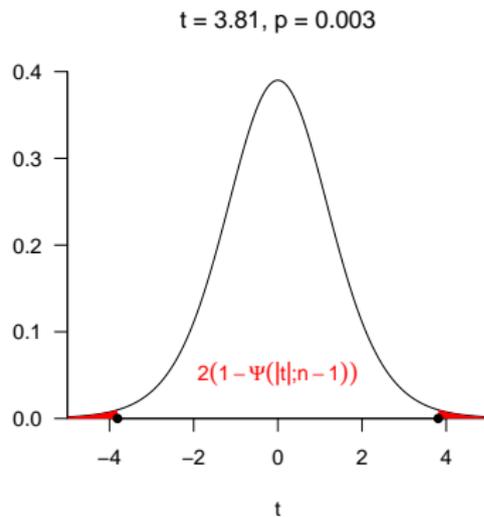
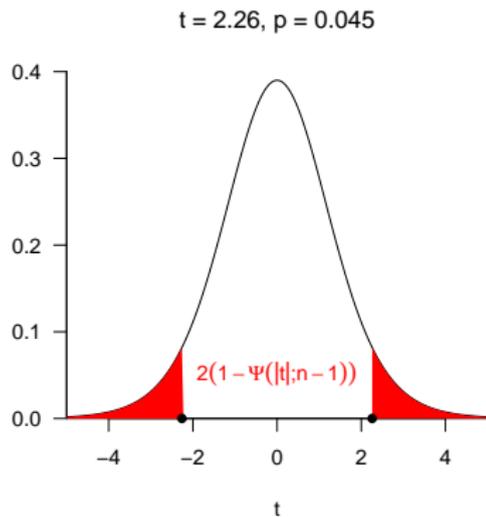
abgelehnt werden. Für diese α_0 gilt, dass

$$\alpha_0 \geq 2(1 - \Psi(|t|; n - 1)), \quad (42)$$

denn

$$\begin{aligned} & |t| \geq \Psi^{-1}\left(1 - \frac{\alpha_0}{2}; n - 1\right) \\ \Leftrightarrow & \Psi(|t|; n - 1) \geq \Psi\left(\Psi^{-1}\left(1 - \frac{\alpha_0}{2}; n - 1\right); n - 1\right) \\ \Leftrightarrow & \Psi(|t|; n - 1) \geq 1 - \frac{\alpha_0}{2} \\ \Leftrightarrow & \mathbb{P}(T \leq |t|) \geq 1 - \frac{\alpha_0}{2} \\ \Leftrightarrow & \frac{\alpha_0}{2} \geq 1 - \mathbb{P}(T \leq |t|) \\ \Leftrightarrow & \frac{\alpha_0}{2} \geq \mathbb{P}(T \geq |t|) \\ \Leftrightarrow & \alpha_0 \geq 2\mathbb{P}(T \geq |t|) \\ \Leftrightarrow & \alpha_0 \geq 2(1 - \Psi(|t|; n - 1)). \end{aligned} \quad (43)$$

Das kleinste $\alpha_0 \in [0, 1]$ mit $\alpha_0 \geq 2\mathbb{P}(T \geq |t|)$ ist dann entsprechend $\alpha_0 = 2(1 - \Psi(|t|; n - 1))$.



Definition (Powerfunktion des Einstichproben-T-Tests)

Gegeben sei der zweiseitige Einstichproben-T-Test mit einfacher Nullhypothese und zusammengesetzter Alternativhypothese. Dann ist die *Powerfunktion* des Tests gegeben durch

$$\pi : \mathbb{R} \times \mathbb{N} \rightarrow [0, 1], (d, n) \mapsto \pi(d, n) := 1 - \Psi(k_{\alpha_0}; d, n - 1) + \Psi(-k_{\alpha_0}; d, n - 1) \quad (44)$$

Bemerkungen

- Wir betrachten hier lediglich die Testgütefunktion

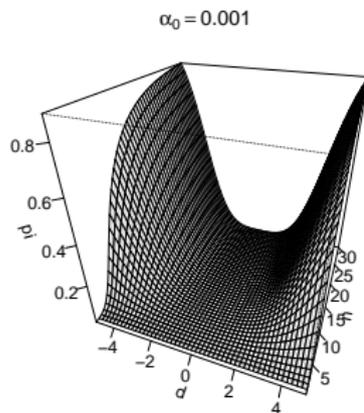
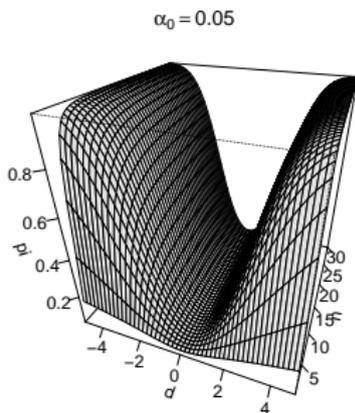
$$q_\phi : \mathbb{R} \rightarrow [0, 1], \mu \mapsto q_\phi(\mu) := 1 - \Psi(k_{\alpha_0}; d_\mu, n - 1) + \Psi(-k_{\alpha_0}; d_\mu, n - 1) \quad (45)$$

bei kontrolliertem Testumfang, also für

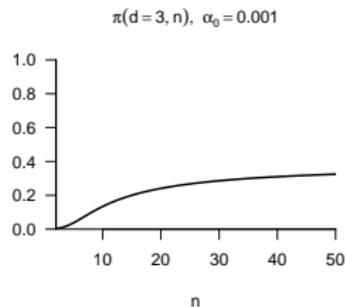
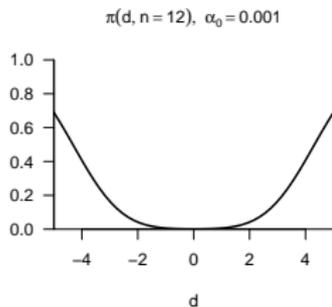
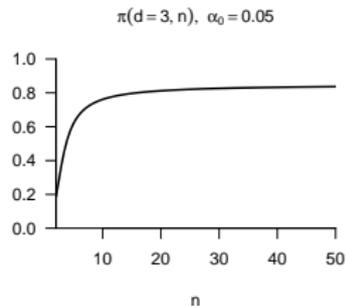
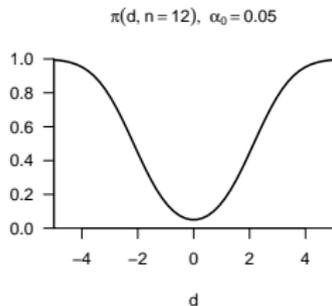
$$k_{\alpha_0} := \Psi^{-1}(1 - \alpha_0/2; n - 1) \quad (46)$$

mit festem α_0 als Funktion des Nichtzentralitätsparameters und des Stichprobenumfangs.

Powerfunktionen des Einstichproben-T-Tests



Powerfunktionsschnitte des Einstichproben-T-Tests



Man nimmt an, dass ein vorliegender univariater Datensatz y_1, \dots, y_n eine Realisierung des Frequentistischen Inferenzmodells $v_1, \dots, v_n \sim N(\mu, \sigma^2)$ des Einstichproben-T-Tests mit wahren, aber unbekanntem Parameter μ und $\sigma^2 > 0$ ist.

Man nimmt ferner an, dass man entscheiden muss ob für einen gewählten Nullhypothesenparameter μ_0 eher die Nullhypothese $H_0 : \mu = \mu_0$ oder die Alternativhypothese $H_1 : \mu \neq \mu_0$ zutrifft.

Um den Testumfang über viele Wiederholungen dieser Testprozedur zu kontrollieren, wählt ein Signifikanzlevel α_0 und bestimmt den zugehörigen kritischen Wert k_{α_0} , so dass zum Beispiel bei einem Stichprobenumfang von $n = 12$ und der Wahl von $\alpha_0 := 0.05$ ein kritischer Wert von $k_{0,05} = 2.20$ gewählt wird.

Anhand des Stichprobenumfangs n , des Nullhypothesenparameters μ_0 , des Stichprobenmittels \bar{y} und der Stichprobenstandardabweichung s berechnet man sodann den Wert der Einstichproben-T-Test-Statistik durch

$$t := \sqrt{n} \frac{\bar{y} - \mu_0}{s}. \quad (47)$$

Wenn dieses für den vorliegenden Datensatz so bestimmte t größer als k_{α_0} ist oder wenn t kleiner als $-k_{\alpha_0}$ ist, lehnt man die Nullhypothese ab, andernfalls lehnt man sie nicht ab. Die oben entwickelte Theorie garantiert dann, dass man im langfristigen Mittel in höchstens $\alpha_0 \cdot 100$ von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.

Weiterhin bestimmt man basierend auf dem vorliegenden Wert der Einstichproben-T-Test-Statistik den zugehörigen p-Wert durch

$$p\text{-Wert} = 2(1 - \Psi(|t|; n - 1)). \quad (48)$$

Folgender **R** Code demonstriert dieses Vorgehen bei Annahme eines vorliegenden Datenvektors y der Länge n .

```
n          = length(y)           # Stichprobenumfang
mu_0       = 0                   # Nullhypothesenparameter
alpha_0    = 0.05                # Signifikanzlevel
k_alpha_0  = qt(1-alpha_0/2,n-1) # kritischer Wert
Tee        = sqrt(n)*((mean(y) - mu_0)/sd(y)) # Einstichproben-T-Test-Statistik
if(abs(Tee) >= k_alpha_0){phi = 1} else {phi = 0} # Testauswertung
p = 2*(1 - pt(Tee,n-1))          # p-Wert Evaluation
```

Will man im Rahmen einer Studienplanung eine Poweranalyse zur Optimierung des Stichprobenumfangs im vorliegenden Testscenario durchführen, so gilt natürlich zunächst grundsätzlich, dass mit steigendem Stichprobenumfang die Powerfunktion des Tests ansteigt. Vor dem Gesichtspunkt der Power des Tests ist ein größerer Stichprobenumfang also immer besser als ein kleinerer Stichprobenumfang.

Allerdings bleiben dabei mögliche Kosten für die Erhöhung des Stichprobenumfangs, wie zum Beispiel mögliche Risiken für die Studienteilnehmer:innen, unberücksichtigt. Weiterhin ist der Wert, den die Powerfunktion bei einem gewählten Stichprobenumfang immer von den wahren, aber unbekanntenen, Parameterwerten μ und σ , die in den Wert des Nichtzentralitätsparameters d einfließen, abhängig. Würde man diese Werte in einem gegebenen Anwendungskontext schon sehr genau kennen, so würde man vermutlich keine Studie durchführen wollen.

Generell wird im Rahmen der Studienplanung deshalb folgendes Vorgehen favorisiert. Zunächst entscheidet man sich für ein Signifikanzlevel α_0 zur Kontrolle des Testumfangs und evaluiert die Powerfunktion. Man überlegt sich dann einen Nichtzentralitätswert d^* , den man mit einer Power von mindestens β detektieren möchte, wobei ein typischer konventioneller $\beta = 0.8$ ist. Man wertet dann die für einen Powerfunktionswert

$$\pi(d = d^*, n) = \beta \tag{49}$$

nötige Stichprobengröße aus.

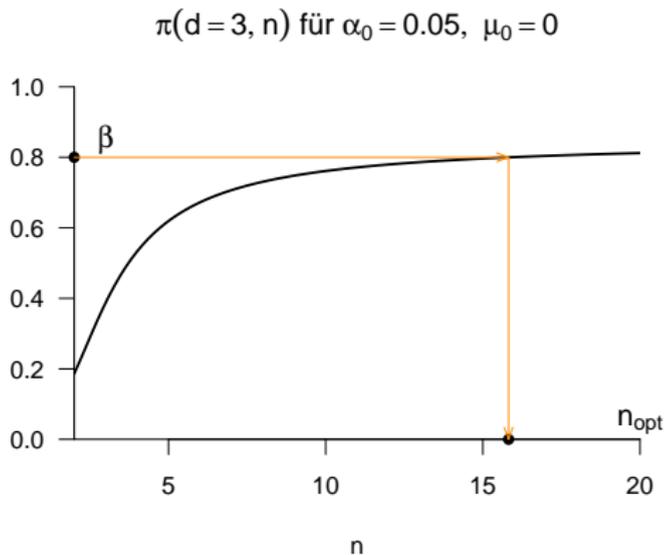
Einstichproben-T-Test | (11) Praktische Durchführung

Aufgrund der Monotonie der Powerfunktion des zweiseitigen Einstichproben-T-Tests mit einfacher Nullhypothese und zusammengesetzter Alternativhypothese im Bereich nicht-negativer Nichtzentralitätsparameter ist dann gewährleistet, dass die Power des Tests für Nichtzentralitätsparameter, die größer als d^* sind, größer oder gleich β sind. Folgender R Code implementiert dieses Vorgehen zur Optimierung des Stichprobenumfangs

```
# Powerfunktionsbasierte Stichprobenumfangsoptimierung
alpha_0 = 0.05 # Signifikanzlevel
beta = 0.8 # gewünschter Powerfunktionswert
d_stern = 3 # fester Nichtzentralitätsparameter
n_min = 2 # minimal betrachteter Stichprobenumfang
n_max = 20 # maximal betrachteter Stichprobenumfang
n_res = 1e2 # Auflösung des Stichprobenumfangsraums
n = seq(n_min, n_max, len = n_res) # Stichprobenumfangraum
k_alpha_0 = qt(1-alpha_0/2, n-1) # kritische Werte
pi_n = 1-pt(k_alpha_0, n-1, d_stern)+pt(-k_alpha_0, n-1, d_stern) # Powerfunktion
i = 1 # Indexinitialisierung
n_min = NaN # minimales n Initialisierung
while(pi_n[i] < beta){ # Solange \pi(d*,n) < \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 = 16

Powerfunktionsbasierte Stichprobenumfangsoptimierung



Testhypothesen und Tests

Testgütekriterien und Testkonstruktion

Einstichproben-T-Test

Anwendungsbeispiel

Konfidenzintervalle und Hypothesentests

Selbstkontrollfragen

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression



BDI-II Fragebogen

Name: _____ Alter: _____ Geschlecht: _____ Datum: _____

Anleitung: Dieser Fragebogen enthält 21 Gruppen von Aussagen. Bitte lesen Sie jede dieser Gruppen von Aussagen sorgfältig durch und wählen Sie sich dann in jeder Gruppe eine Aussage heraus, die am besten beschreibt, wie Sie sich in den letzten zwei Wochen **überwiegend** gefühlt haben. **Wichtig!** Geben Sie die Zahl neben der Aussage an, die für sich am besten ausreicht (haben 0, 1, 2 oder 3). Falls in einer Gruppe mehrere Aussagen gleichberechtigt auf Sie zuzutreffen, können Sie die Aussage wählen, die für Sie den besten Fall ist. Achten Sie bitte darauf, dass Sie in jeder Gruppe nicht mehr als eine Aussage auswählen, die gilt noch für Gruppen in verschiedenen der Subskalienergebnisse oder Gruppen in verschiedenen der Äquivalente.

1.) Traurigkeit 0 Ich bin nicht traurig. 1 Ich bin ein bisschen traurig. 2 Ich bin so traurig, wie gewöhnlich, dass ich mir nicht darüber Gedanken mache. 3 Ich bin sehr traurig.	6.) Besorgungsgefühle 0 Ich habe nicht das Gefühl, für etwas Besorgnis zu sein. 1 Ich habe das Gefühl, vielleicht Besorgnis zu sein. 2 Ich bin etwas besorgt zu sein. 3 Ich habe das Gefühl, Besorgnis zu sein.
2.) Zukunftsdenken 0 Ich sehe nicht mehr in die Zukunft, ich sehe nur auf das Jetzt. 1 Ich bin etwas optimistisch, aber ich sehe nicht mehr in die Zukunft. 2 Ich bin optimistisch und glaube, dass meine Situation besser wird. 3 Ich bin optimistisch, dass meine Zukunft hoffnungsvoll ist und nur noch schlechter wird.	7.) Selbstabwertung 0 Ich habe mir nie gemacht und nie etwas. 1 Ich habe mir manchmal ein bisschen angetan. 2 Ich bin ein bisschen selbstverleugend. 3 Ich bin mir selbstverleugend.
3.) Verantwortungslosigkeit 0 Ich fühle mich nicht als Versager. 1 Ich habe manchmal Verantwortungslosigkeit. 2 Wenn ich Verantwortung habe, fühle ich eine Menge Verantwortung. 3 Ich habe das Gefühl, als Mensch ein völliger Versager zu sein.	8.) Selbstverleugung 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 Fehler. 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.) Selbstverleugung 0 Ich bin stolz auf das, was ich erreicht habe. 1 Ich bin stolz auf das, was ich erreicht habe, aber ich bin nicht so stolz wie früher. 2 Ich bin stolz auf das, was ich erreicht habe, aber ich bin nicht so stolz wie früher. 3 Ich bin stolz auf das, was ich erreicht habe, aber ich bin nicht so stolz wie früher.
5.) Schuldgefühle 0 Ich habe keine besonderen Schuldgefühle. 1 Ich habe ein bisschen Schuldgefühle, die ich gegen habe oder haben kann. 2 Ich habe ein bisschen Schuldgefühle, die ich gegen habe oder haben kann. 3 Ich habe ein bisschen Schuldgefühle, die ich gegen habe oder haben kann.	10.) Wut 0 Ich wütet nicht mehr als früher. 1 Ich wütet ein bisschen mehr als früher. 2 Ich wütet ein bisschen mehr als früher. 3 Ich wütet ein bisschen mehr als früher.

PROBEN
© 2001
© 2001
© 2001

⇒ Pre-Post BDI Score Reduktion

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

—
dBDI
—
-1
3
-2
9
3
-2
4
5
5
1
9
4
—

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Für die Pre-Post BDI Score Reduktion v_i der i ten von n Patient:innen legen wir das Modell

$$v_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, \dots, n \quad (50)$$

zugrunde. Dabei wird die Pre-Post BDI Reduktion v_i der i ten Patient:in also mithilfe einer über die Gruppe von Patient:innen identischen Pre-Post BDI Score Reduktion $\mu \in \mathbb{R}$ und einer Patient:innen-spezifischen normalverteilten Pre-Post BDI Score Reduktionsabweichung ε_i erklärt

Wie gezeigt ist dieses Modell äquivalent zum Normalverteilungsmodell

$$v_1, \dots, v_n \sim N(\mu, \sigma^2). \quad (51)$$

Die Standardproblemstellungen der Frequentistischen Inferenz führen dann auf folgende Fragen:

- (1) Was sind sinnvolle Tipps für die wahren, aber unbekanntenen, Parameterwerte μ und σ^2 ?
- (2) Wie kann im Sinne einer Intervallschätzung eine möglichst sichere Schätzung von μ gelingen?
- (3) Entscheiden wir uns sinnvollerweise für die Hypothese, dass gilt $\mu \neq 0$?

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

```
D      = read.csv("./11_Daten/11_Hypothesentests.csv") # Datensatzeinlesen
y      = D$dBDI                                     # Datenauswahl
n      = length(y)                                 # Stichprobenumfang
mu_hat = mean(y)                                   # Erwartungswertparameterschätzer
delta  = 0.95                                     # Konfidenzlevel
t_delta = qt((1+delta)/2,n-1)                     # \Psi^{-1}((\delta + 1)/2, n-1)
G_u    = mean(y) - (sd(y)/sqrt(n))*t_delta        # untere Konfidenzintervallgrenze
G_o    = mean(y) + (sd(y)/sqrt(n))*t_delta        # obere Konfidenzintervallgrenze
mu_0   = 0                                         # Nullhypothesenparameter, hier \mu = \mu_0
alpha_0 = 0.05                                    # Signifikanzlevel
k_alpha_0 = qt(1-alpha_0/2,n-1)                  # kritischer Wert
Tee    = sqrt(n)*((mean(y) - mu_0)/sd(y))        # T-Teststatistik
if(abs(Tee) >= k_alpha_0){phi = 1} else {phi = 0} # Test 1_{|t| \ge k_alpha_0}
p      = 2*(1 - pt(Tee,n-1))                     # p-Wert
cat("Parameterschätzwert   =", mu_hat,
    "\n95%-Konfidenzintervall =", G_u, G_o,
    "\nSignifikanzlevel     =", alpha_0,
    "\nKritischer Wert      =", k_alpha_0,
    "\nTeststatistik        =", Tee,
    "\nTestwert             =", phi,
    "\np-Wert                =", p)
```

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Parameterschätzwert	= 3.166667
95%-Konfidenzintervall	= 0.8074098 5.525923
Signifikanzlevel	= 0.05
Kritischer Wert	= 2.200985
Teststatistik	= 2.95423
Testwert	= 1
p-Wert	= 0.01310986

Im vorliegenden Fall würde man die Nullhypothese bei einem Signifikanzlevel von $\alpha_0 = 0.05$ ablehnen. Ob die Nullhypothese zutrifft oder nicht bleibt wie der wahre, aber unbekannte, Wert von μ unbekannt. Im langfristigen Mittel würde man, basierend auf den zugrundegelegten Annahmen, die Nullhypothese nur in 5 von 100 Fällen fälschlicherweise ablehnen.

Anwendungsbeispiel

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Durchführung mit der R Funktion `t.test()`

```
D      = read.csv("./11_Daten/11_Hypothesentests.csv") # Datensatzeinlesen
y      = D$dBDI                                     # Datenauswahl
t.test(y)                                           # Einstichproben-T-Test
```

One Sample t-test

```
data: y
t = 2.9542, df = 11, p-value = 0.01311
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.8074098 5.5259235
sample estimates:
mean of x
 3.166667
```

Im vorliegenden Fall würde man die Nullhypothese bei einem Signifikanzlevel von $\alpha_0 = 0.05$ ablehnen. Ob die Nullhypothese zutrifft oder nicht bleibt wie der wahre, aber unbekannt, Wert von μ unbekannt. Im langfristigen Mittel würde man, basierend auf den zugrundegelegten Annahmen, die Nullhypothese nur in 5 von 100 Fällen fälschlicherweise ablehnen.

Testhypothesen und Tests

Testgütekriterien und Testkonstruktion

Einstichproben-T-Test

Anwendungsbeispiel

Konfidenzintervalle und Hypothesentests

Selbstkontrollfragen

Theorem (Dualität von Konfidenzintervallen und Hypothesentests)

Es sei v die Stichprobe eines Frequentistischen Inferenzmodells mit Ergebnisraum \mathcal{Y} und Parameterraum Θ . Weiterhin sei für ein $\delta \in]0, 1[$ mit $[G_u(v), G_o(v)]$ ein δ -Konfidenzintervall für den wahren, aber unbekanntem, Parameter $\theta \in \Theta$ definiert. Dann gilt, dass der Hypothesentest definiert durch

$$\phi_\theta : \mathcal{Y} \rightarrow \{0, 1\}, y \mapsto \phi(y) := \begin{cases} 0, & [G_u(y), G_o(y)] \ni \theta_0 \\ 1, & [G_u(y), G_o(y)] \not\ni \theta_0 \end{cases} \quad (52)$$

ein Hypothesentest vom Signifikanzlevel $\alpha_0 = 1 - \delta$ für die Hypothesen

$$\Theta_0 := \{\theta_0\} \text{ und } \Theta_1 := \Theta \setminus \{\theta_0\}. \quad (53)$$

Beweis

Aufgrund der einfachen Nullhypothese und somit $\alpha_0 = \alpha$ folgt

$$\alpha_0 = \alpha = \mathbb{P}_{\theta_0}(\phi(v) = 1) = \mathbb{P}_{\theta_0}([G_u(v), G_o(v)] \not\ni \theta_0) = 1 - \mathbb{P}_{\theta_0}([G_u(v), G_o(v)] \ni \theta_0) = 1 - \delta. \quad (54)$$

□

Bemerkung

- Mit δ -Konfidenzintervallen kann man also Hypothesentests mit Signifikanzlevel $\alpha_0 = 1 - \delta$ konstruieren.

Theorem (Dualität von Konfidenzintervall und Einstichproben-T-Test)

Gegeben sei das Normalverteilungsmodell und es sei

$$\kappa := \left[\bar{v} - \frac{S}{\sqrt{n}} t_\delta, \bar{v} + \frac{S}{\sqrt{n}} t_\delta \right]. \quad (55)$$

das mithilfe von

$$t_\delta := \Psi^{-1} \left(\frac{1+\delta}{2}; n-1 \right) \quad (56)$$

in (9) Konfidenzintervalle definierte δ -Konfidenzintervall für den Erwartungswertparameter. Dann ist der Test

$$\phi : \mathcal{Y} \rightarrow \{0, 1\}, y \mapsto \phi(y) := \begin{cases} 0, & \left[\bar{v} - \frac{S}{\sqrt{n}} t_\delta, \bar{v} + \frac{S}{\sqrt{n}} t_\delta \right] \ni \mu_0 \\ 1, & \left[\bar{v} - \frac{S}{\sqrt{n}} t_\delta, \bar{v} + \frac{S}{\sqrt{n}} t_\delta \right] \not\ni \mu_0 \end{cases} \quad (57)$$

ein Test der einfachen Nullhypothese $H_0 : \mu = \mu_0$ und der zusammengesetzten Alternativhypothese $H_1 := \mu \neq \mu_0$ mit Signifikanzlevel $\alpha_0 = 1 - \delta$.

Beweis

Es gilt

$$\mathbb{P}_{\mu_0}(\phi(v) = 1) = 1 - \mathbb{P}_{\mu_0}(\phi(v) = 0) = 1 - \mathbb{P}_{\mu_0} \left(\left[\bar{v} - \frac{S}{\sqrt{n}} t_\delta, \bar{v} + \frac{S}{\sqrt{n}} t_\delta \right] \ni \mu_0 \right) = 1 - \delta. \quad (58)$$

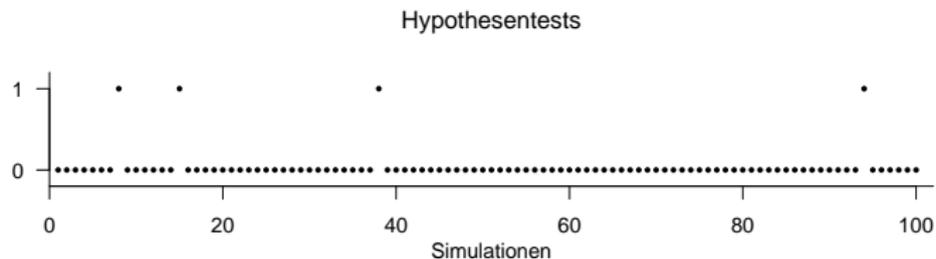
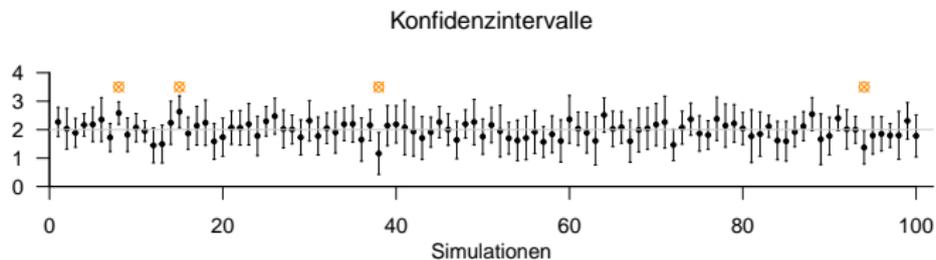
Simulation der Dualität von Konfidenzintervall und Einstichproben-T-Test

```
n      = 12                                # Stichprobenumfang
mu     = 2                                 # wahrer, aber unbekannter, Erwartungswertparameter
sigsqr = 1                                 # wahrer, aber unbekannter, Varianzparameter
delta  = 0.95                              # Konfidenzlevel
t_delta = qt((1+delta)/2, n-1)            # \Psi^{-1}((\delta + 1)/2, n-1)
mu_0   = mu                                # Nullhypothesenparameter bei Zutreffen von H_0
set.seed(1)                                # random number generator seed
ns     = 1e2                               # Anzahl Simulationen
y_bar  = rep(NaN, ns)                      # Stichprobenmittelarray
s      = rep(NaN, ns)                      # Stichprobenstandardabweichungarray
kappa  = matrix(rep(NaN, 2*ns), ncol = 2)  # Konfidenzintervallarray
kfn    = rep(NaN, ns)                     # Überdeckungsindikatorarray
phi    = rep(NaN, ns)                     # Testarray
for(i in 1:ns){                             # Simulationsiterationen
  y      = rnorm(n, mu_0, sqrt(sigsqr))     # Stichprobenrealisierung
  y_bar[i] = mean(y)                       # Stichprobenmittel
  s[i]   = sd(y)                           # Stichprobenstandardabweichung
  kappa[i,1] = y_bar[i] - (s[i]/sqrt(n))*t_delta # untere Konfidenzintervallgrenze
  kappa[i,2] = y_bar[i] + (s[i]/sqrt(n))*t_delta # obere Konfidenzintervallgrenze
  if(kappa[i,1] <= mu_0 & mu_0 <= kappa[i,2]){
    kfn[i] = 1} else{kfn[i] = 0}           # Überdeckungsindikatorevaluation
  if(kappa[i,1] <= mu_0 & mu_0 <= kappa[i,2]){
    phi[i] = 0} else{phi[i] = 1}         # Testevaluation
}
cat( "Geschätztes Konfidenzniveau =", mean(kfn),
     "\nGeschätzter Testumfang      =", mean(phi)) # Ausgabe
```

```
Geschätztes Konfidenzniveau = 0.96
Geschätzter Testumfang      = 0.04
```

Konfidenzintervalle und Hypothesentests

Simulation der Dualität von Konfidenzintervall und Einstichproben-T-Test



Testhypothesen und Tests

Testgütekriterien und Testkonstruktion

Einstichproben-T-Test

Anwendungsbeispiel

Konfidenzintervalle und Hypothesentests

Selbstkontrollfragen

Selbstkontrollfragen

1. Erläutern Sie die grundlegende Logik Frequentistischer Hypothesentests.
2. Geben Sie die Definition der Begriffe der Testhypothesen und des Testszenario wieder.
3. Geben Sie die Definition der Begriffe der einfachen und zusammengesetzten Testhypothesen wieder.
4. Geben Sie die Definition der Begriffe einseitigen und zweiseitigen Testhypothesen wieder.
5. Geben Sie die Definition des Begriff des Tests wieder.
6. Geben Sie die Definition des Begriffs des Standardtests wieder.
7. Geben Sie die Definition des Begriffs des kritischen Bereichs wieder.
8. Geben Sie die Definition des Begriffs des Ablehnungsbereichs wieder
9. Geben Sie die Definition des Begriffs des kritischen Wert-basierten Tests wieder.
10. Geben Sie die Definition der Begriffe der richtigen Testentscheidungen und der Testfehler wieder.
11. Geben Sie die Definition des Begriffs der Testgütefunktion wieder.
12. Erläutern Sie die Bedeutung der Testgütefunktion im Rahmen der Konstruktion von Hypothesentests.
13. Geben Sie die Definition der Begriffe des Level- α_0 -Tests und des Signifikanzlevels α_0 wieder
14. Geben Sie die Definition des Begriffs des Testumfangs α wieder.
15. Geben Sie die Definition des Begriffs des p-Werts wieder.

Selbstkontrollfragen - Lösungen

1. Siehe [Grundlegende Logik Frequentistischer Hypothesentests](#) auf Folie 6
2. Siehe Definition (Testhypothesen und Testszenario).
3. Siehe Definition (Einfache und zusammengesetzte Testhypothesen).
4. Siehe Definition (Einseitige und zweiseitige Testhypothesen).
5. Siehe Definition (Test).
6. Siehe Definition (Standardtest).
7. Siehe Definition (Kritischer Bereich).
8. Siehe Definition (Ablehnungsbereich).
9. Siehe Definition (Kritischer Wert-basierte Tests).
10. Siehe Definition (Richtigen Testentscheidungen und Testfehler).
11. Siehe Definition (Testgütefunktion).
12. Siehe [Intuition zur Testkonstruktion](#) auf Folien 20 und 21.
13. Siehe Definition (Level- α_0 -Test, Signifikanzlevel α_0 , Testumfang α).
14. Siehe Definition (Level- α_0 -Test, Signifikanzlevel α_0 , Testumfang α).
15. Siehe Definition (p-Wert).

- Lehmann, E. L. 1986. *Testing Statistical Hypotheses*. Wiley Series in Probability and Statistics.
- 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.
- Wasserstein, Ronald L., Allen L. Schirm, and Nicole A. Lazar. 2019. "Moving to a World Beyond ' $p < 0.05$!'" *The American Statistician* 73 (sup1): 1–19. <https://doi.org/10.1080/00031305.2019.1583913>.