Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2022/23

Prof. Dr. Dirk Ostwald

Aufnahme läuft!

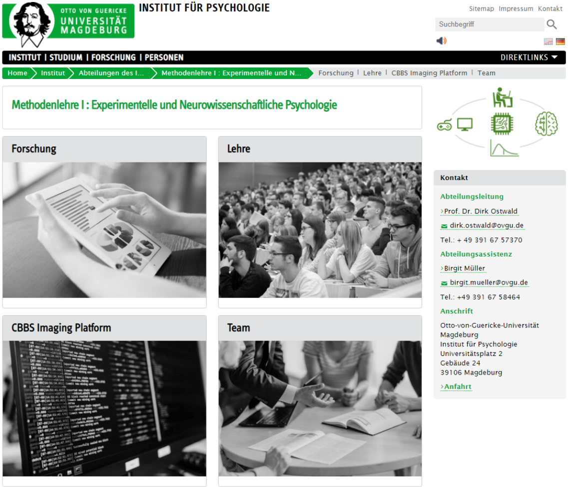
(1) Einführung

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

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

Forschung	Komputationale Kognitive Neurowissenschaften
Lehre	Datenwissenschaft

Homepage


OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG
INSTITUT FÜR PSYCHOLOGIE
Sitemap Impressum Kontakt
Suchbegriff
INSTITUT | STUDIUM | FORSCHUNG | PERSONEN
DIREKTLINKS
Home Institut Abteilungen des I... Methodenlehre I : Experimentelle und N...
Forschung | Lehre | CBBS Imaging Platform | Team
Methodenlehre I : Experimentelle und Neurowissenschaftliche Psychologie
Forschung
Lehre
CBBS Imaging Platform
Team
Kontakt
Abteilungsleitung
Prof. Dr. Dirk Ostwald
dirk.ostwald@ovgu.de
Tel.: + 49 391 67 57370
Abteilungsassistenz
Birgit Müller
birgit.mueller@ovgu.de
Tel.: +49 391 67 58464
Anschrift
Otto-von-Guericke-Universität Magdeburg
Institut für Psychologie
Universitätsplatz 2
Gebäude 24
39106 Magdeburg
Anfahrt

Datenwissenschaft und Statistik

Formalia

Studium und Diskussion

Selbstkontrollfragen

Datenwissenschaft und Statistik

Formalia

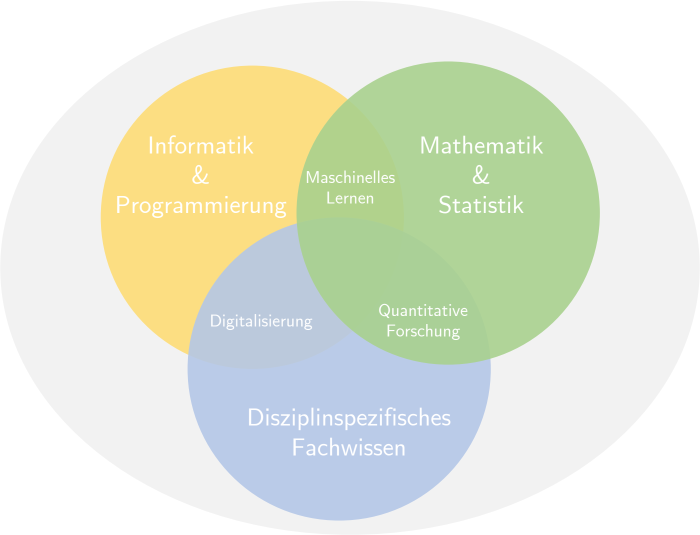
Studium und Diskussion

Selbstkontrollfragen

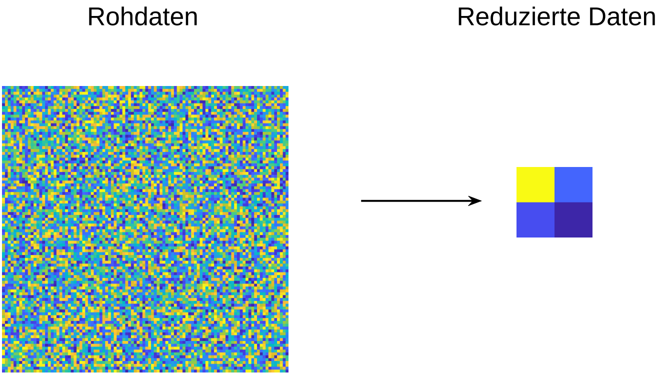
Datenwissenschaft

Die Kunst, aus Daten Sinn zu generieren

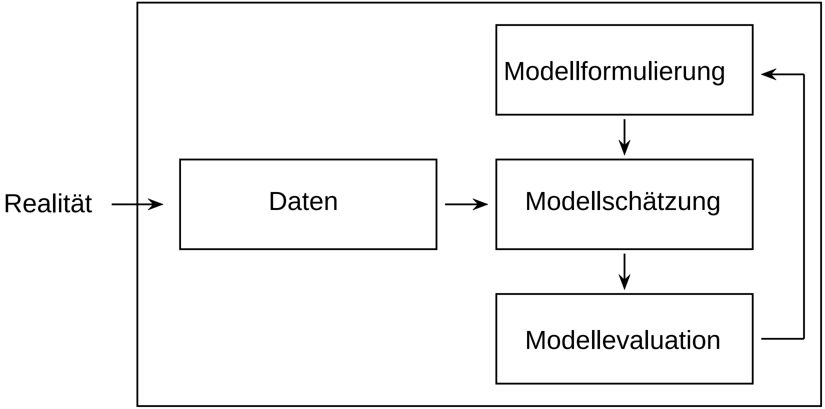
Datenwissenschaft und Statistik



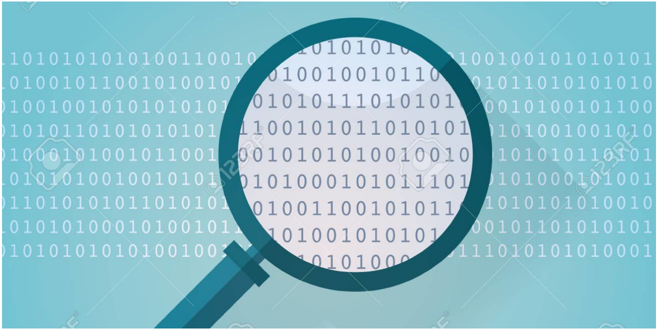
Datenwissenschaft ist Datenreduktion



Datenwissenschaft ist Naturwissenschaft



Datenwissenchaft ist Dateninterpretation



Terminologie der Datenwissenschaft

Statistik = Maschinelles Lernen = Künstliche Intelligenz

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

Datenwissenschaft in der Psychologie

Die Kunst, aus Verhaltens- und Neurophysiologiedaten

psychologischen Sinn zu generieren

Statistik

Die Kunst, aus Daten Sinn zu generieren

und seine assoziierte Unsicherheit zu quantifizieren

Asymptotische Inferenz

Maßtheorie

Deskriptive Statistik

Wahrscheinlichkeitstheorie

Schätztheorie

Konfidenzintervalle

Korrelation und Regression

Frequentistische Inferenz

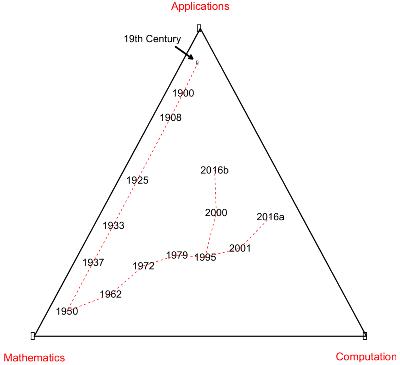
Bayesianische Inferenz

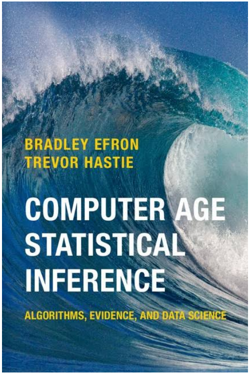
Hypothesentests

Varianzanalyse

Poweranalyse

Datenwissenschaft und Statistik





Statistik in der Psychologie

Die Kunst, aus Verhaltens- und Neurophysiologiedaten

psychologischen Sinn zu generieren

und seine assoziierte Unsicherheit zu quantifizieren

Datenwissenschaft und Statistik

Klassische Partition der Statistik in der Psychologie

- Deskriptivstatistik
- Inferenzstatistik
- Multivariate Statistik

Sinnvolle Partition der psychologischen Datenwissenschaft

- Wahrscheinlichkeitstheoretische Grundlagen
- Frequentistische Inferenz
- Allgemeines Lineares Modell
- Bayesianische Inferenz
- Multivariate Datenanalyse

Fundamentale Annahmen der Wahrscheinlichkeitstheorie

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

© 2023 Dirk Ostwald CC BY-NC-SA 4.0

Fundamentale Annahmen der Frequentistischen Inferenz

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

Fundamentale Annahmen der Bayesianischen Inferenz

- Wahrscheinlichkeiten werden als Grade der Sicherheit, nicht als langfristige relative Häufigkeiten interpretiert. Aussagen der Form "Die Wahrscheinlichkeit, dass das Wintersemester 2022/23 vollständig in Präsenzlehre stattfindet, ist 0.9." haben eine Bedeutung.

- Die Parameter probabilistischer Modelle sind feste, unbekannte Konstanten, die als *wahre, aber unbekannte, Parameterwerte* bezeichnet werden. Über Parameterwerte und Modelle werden probabilistische Aussagen getroffen, die unseren Grad an Sicherheit hinsichtlich ihrer quantitativen Ausprägung und Validität widerspiegeln.

- Probabilistische Aussagen über Parameter werden mithilfe von Wahrscheinlichkeitsverteilungen getroffen, auf deren Grundlage optimale Entscheidungen im Sinne von Kosten-Nutzenfunktionen getroffen werden können.

Datenwissenschaft und Statistik

Datenwissenschaftliches Curriculum der OVGU Psychologie

- Mathematische Grundlagen
  - Mengen, Funktionen, Differentialrechnung, Integralrechnung, ...
- Wahrscheinlichkeitstheorie
  - Maßtheoretische Grundlagen, Zufallsvariablen, Erwarungswert, Varianz, ...
- Frequentistische Inferenz
  - Statistische Modelle, Schätztheorie, Konfidenzintervalle, Hypothesentesten,...
- Allgemeines Lineares Modell
  - Matrizen, multivariate Normalverteilung, Schätztheorie, Studiendesigns, ...
- Bayesianische Inferenhz
  - Konjugierte Modelle, numerische Inferenz, variational inference, ...
- Multivariate Methoden
  - Multivariates ALM, Faktoranalyse, Neuronale Netze, ...
- Programmierung
  - Grundlagen von R, Python, Matlab, Linux, Parallel computing, ...

Formalia

Modul B1 Deskriptive Statistik | Wahrscheinlichkeitstheorie und Frequentistische Inferenz

- Donnerstags 13-16 Uhr in Raum G40B-231
- Kursmaterialien (Folien, Videos, RMarkdown Code) auf der Kurswebseite
- Code auf Github
- Ankündigungen über die Moodleseite
- Empfohlene Lektüre ist PDWP
- Link zu vorheriger Iteration des Kurses
- Link zum Kurs Mathematische Grundlagen
- Tutorium Mittwochs 11-13 Uhr mit Belinda Fleischmann

Modul B1 Deskriptive Statistik | Wahrscheinlichkeitstheorie und Frequentistische Inferenz

Datum	Einheit	Thema
13.10.2022	Einführung	(1) Einführung
20.10.2022	Wahrscheinlichkeitstheorie	(2) Wahrscheinlichkeitsräume
27.10.2022	Wahrscheinlichkeitstheorie	(3) Zufallsvariablen
03.11.2022	Wahrscheinlichkeitstheorie	(4) Zufallsvektoren
10.11.2022	Wahrscheinlichkeitstheorie	(5) Erwartungswert und Kovarianz
17.11.2022	Wahrscheinlichkeitstheorie	(6) Ungleichungen und Grenzwerte
24.11.2022	Wahrscheinlichkeitstheorie	(7) Normalverteilungstransformationen
01.12.2022	Frequentistische Inferenz	(8) Statistische Modelle, Statistiken, Schätzer
08.12.2022	Frequentistische Inferenz	(9) Schätzereigenschaften
15.12.2022	Frequentistische Inferenz	(10) Konfidenzintervalle
	Weihnachtspause	
05.01.2023	Frequentistische Inferenz	(11) Hypothesentests
12.01.2023	Frequentistische Inferenz	(12) T-Tests
19.01.2023	Frequentistische Inferenz	(13) Einfaktorielle Varianzanalyse
26.01.2023	Frequentistische Inferenz	(14) Zweifaktorielle Varianzanalyse
Feb 2023	Klausurtermin	
Jul 2023	Klausurwiederholungstermin	

Formalia

Modul B1 Deskriptive Statistik | Wahrscheinlichkeitstheorie und Frequentistische Inferenz

- Vorlesungsfolien inklusive Selbstkontrollfragen sind klausurrelevant
- Altklausuren finden sich auf den Kurswebseiten früherer Jahre
- Benotete Multiple Choice Klausur (30 Fragen) Ende Wintersemester 2022/23
- Klausurwiederholungstermin am Ende des Sommersemesters 2023
- Klausurtermin und Klausurort gemäß Prüfungsplan des FNW Prüfungsamtes

Datenwissenschaft und Statistik

Formalia

Studium und Diskussion

Selbstkontrollfragen

Umfrage zum Studienstart

Studium $\neq$ Schule

- Schule ist Pflicht, Studium ist freiwillig.
- Sie wollen nicht studiert werden, Sie wollen studieren.
- Sie sind motiviert.
- Studium ist Arbeit mit 40-Stundenwoche.
- Wir machen keinen Osterhasenunterricht.
- Klausuren dienen Ihnen, nicht den Lehrenden.
- Veranstaltungen dienen der Organisation, nicht des Erwerbs von Wissen.

Studium $\neq$ Berufsausbildung

- Das Studium dient dem Erwerb theoretischen Wissens.
- Studium = Reproduktion, Praxis = Translation, Wissenschaft = Reflexion.
- Sie werden nie wieder so viel Zeit zum Erwerb theoretischen Wissens haben.
- Nach Studienabschluss sind Sie keine Psychotherapeut:in.
- Nach Studienabschluss haben sie viel über Psychologie gelesen.
- Praktische Fähigkeiten lernt man in der Praxis, nicht in der Theorie.
- Denken und lernen Sie interdisziplinär, Fachgrenzen sind für Faule.

Lernphasen

Phase 1: Überblicken

- Überblick durch Vorlesung/Überfliegen der Materialien.
- Verstehen einfacher Zusammenhänge.
- Verstehen, was man nicht versteht.

Phase 2: Verstehen

- Erarbeiten des Verstehens komplexer Zusammenhänge.
- Schriftliche Beantwortung der Selbstkontrollfragen.
- Klärung von Details.

Phase 3: Memorisieren

- Auswendiglernen aller Inhalte.
- Aktive Wiedergabe der Inhalte, schriftlich oder mündlich.
- Teilnahme an der Klausur.

Teilen Sie große Aufgaben immer in viele kleine, gut zu bewältigende Aufgaben!

Sie machen Schreibtischarbeit, treiben Sie also täglich Sport!

Verschiedenes

Ist Statistik schwer?

Ich kann kein Mathe, Statistik macht mich fertig. Was soll ich bloß tun?

Psychotherapeut:in wollte ich eigentlich jetzt erstmal nicht werden, sondern ich will Menschen verstehen. Wozu brauche ich da Statistik?

Ich würde gerne verstehen, wie das Gehirn funktioniert. In welchem Kurs bekomme ich die Antwort?

Warum muss ich etwas über wissenschaftliche Methoden lernen, ich will doch viel lieber Menschen helfen?

Approbationsordnung für Psychotherapeutinnen und Psychotherapeuten (2020)

Inhalte, die im Bachelorstudiengang im Rahmen der hochschulischen Lehre zu vermitteln und bei dem Antrag auf Zulassung zur psychotherapeutischen Prüfung nachzuweisen sind.

9. wissenschaftliche Methodenlehre

Die studierenden Personen (. . .)

c) wenden Begriffe, Methoden und Ergebnisse der qualitativen und quantitativen Forschung in der psychologischen Grundlagen- und Anwendungsforschung an,

d) beurteilen die Auswirkungen von Forschungsmethoden auf Untersuchungspopulationen und wenden deskriptive und inferenzstatistische Methoden sowie weitere statistische Verfahren zur Auswertung von Ergebnissen grundlagen- und anwendungsbezogener Studien in verschiedenen Bereichen der psychologischen und psychotherapeutischen Forschung an,

e) planen wissenschaftliche Untersuchungen, führen diese Untersuchungen durch und werten sie aus, (. . .)

$\Rightarrow$ Bachelorarbeit

Zur Vermitllung der Inhalte der wissenschaftlichen Methodenlehre sind bei der Planung der hochschulischen Lehre (. . .) die folgenden Wissensbereiche abzudecken (. . .)

c) deskriptive und Inferenz-Statistik (. . .)

d) Datenerhebung und Datenanalyse unter Nutzung digitaler Technologien.

Studium und Diskussion

Approbationsordnung für Psychotherapeutinnen und Psychotherapeuten (2020)

Inhalte, die im Masterstudiengang im Rahmen der hochschulischen Lehre zu vermitteln und bei dem Antrag auf Zulassung zur psychotherapeutischen Prüfung nachzuweisen sind.

2. vertiefte Forschungsmethodik

Die studierenden Personen

a) wenden komplexe und multivariate Erhebungs- und Auswertungsmethoden zur Evaluierung und Qualitätssicherung von Interventionen an,

b) nutzen und beurteilen einschlägige Forschungsstudien und deren Ergebnisse für die Psychotherapie

c) planen selbstständig Studien zur Neu- oder Weiterentwicklung der Psychotherapieforschung oder der Forschung in angrenzenden Bereichen, führen solche Studien durch, werten sie aus und fassen sie zusammen, (...)

$\Rightarrow$ Masterarbeit

Zur Vermitllung der Inhalte der vertieften Forschnungsmethodik sind bei der Planung der hochschulischen Lehre (...) die folgenden Wissensbereiche abzudecken (...)

a) multivariate Verfahren und Messtheorie

Q & A

Datenwissenschaft und Statistik

Formalia

Studium und Diskussion

Selbstkontrollfragen

Selbstkontrollfragen

1. Nennen Sie die drei Hauptkomponenten der Datenwissenschaft.
2. Nennen Sie drei Grundannahmen der Wahrscheinlichkeitstheorie.
3. Nennen Sie drei Grundannahmen der Frequentistischen Inferenz.
4. Nennen Sie drei Grundannahmen der Bayesianischen Inferenz.

Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2021/22

Prof. Dr. Dirk Ostwald

(2) Wahrscheinlichkeitsräume

Statistik

Die Kunst, aus Daten Sinn zu generieren

und seine assoziierte Unsicherheit zu quantifizieren

Repräsentation zentraler Eigenschaften

Modellierung

Modell

Wahrscheinlichkeitstheorie

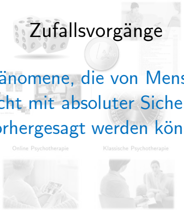
Probabilistisches Modell

$(\Omega, \mathcal{A}, \mathbb{P}), \xi : \Omega \to \mathbb{R}$

Wahrscheinlichkeitsrechnung

$\mathbb{P}_\xi(S) = \mathbb{P}(\xi^{-1}(S))$

Realität



Zufallsvorgänge

Phänomene, die von Menschen nicht mit absoluter Sicherheit vorhergesagt werden können.

Vorhersagen

Quantifizierung von Unsicherheit

Jedes Augenzahlpaar kommt im Mittel gleich häufig vor.

Basierend auf der Physik sollte jedes Augenzahlpaar die gleiche Wahrscheinlichkeit haben.

Modellierung

Modell

Wahrscheinlichkeitstheorie

$$\Omega := \{(r,b) | r \in \mathbb{N}_6, b \in \mathbb{N}_6\}$$

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

$$\mathbb{P}(\{\omega\}) := \frac{1}{36}$$

Realität

Zufallsvorgang

Gleichzeitiges Werfen eines roten und eines blauen Würfels

Vorhersagen

Die Wahrscheinlichkeit dafür, dass die Summe der gefallenen Augenzahlen 4 ist, ist $0.08\bar{3}$.

Bei 100 Würfelwürfen ist die Summe der gefallenen Augenzahlen im Durchschnitt 8.3 Mal gleich 4.

Definition

Erste Eigenschaften

Wahrscheinlichkeitsfunktionen

Beispiele

Selbstkontrollfragen

Definition

Erste Eigenschaften

Wahrscheinlichkeitsfunktionen

Beispiele

Selbstkontrollfragen

Definition

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 $\mathbb{P} : \mathcal{A} \to [0, 1]$ mit den Eigenschaften
    - o *Nicht-Negativität* $\mathbb{P}(A) \geq 0$ für alle $A \in \mathcal{A}$,
    - o *Normiertheit* $\mathbb{P}(\Omega) = 1$ und
    - o *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 $(\Omega, \mathcal{A})$ aus Ergebnismenge und Ereignissystem wird als *Messraum* bezeichnet.

Bemerkung

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

Definition

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 Komplementärmengen 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 $A_1, A_2, ... \in \mathcal{A}$ folgt, dass auch $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$ 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 DER SCHRIFTLEITUNG
DES
„ZENTRALBLATT FÜR MATHEMATIK"
ZWEITER BAND
3

GRUNDBEGRIFFE DER
WAHRSCHEINLICHKEITS-
RECHNUNG

VON

A. KOLMOGOROFF

BERLIN
VERLAG VON JULIUS SPRINGER
1933

205] § 3. Terminologische Vorbemerkungen. 5

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

Bemerkung II. Dem unmöglichen Ereignis (der leeren Menge) entspricht kraft unserer Axiome die Wahrscheinlichkeit $\mathsf{P}(0) = 0$ *, während umgekehrt aus $\mathsf{P}(A) = 0$ die Unmöglichkeit des Ereignisses A durchaus nicht zu folgen braucht; nach dem Prinzip B folgt aus dem Nullwerden der Wahrscheinlichkeit nur, daß bei einer einmaligen Realisation der Bedingungen $\mathfrak{S}$ 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 $\mathsf{P}(A) = 0$ und sehr großem n der Quotient m/n sehr klein wird (er kann z. B. gleich $1/n$ sein).

§ 3. Terminologische Vorbemerkungen.

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

Mengentheoretisch.	*Im Falle der zufälligen Ereignisse.*
1. A und B sind disjunkt, d. h. $AB = 0$.	1. Die Ereignisse A und B sind unvereinbar.
2. $AB \ldots N = 0$.	2. Die Ereignisse $A, B, \ldots, N$ sind unvereinbar.
3. $AB \ldots N = X$.	3. Das Ereignis X besteht in der gleichzeitigen Realisation aller Ereignisse $A, B, \ldots, N$.
4. $A \dotplus B \dotplus \cdots \dotplus N = X$.	4. Das Ereignis X besteht in der Erscheinung mindestens eines unter den Ereignissen $A, B, \ldots, N$.
5. Die Komplementärmenge $\bar{A}$.	5. Das entgegengesetzte Ereignis $\bar{A}$ besteht in der Nichterscheinung des Ereignisses A.
6. $A = 0$.	6. A ist unmöglich.
7. $A = E$.	7. A muß notwendig vorkommen.

* Vgl. § 3, Formel (3).

6 I. Die elementare Wahrscheinlichkeitsrechnung. [206

8. Ein System $\mathfrak{A}$ der Mengen $A_1, A_2, \ldots, A_n$ bildet eine *Zerlegung* der Menge E, wenn $A_1 + A_2 + \cdots + A_n = E$ ist (das setzt bereits voraus, daß die Mengen A_i paarweise disjunkt sind).	8. Ein *Versuch* $\mathfrak{A}$ besteht darin, daß man feststellt, welches unter den Ereignissen $A_1, A_2, \ldots, A_n$ vorkommt. $A_1, A_2, \ldots, A_n$ sind die möglichen Ausgänge des Versuches $\mathfrak{A}$.
9. B ist eine Untermenge von A: $B \subset A$.	9. Aus der Realisation des Ereignisses B folgt notwendig dieselbe von A.

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

Aus $A + \bar{A} = E$ und den Axiomen IV und V folgt

$$\mathsf{P}(A) + \mathsf{P}(\bar{A}) = 1, \tag{1}$$

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

Da $\bar{E} = 0$ ist, erhält man insbesondere

$$\mathsf{P}(0) = 0. \tag{3}$$

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

$$\mathsf{P}(A + B + \cdots + N) = \mathsf{P}(A) + \mathsf{P}(B) + \cdots + \mathsf{P}(N) \tag{4}$$

(der Additionssatz).

Wenn $\mathsf{P}(A) > 0$ ist, so nennt man den Quotienten

$$\mathsf{P}_A(B) = \frac{\mathsf{P}(AB)}{\mathsf{P}(A)} \tag{5}$$

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

$$\mathsf{P}(AB) = \mathsf{P}(A)\,\mathsf{P}_A(B). \tag{6}$$

Ein Induktionsschluß ergibt sodann die allgemeine Formel

$$\mathsf{P}(A_1 A_2 \ldots A_n) = \mathsf{P}(A_1)\,\mathsf{P}_{A_1}(A_2)\,\mathsf{P}_{A_1 A_2}(A_3) \ldots \mathsf{P}_{A_1 A_2 \ldots A_{n-1}}(A_n) \tag{7}$$

(der Multiplikationssatz).

Man beweist auch leicht folgende Formeln:

$$\mathsf{P}_A(B) \geqq 0, \tag{8}$$

$$\mathsf{P}_A(E) = 1, \tag{9}$$

$$\mathsf{P}_A(B + C) = \mathsf{P}_A(B) + \mathsf{P}_A(C). \tag{10}$$

Vergleicht man diese Formeln (8) bis (10) mit den Axiomen III bis V, so ergibt sich, daß *das Mengensystem* $\mathfrak{F}$ *mit der Mengenfunktion* $\mathsf{P}_A(B)$

"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]

Definition

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 $A_1, A_2, \ldots \in \mathcal{A}$ folgen soll, dass auch $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$.
- Für endliches Ω und für $\Omega := \mathbb{R}$ sind passende Ereignissysteme schon lange bekannt.

$$
\begin{aligned}
\Omega \text{ ist endlich} \quad &\Rightarrow \text{Man wählt für } \mathcal{A} \text{ die Potenzmenge von } \Omega \\
\Omega \text{ ist } \mathbb{R} \quad &\Rightarrow \text{Man wählt für } \mathcal{A} \text{ die } \textit{Borelsche } \sigma\textit{-Algebra } \mathcal{B}(\mathbb{R}) \\
\Omega \text{ ist } \mathbb{R}^n \quad &\Rightarrow \text{Man wählt für } \mathcal{A} \text{ die } \textit{Borelsche } \sigma\textit{-Algebra } \mathcal{B}(\mathbb{R}^n)
\end{aligned}
$$

Definition

Theorem (Ereignissystem bei endlicher Ergebnismenge)

$\Omega := \{\omega_1, \omega_2, ..., \omega_n\}$ 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 Ereignisssytem 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 $A_1, A_2, ..., A_n \subseteq \Omega$. Dann ist $\cup_{i=1}^n A_i$ 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 $\cup_{i=1}^n A_i$ eine Teilmenge von Ω und damit auch die dritte σ-Algebra Eigenschaft erfüllt.

Definition

Wahrscheinlichkeitsmaß $\mathbb{P}$

- $(\Omega, \mathcal{A})$ ist die *strukturelle Basis* eines Wahrscheinlichkeitsraummodells.
- $\mathbb{P}$ repräsentiert die probabilistischen Charakteristika eines Wahrscheinlichkeitsraummodells.
- $\mathbb{P}$ entspricht also der *funktionellen Basis* eines Wahrscheinlichkeitsraummmodells.
- Es gilt $\mathbb{P} : \mathcal{A} \to [0, 1]$, $\mathbb{P}$ 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.

Definition

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 $\Omega, \mathcal{A}$ 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

Erste Eigenschaften

Theorem (Wahrscheinlichkeit des unmöglichen Ereignisses)

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

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

Beweis

Für $i = 1, 2, ...$ sei $A_i := \emptyset$. Dann ist $A_1, A_2, ...$ 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

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

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

$\square$

Theorem (σ-Additivität bei endlichen Folgen disjunkter Ereignisse)

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

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

<u>Beweis</u>

Wir betrachten eine unendliche Folge von paarweise disjunkten Ereignissen $A_1, A_2, \ldots$ 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

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

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

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

$\square$

Definition

Erste Eigenschaften

Wahrscheinlichkeitsfunktionen

Beispiele

Selbstkontrollfragen

Definition (Wahrscheinlichkeitsfunktion)

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

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

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

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

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

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

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

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

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

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 $A_1, A_2, \ldots \in \mathcal{A}$. Dann gilt

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

und damit die σ-Addivität von $\mathbb{P}$. $\square$

Definition

Erste Eigenschaften

Wahrscheinlichkeitsfunktionen

Beispiele

Selbstkontrollfragen

Beispiele

Aus dem bis hierin Gesagtem lässst 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 geignetes Ereignissystem; im Falle einer endlichen Ergebnismenge bietet sich die Potenzmenge $\mathcal{P}(\Omega)$, 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 inbesondere 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 Werfen eines Würfels. Es ist sicherlich sinnvoll, die Ergebnismenge als $\Omega := \{1, 2, 3, 4, 5, 6\}$ zu definieren. Allerdings wäre auch die Definition von $\Omega := \{\cdot, \cdot\cdot, \cdot\cdot\cdot, \cdot\cdot\cdot\cdot, \cdot\cdot\cdot\cdot\cdot, \cdot\cdot\cdot\cdot\cdot\cdot\}$ 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.

Tabelle 1: Ausgewählte Ereignisse beim Modell des Werfen 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 $(\Omega, \mathcal{A})$ 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

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

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

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

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

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

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

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

Beispiele

Gleichzeitiges Würfeln mit einem blauem und einem roten Würfel

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

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

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.

Tabelle 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 gefallenen Zahlen ist Vier	$\omega \in A = \{(1,3),(3,1),(2,2)\}$

Beispiele

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

Die Definition des Messraum $(\Omega, \mathcal{A})$ 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

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

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

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

Beispiele

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.

Tabelle 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 $(\Omega, \mathcal{A})$ ist damit abgeschlossen.

Beispiele

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

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

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

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

Gleichzeitiges Werfen von zwei Münzen

Wir modellieren das gleichzeitige Werfen zweier Münzen. Basierend auf dem Modell des einfachen Münzwurfs ist es sinnvoll, die Ergebnismenge als $\Omega := \{HH, HT, TH, TT\}$ zu definieren. Die Potenzmenge $\mathcal{A} := \mathcal{P}(\Omega)$ enthält wiederum alle $2^{|\Omega|} = 2^4 = 16$ möglichen Ereignisse. In untenstehender Tabelle listen wir vier davon.

Tabelle 4: Ausgewählte Ereignisse beim Modell des zweifachen Werfens einer Münze

Beschreibung	Mengenform
Kopf fällt im ersten Wurf	$\omega \in A = \{HH, HT\}$
Kopf fällt im zweiten Wurf	$\omega \in A = \{HH, TH\}$
Kopf fällt nicht	$\omega \in A = \{TT\}$
Zahl fällt mindestens einmal	$\omega \in A = \{HT, TH, TT\}$

Die Definition des Messraum $(\Omega, \mathcal{A})$ 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 fairer Münzen könnte man etwa

$$\mathbb{P}(\{HH\}) = \mathbb{P}(\{HT\}) = \mathbb{P}(\{TH\}) = \mathbb{P}(\{TT\}) = \frac{1}{4} \tag{19}$$

wählen.

© 2023 Dirk Ostwald CC BY-NC-SA 4.0

Definition

Erste Eigenschaften

Wahrscheinlichkeitsfunktionen

Beispiele

Selbstkontrollfragen

© 2023 Dirk Ostwald CC BY-NC-SA 4.0

Selbstkontrollfragen

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

References

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 2022/23

Prof. Dr. Dirk Ostwald

(3) Elementare Wahrscheinlichkeiten

Repräsentation zentraler Eigenschaften

Modellierung

Modell

Wahrscheinlichkeitstheorie

Probabilistisches Modell

$(\Omega, \mathcal{A}, \mathbb{P}), \xi : \Omega \to \mathbb{R}$

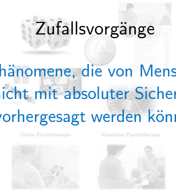
Wahrscheinlichkeitsrechnung

$\mathbb{P}_\xi(S) = \mathbb{P}(\xi^{-1}(S))$

Realität

Zufallsvorgänge

Phänomene, die von Menschen nicht mit absoluter Sicherheit vorhergesagt werden können.



Vorhersagen

Quantifizierung von Unsicherheit

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

Modellierung

Modell

Wahrscheinlichkeitstheorie

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

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

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

Realität

Zufallsvorgang

Werfen eines fairen Würfels

Vorhersagen

Wenn ich **nicht weiß**, ob eine Augenzahl größer als Drei gefallen ist, dann ist die Wahrscheinlichkeit, dass die Augenzahl gerade ist $\mathbf{1/2}$.

Wenn ich **weiß**, dass eine Augenzahl größer als Drei gefallen ist, dann ist die Wahrscheinlichkeit dafür, dass die Augenzahl gerade ist $\mathbf{2/3}$.

Interpretation

Gemeinsame Wahrscheinlichkeiten

Weitere Eigenschaften

Unabhängige Ereignisse

Bedingte Wahrscheinlichkeiten

Selbstkontrollfragen

Interpretation

Gemeinsame Wahrscheinlichkeiten

Weitere Eigenschaften

Unabhängige Ereignisse

Bedingte Wahrscheinlichkeiten

Selbstkontrollfragen

Wiederholung der Definition

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

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

mit den Eigenschaften

- $\mathbb{P}(A) \geq 0$ für alle $A \in \mathcal{A}$ (Nicht-Negativität),
- $\mathbb{P}(\Omega) = 1$ (Normiertheit),
- $\mathbb{P}\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ für paarweise disjunkte $A_i \in \mathcal{A}$ (σ-Additivität)

heißt ein *Wahrscheinlichkeitsmaß* oder einfach *Wahrscheinlichkeit*. Der Funktionswert $\mathbb{P}(A)$ heißt *Wahrscheinlichkeit des Ereignisses* $A \in \mathcal{A}$.

Interpretation

In der W-Theorie sind Wahrscheinlichkeiten anhand ihrer mathematischen Eigenschaften definiert.

Die Interpretation des mathematischen Wahrscheinlichkeitsbegriffs $\mathbb{P}(A)$ ist nicht eindeutig.

Es gibt mindestens zwei unterschiedliche Interpretationen:

Nach der **Frequentistischen Interpretation** ist $\mathbb{P}(A)$ die idealisierte relative Häufigkeit, mit der das Ereignis A unter den gleichen äußeren Bedingungen einzutreten pflegt. Zum Beispiel ist die frequentistische Interpretation von $\mathbb{P}(\{6\})$ im Modell des Werfen eines Würfels "Wenn man einen Würfel unendlich oft werfen würde und die relative Häufigkeit des Elementareignisses $\{6\}$ bestimmen würde, dann wäre diese relative Häufigkeit gleich $\mathbb{P}(\{6\})$."

Nach der **Bayesianischen Interpretation** ist $\mathbb{P}(A)$ der Grad der Sicherheit, den eine Beobachter:in aufgrund ihrer subjektiven Einschätzung der Lage dem Eintreten des Ereignisses A zumisst. Zum Beispiel ist die Bayesianische Interpretation von $\mathbb{P}(\{6\})$ im Modell des Werfen eines Würfels "Basierend auf meinen Erfahrung mit dem Werfen eines Würfels bin ich mir zu $\mathbb{P}(\{6\}) \cdot 100$ Prozent sicher, dass der Würfel beim nächsten Wurf eine 6 zeigt."

Interpretation

In Modellen von tatsächlich wiederholbaren Zufallsvorgängen wie dem Werfen eines Würfels ist der Unterschied zwischen Frequentistischer und Bayesianischer Interpretation eher subtil. Es gibt aber viele Zufallsvorgänge, die mit Wahrscheinlichkeiten beschrieben werden, bei denen aufgrund ihrer Einmaligkeit eine frequentistische Interpretation jedoch nicht sinnvoll ist. Zum Beispiel machen Aussagen der Form "Die Wahrscheinlichkeit, dass sich die Temperatur der Erde bis zum Jahr 2100 nur um 2° Celsius erhöht, wenn der CO_2 Ausstoß sofort auf Null gesetzt würde, ist 0.5" nur unter der Bayesianischen Interpretation Sinn.

In Hinblick auf die Definitionen und Rechenregeln für Wahrscheinlichkeiten unterscheiden sich beide Interpretationen allerdings nicht. Sowohl die frequentistische als auch die Bayesianisch geprägte probabilistische Datenanalyse haben mit der Wahrscheinlichkeitstheorie ein identisches mathematisches Bezugssystem.

Wir werden also erst an späterer Stelle wieder auf unterschiedliche Herangehensweisen in der probabilistischen Datenanalyse, die sich aus den unterschiedlichen Interpretation des Wahrscheinlichkeitsbegriffes ergeben, zurückkommen.

Generell kann man vielleicht sagen, dass die Bayesianische Interpretation mathematischer Wahrscheinlichkeiten logisch schlüssiger ist, in vielen wichtigen Anwendungsfeldern wie der Psychologie oder Biomedizin, frequentistisch geprägte Datenanalysen aber weiterhin vorherrschen.

Interpretation

Gemeinsame Wahrscheinlichkeiten

Weitere Eigenschaften

Unabhängige Ereignisse

Bedingte Wahrscheinlichkeiten

Selbstkontrollfragen

Definition (Gemeinsame Wahrscheinlichkeit)

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

$$\mathbb{P}(A \cap B) \tag{2}$$

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

$$A := \{2,4,6\} \quad \text{Es fällt eine gerade Augenzahl}$$
$$B := \{4,5,6\} \quad \text{Es fällt eine Augenzahl größer als Drei}$$

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

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

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 *inklusiven 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

$$A \cup B = \{2, 4, 6\} \cup \{4, 5, 6\} = \{2, 4, 5, 6\} \tag{4}$$

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

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

Interpretation

Gemeinsame Wahrscheinlichkeiten

Weitere Eigenschaften

Unabhängige Ereignisse

Bedingte Wahrscheinlichkeiten

Selbstkontrollfragen

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. $A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

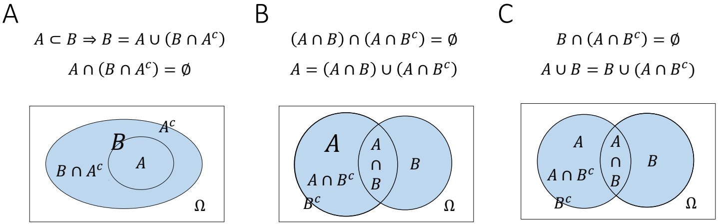
Bemerkungen

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

Weitere Eigenschaften

<u>Beweis</u>

Die zweite, dritte, und vierte Aussage dieses Theorems basieren auf elementaren mengentheoretischen Aussagen und der σ-Addivität 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.



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

$$\mathbb{P}(\Omega) = 1 \Leftrightarrow \mathbb{P}(A^c \cup A) = 1 \Leftrightarrow \mathbb{P}(A^c) + \mathbb{P}(A) = 1 \Leftrightarrow \mathbb{P}(A^c) = 1 - \mathbb{P}(A). \tag{6}$$

Weitere Eigenschaften

Beweis (fortgeführt)

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

$$A \subset B \Rightarrow B = A \cup (B \cap A^c) \text{ mit } A \cap (B \cap A^c) = \emptyset. \tag{7}$$

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

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c). \tag{8}$$

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)

$$(A \cap B) \cap (A \cap B^c) = \emptyset \text{ und } A = (A \cap B) \cup (A \cap B^c). \tag{9}$$

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

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) \Leftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A \cap B^c) \tag{10}$$

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

$$B \cap (A \cap B^c) = \emptyset \text{ und } A \cup B = B \cup (A \cap B^c). \tag{11}$$

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

$$\mathbb{P}(A \cup B) = \mathbb{P}(B) + \mathbb{P}(A \cap B^c). \tag{12}$$

Mit 3. folgt dann

$$\mathbb{P}(A \cup B) = \mathbb{P}(B) + \mathbb{P}(A) - \mathbb{P}(A \cap B) \tag{13}$$

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

Interpretation

Gemeinsame Wahrscheinlichkeiten

Weitere Eigenschaften

Unabhängige Ereignisse

Bedingte Wahrscheinlichkeiten

Selbstkontrollfragen

Unabhängige Ereignisse

Definition (Unabhängige Ereignisse)

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

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \tag{14}$$

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

$$\mathbb{P}\left(\cap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j). \tag{15}$$

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:

  $\mathbb{P}(A)\mathbb{P}(B) > 0$, aber $\mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0$, 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 Unabhängkeit erschließt sich im Kontext *bedingter Wahrscheinlichkeiten*.

Interpretation

Gemeinsame Wahrscheinlichkeiten

Weitere Eigenschaften

Unabhängige Ereignisse

Bedingte Wahrscheinlichkeiten

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

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

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

$$\mathbb{P}(\ |B) : \mathcal{A} \to [0,1], A \mapsto \mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{17}$$

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
    - $\mathbb{P}\left(\cup_{i=1}^{\infty} A_i|B\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i|B)$ für paarweise disjunkte $A_1, A_2, ... \in \mathcal{A}$.
- $\Rightarrow$ 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

$$\mathbb{P}(\{2,4,6\}) = \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = 1/6 + 1/6 + 1/6 = 3/6 \tag{18}$$

$$\mathbb{P}(\{4,5,6\}) = \mathbb{P}(\{4\}) + \mathbb{P}(\{5\}) + \mathbb{P}(\{6\}) = 1/6 + 1/6 + 1/6 = 3/6. \tag{19}$$

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

$$\mathbb{P}(A \cap B) = \mathbb{P}(\{2,4,6\} \cap \{4,5,6\}) = \mathbb{P}(\{4,6\}) = \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = 1/6 + 1/6 = 2/6 \tag{20}$$

hat. Nach Definition der bedingten Wahrscheinlichkeit ergibt sich dann direkt

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\{4,6\})}{\mathbb{P}(\{4,5,6\})} = \frac{2/6}{3/6} = 2/6 \cdot 6/3 = 2/3. \tag{21}$$

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 (Bedingte Wahrscheinlichkeit unter Unabhängigkeit)

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

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A). \tag{22}$$

Bemerkungen

- Bei Unabhängigkeit von A und B ist es irrelevant, ob auch B eintritt, $\mathbb{P}(A)$ bleibt gleich.
- Stochastische Unabhängigkeit wird also als $\mathbb{P}(A)\mathbb{P}(B)$ modelliert, damit $\mathbb{P}(A|B) = \mathbb{P}(A)$ gilt.

$\Rightarrow$ Die stochastische Unabhängigkeit zweier Ereignisse bedeutet, dass das Wissen um das Eintreten eines der beiden Ereignisse die Wahrscheinlichkeit für das Eintreten des anderen Ereignisses nicht ändert. Andersherum bedeutet stochastische Abhängigkeit, dass das Wissen um das Eintreten eines der beiden Ereignisse die Wahrscheinlichkeit für das Eintreten des anderen Ereignisses verändert, also erhöht oder verringert.

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 $P(B) > 0$. Dann gilt

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A). \tag{23}$$

Beweis

Mit der Definition der jeweiligen bedingten Wahrscheinlichkeit folgen direkt

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \Leftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) \tag{24}$$

und

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} \Leftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A). \tag{25}$$

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.

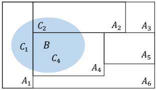
Theorem (Gesetz der totalen Wahrscheinlichkeit)

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

$$\mathbb{P}(B) = \sum_{i=1}^{k} \mathbb{P}(B|A_i)\mathbb{P}(A_i). \tag{26}$$

Beweis

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



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

$\square$

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

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{i=1}^k \mathbb{P}(B|A_i)\mathbb{P}(A_i)}. \tag{27}$$

Beweis

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

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{i=1}^k \mathbb{P}(B|A_i)\mathbb{P}(A_i)}. \tag{28}$$

□

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 Unabhängikeit 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.

© 2023 Dirk Ostwald CC BY-SA 4.0

Interpretation

Gemeinsame Wahrscheinlichkeiten

Weitere Eigenschaften

Unabhängige Ereignisse

Bedingte Wahrscheinlichkeiten

Selbstkontrollfragen

Selbstkontrollfragen

1. Erläutern Sie die Frequentistische Interpretation der Wahrscheinlichkeit eines Ereignisses.
2. Erläutern Sie die Bayesianische Interpretation der Wahrscheinlichkeit eines Ereignisses.
3. Geben Sie die Definition der gemeinsamen Wahrscheinlichkeit zweier Ereignisse wieder.
4. Erläutern Sie die intuitive Bedeutung der gemeinsamen Wahrscheinlichkeit zweier Ereignisse.
5. Geben Sie das Theorem zu weiteren Eigenschaften von Wahrscheinlichkeiten wieder.
6. Geben Sie die Definition der Unabhängigkeit zweier Ereignisse wieder.
7. Geben Sie die Definition der bedingten Wahrscheinlichkeit eines Ereignisses und der bedingten Wahrscheinlichkeit wieder.
8. Geben Sie das Theorem zur bedingten Wahrscheinlichkeit unter Unabhängigkeit wieder.
9. Erläutern Sie den Begriff der stochastischen Unabhängigkeit vor dem Hintergrund des Theorems zur bedingten Wahrscheinlichkeit unter Unabhängigkeit.
10. Geben Sie das Theorem zu gemeinsamen und bedingten Wahrscheinlichkeiten wieder.
11. Geben Sie das Gesetz von der totalen Wahrscheinlichkeit wieder.
12. Geben Sie das Theorem von Bayes wieder.
13. Erläutern Sie das Theorem von Bayes im Rahmen der Bayesianischen Inferenz.
14. Beweisen Sie das Theorem von Bayes.

Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2022/23

Prof. Dr. Dirk Ostwald

(4) Zufallsvariablen

Repräsentation zentraler Eigenschaften

Modellierung

Modell

Wahrscheinlichkeitstheorie

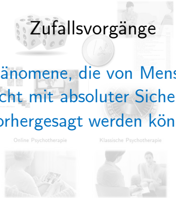
Probabilistisches Modell

$(\Omega, \mathcal{A}, \mathbb{P}), \xi : \Omega \to \mathbb{R}$

Wahrscheinlichkeitsrechnung

$\mathbb{P}_\xi(S) = \mathbb{P}(\xi^{-1}(S))$

Realität



Zufallsvorgänge

Phänomene, die von Menschen nicht mit absoluter Sicherheit vorhergesagt werden können.

Vorhersagen

Quantifizierung von Unsicherheit

Jedes Augenzahlpaar kommt im Mittel gleich häufig vor.

Basierend auf der Physik sollte jedes Augenzahlpaar die gleiche Wahrscheinlichkeit haben.

Modellierung

Modell

Wahrscheinlichkeitstheorie

$\Omega \coloneqq \{(r, b) | r \in \mathbb{N}_6, b \in \mathbb{N}_6\}$

$\mathcal{A} \coloneqq \mathcal{P}(\Omega)$ $\mathbb{P}(\{\omega\}) \coloneqq 1/36$

$\xi: \Omega \to \mathbb{R}, \xi(r, b) \coloneqq r + b$

Realität

Zufallsvorgang

Gleichzeitiges Werfen eines
roten und eines blauen Würfels

Vorhersagen

Die Summe der Augenzahlen ist eine Zufallsvariable mit Verteilung $\mathbb{P}_\xi$.

Konstruktion, Definition, Notation, Intuition

Wahrscheinlichkeitsmassefunktionen

Wahrscheinlichkeitsdichtefunktionen

Kumulative Verteilungsfunktionen

Selbstkontrollfragen

Konstruktion, Definition, Notation, Intuition

Wahrscheinlichkeitsmassefunktionen

Wahrscheinlichkeitsdichtefunktionen

Kumulative Verteilungsfunktionen

Selbstkontrollfragen

Konstruktion von Zufallsvariablen und Verteilungen

- Es seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $\xi : \Omega \to \mathcal{X}$ 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

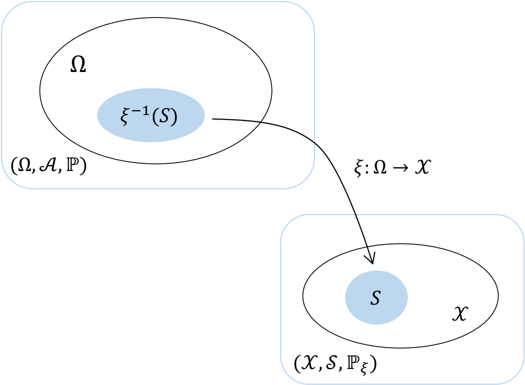
$$\xi^{-1}(S) := \{\omega \in \Omega | \xi(\omega) \in S\}. \tag{1}$$

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

$$\mathbb{P}_\xi : \mathcal{S} \to [0,1], S \mapsto \mathbb{P}_\xi(S) := \mathbb{P}\left(\xi^{-1}(S)\right) = \mathbb{P}\left(\{\omega \in \Omega | \xi(\omega) \in S\}\right) \tag{2}$$

  zugeordnet werden.
- ξ heißt nun *Zufallsvariable* und $\mathbb{P}_\xi$ 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.

Konstruktion, Definition, Notation, Intuition



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

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 $\xi : \Omega \to \mathcal{X}$ mit der *Messbarkeitseigenschaft*

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

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

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

  beschrieben, wobei $\mathcal{X} := \{2, 3, ..., 12\}$.
- $\mathcal{S} := \mathcal{P}(\mathcal{X})$ ist eine sinnvolle σ-Algebra auf $\mathcal{X}$.
- Mithilfe der σ-Addivität von $\mathbb{P}$ können wir die Verteilung $\mathbb{P}_\xi$ 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 zweier Würfel)

Bestimmung der Verteilung $\mathbb{P}_\xi$ von ξ

$$
\begin{aligned}
\mathbb{P}_\xi(\{2\}) &= \mathbb{P}\left(\xi^{-1}(\{2\})\right) &&= \mathbb{P}(\{(1,1)\}) &&= \tfrac{1}{36} \\
\mathbb{P}_\xi(\{3\}) &= \mathbb{P}\left(\xi^{-1}(\{3\})\right) &&= \mathbb{P}(\{(1,2),(2,1)\}) &&= \tfrac{2}{36} \\
\mathbb{P}_\xi(\{4\}) &= \mathbb{P}\left(\xi^{-1}(\{4\})\right) &&= \mathbb{P}(\{(1,3),(3,1),(2,2)\}) &&= \tfrac{3}{36} \\
\mathbb{P}_\xi(\{5\}) &= \mathbb{P}\left(\xi^{-1}(\{5\})\right) &&= \mathbb{P}(\{(1,4),(4,1),(2,3),(3,2)\}) &&= \tfrac{4}{36} \\
\mathbb{P}_\xi(\{6\}) &= \mathbb{P}\left(\xi^{-1}(\{6\})\right) &&= \mathbb{P}(\{(1,5),(5,1),(2,4),(4,2),(3,3)\}) &&= \tfrac{5}{36} \\
\mathbb{P}_\xi(\{7\}) &= \mathbb{P}\left(\xi^{-1}(\{7\})\right) &&= \mathbb{P}(\{(1,6),(6,1),(2,5),(5,2),(3,4),(4,3)\}) &&= \tfrac{6}{36} \\
\mathbb{P}_\xi(\{8\}) &= \mathbb{P}\left(\xi^{-1}(\{8\})\right) &&= \mathbb{P}(\{(2,6),(6,2),(3,5),(5,3),(4,4)\}) &&= \tfrac{5}{36} \\
\mathbb{P}_\xi(\{9\}) &= \mathbb{P}\left(\xi^{-1}(\{9\})\right) &&= \mathbb{P}(\{(3,6),(6,3),(4,5),(5,4)\}) &&= \tfrac{4}{36} \\
\mathbb{P}_\xi(\{10\}) &= \mathbb{P}\left(\xi^{-1}(\{10\})\right) &&= \mathbb{P}(\{(4,6),(6,4),(5,5)\}) &&= \tfrac{3}{36} \\
\mathbb{P}_\xi(\{11\}) &= \mathbb{P}\left(\xi^{-1}(\{11\})\right) &&= \mathbb{P}(\{(5,6),(6,5)\}) &&= \tfrac{2}{36} \\
\mathbb{P}_\xi(\{12\}) &= \mathbb{P}\left(\xi^{-1}(\{12\})\right) &&= \mathbb{P}(\{(6,6)\}) &&= \tfrac{1}{36}
\end{aligned}
$$

Definition (Notation für Zufallsvariablen)

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

$$
\begin{aligned}
\{\xi \in S\} &:= \{\omega \in \Omega | \xi(\omega) \in S\}, S \subset \mathcal{X}, \\
\{\xi = x\} &:= \{\omega \in \Omega | \xi(\omega) = x\}, x \in \mathcal{X}, \\
\{\xi \le x\} &:= \{\omega \in \Omega | \xi(\omega) \le x\}, x \in \mathcal{X}, \\
\{\xi < x\} &:= \{\omega \in \Omega | \xi(\omega) < x\}, x \in \mathcal{X}.
\end{aligned}
$$

Aus diesen Konventionen folgen exemplarisch die folgenden Konventionen für Verteilungen:

$$
\begin{aligned}
\mathbb{P}_\xi\left(\xi \in S\right) = \mathbb{P}\left(\{\xi \in S\}\right) = \mathbb{P}\left(\{\omega \in \Omega | \xi(\omega) \in S\}\right), S \subset \mathcal{X} \\
\mathbb{P}_\xi\left(\xi \le x\right) = \mathbb{P}\left(\{\xi \le x\}\right) = \mathbb{P}\left(\{\omega \in \Omega | \xi(\omega) \le x\}\right), x \in \mathcal{X}.
\end{aligned}
$$

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

$$
\begin{aligned}
\mathbb{P}\left(\xi \in S\right) = \mathbb{P}_\xi\left(\xi \in S\right), S \subset \mathcal{X}, \\
\mathbb{P}\left(\xi \le x\right) = \mathbb{P}_\xi\left(\xi \le S\right), x \in \mathcal{X}.
\end{aligned}
$$

Definition (Realisierung einer Zufallsvariable)

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

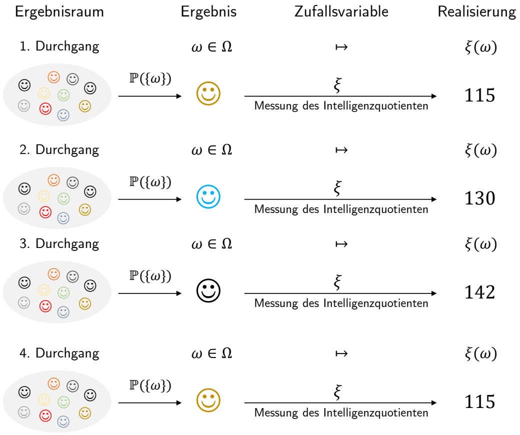
- In der Datenanalyse werden Daten typischerweise als Realisierungen von Zufallsvariablen modelliert.
- Da die Auswahl eines $\omega \in \Omega$ in einem Zufallsvorgang zufällig ist, erscheint $\xi(\omega)$ zufällig.

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     :  4 2
> xi(omega) :  6
```

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, $\xi : \Omega \to \mathbb{R}$, $\upsilon : \Omega \to \mathbb{R}$ seien reellwertige Zufallsvariablen und $c \in \mathbb{R}$ sei eine Konstante. Weiterhin seien

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

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 $\xi + c$, $c\xi$, $\xi + \upsilon$ und $\xi\upsilon$ 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 Hesse (2009), Seite 33-34.

Konstruktion, Definition, Notation, Intuition

Wahrscheinlichkeitsmassefunktionen

Wahrscheinlichkeitsdichtefunktionen

Kumulative Verteilungsfunktionen

Selbstkontrollfragen

Definition (Diskrete ZV, Wahrscheinlichkeitsmassefunktion)

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

$$p : \mathcal{X} \to \mathbb{R}_{\geq 0}, x \mapsto p(x) \tag{6}$$

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 *Wahrscheinlichkeitsmassefunktion (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 bevorzugen wird den Begriff WMF.

© 2023 Dirk Ostwald CC BY-SA 4.0

Definition (Bernoulli-Zufallsvariable)

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

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

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

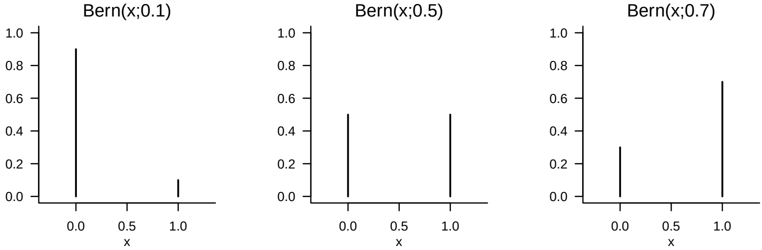
$$\text{Bern}(x; \mu) := \mu^x (1 - \mu)^{1-x}. \tag{8}$$

Bemerkungen

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

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

Bernoulli-Zufallsvariable



Definition (Binomial-Zufallsvariable)

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

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

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

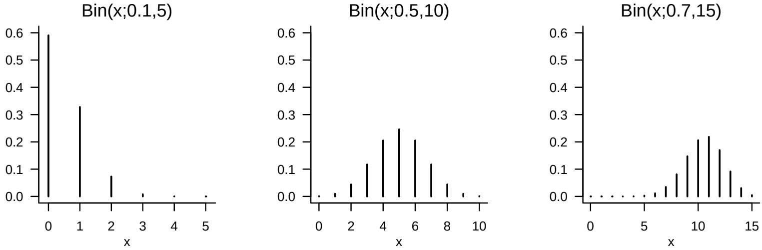
$$\text{Bin}(x; \mu, n) := \binom{n}{x} \mu^x (1-\mu)^{n-x}. \tag{11}$$

Bemerkung

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

Wahrscheinlichkeitsmassefunktionen

Binomial-Zufallsvariable



Definition (Diskret-gleichverteilte Zufallsvariable)

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

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

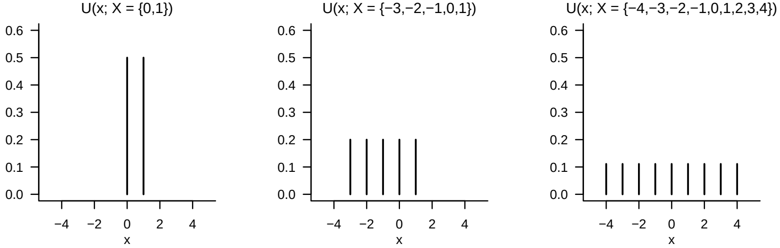
Dann sagen wir, dass ξ einer *diskreten Gleichverteilung* unterliegt und nennen ξ eine *diskret-gleichverteilte Zufallsvariable*. Wir kürzen dies mit $\xi \sim U(|\mathcal{X}|)$ ab. Die WMF einer diskret-gleichverteilten Zufallsvariable bezeichnen wir mit

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

Bemerkungen

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

Diskret-gleichverteilte Zufallsvariable



Konstruktion, Definition, Notation, Intuition

Wahrscheinlichkeitsmassefunktionen

Wahrscheinlichkeitsdichtefunktionen

Kumulative Verteilungsfunktionen

Selbstkontrollfragen

Definition (Kontinuierliche ZV, Wahrscheinlichkeitsdichtefunktion)

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

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

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

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

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 $\xi \sim N\left(\mu, \sigma^2\right)$ ab. Die WDF einer normalverteilten Zufallsvariable bezeichnen wir mit

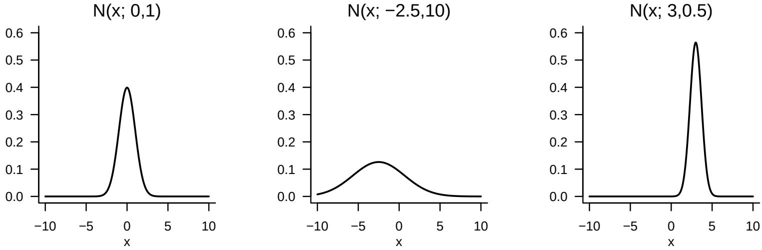
$$N\left(x; \mu, \sigma^2\right) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right). \tag{16}$$

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



Definition (Gamma-Zufallsvariable)

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

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

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 gammaverteilen Zufallsvariable bezeichnen wir mit

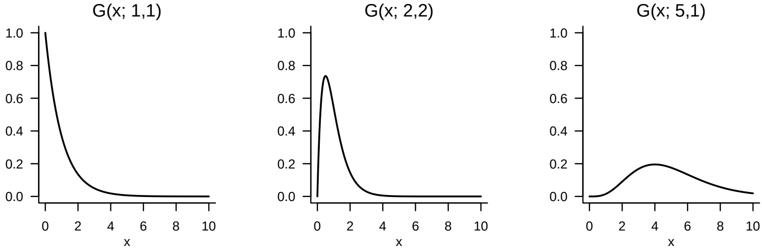
$$G(x; \alpha, \beta) := \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right). \tag{18}$$

Bemerkung

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

Wahrscheinlichkeitsdichtefunktionen

Gamma-Zufallsvariablen



Definition (Beta-Zufallsvariable)

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

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

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

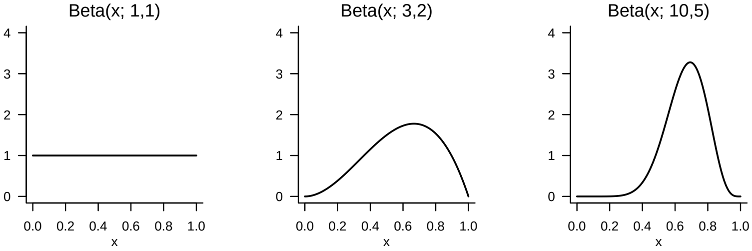
$$\text{Beta}(x; \alpha, \beta) := \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}. \tag{20}$$

Bemerkung

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

Wahrscheinlichkeitsdichtefunktionen

Beta-Zufallsvariable



Definition (Gleichverteilte Zufallsvariable)

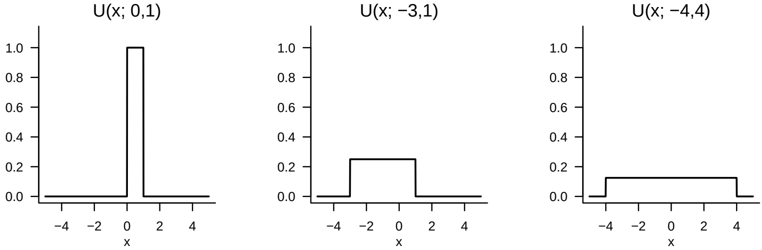
Es sei ξ eine kontinuierliche Zufallsvariable mit Ergebnisraum $\mathbb{R}$ und WDF

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

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

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

Gleichverteilte Zufallsvariablen



Konstruktion, Definition, Notation, Intuition

Wahrscheinlichkeitsmassefunktionen

Wahrscheinlichkeitsdichtefunktionen

Kumulative Verteilungsfunktionen

Selbstkontrollfragen

Definition (Kumulative Verteilungsfunktion)

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

$$P : \mathbb{R} \to [0,1], x \mapsto P(x) := \mathbb{P}(\xi \le x). \tag{23}$$

Bemerkungen

- KVFen 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 KVFen 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

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

Beweis

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

$$\Omega = \{\omega \in \Omega | \xi(\omega) > x\} \cup \{\omega \in \Omega | \xi(\omega) \le x\} = \{\xi > x\} \cup \{\xi \le x\}. \tag{25}$$

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

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

$\square$

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

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

Beweis

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

$$\{\xi \le x_1\} \cap \{x_1 < \xi \le x_2\} = \emptyset \text{ und } \{\xi \le x_1\} \cup \{x_1 < \xi \le x_2\} = \{\xi \le x_2\}. \tag{28}$$

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

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

□

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 $P(x_1) \le P(x_2)$.

(2) $\lim_{x\to-\infty} P(x) = 0$ und $\lim_{x\to\infty} P(x) = 1$.

(3) P ist *rechtsseitig stetig*, d.h., $P(x) = P(x^+) = \lim_{y\to x, y>x} P(y)$ 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 $x_1 < x_2$,

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

Also gilt

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

(2) Wir verzichten auf einen Beweis

(3) Wir definieren

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

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

$$\{\xi \leq x\} = \cap_{n=1}^{\infty} \{\xi \leq y_n\}. \tag{33}$$

Es gilt also

$$P(x) = \mathbb{P}(\{\xi \leq x\}) = \mathbb{P}(\cap_{n=1}^{\infty} \{\xi \leq y_n\}) = \lim_{n \to \infty} \mathbb{P}(\{\xi \leq y_n\}) = P(x^+), \tag{34}$$

wobei wir die dritte Gleichung unbegründet stehen lassen.

Kumulative Verteilungsfunktionen

Kumulative Verteilungsfunktionen von diskreten Zufallsvariablen

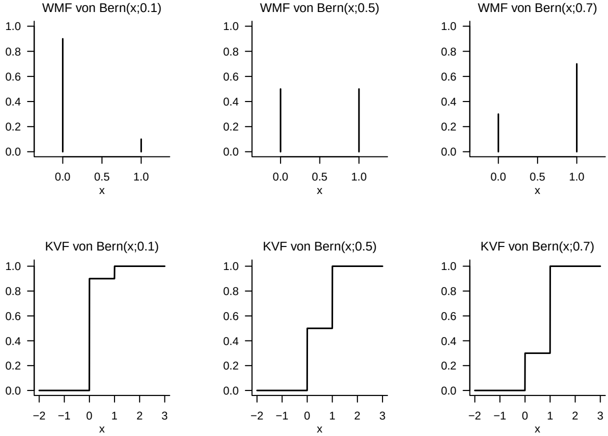
- Wenn $a < b$ und $\mathbb{P}(a < \xi < b) = 0$, dann ist P konstant horizontal auf $]a, b[$.
- An jedem Punkt x mit $\mathbb{P}(\xi = x) > 0$ springt die KVF um den Betrag $\mathbb{P}(\xi = x)$.
- $\Leftrightarrow$ An jedem Punkt x mit $p(x) > 0$ springt die KVF um den Betrag $p(x) > 0$.
- Generell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

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

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

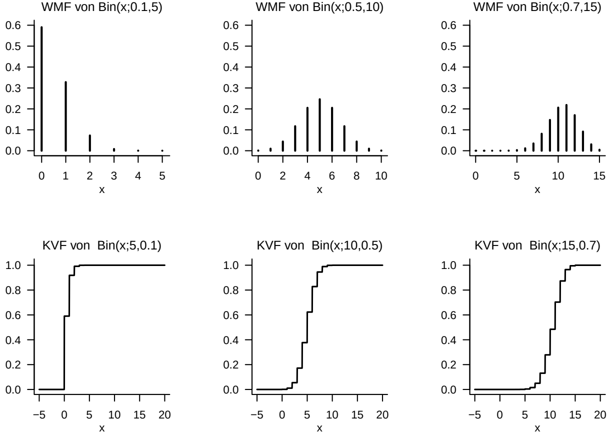
Kumulative Verteilungsfunktionen

Bernoulli-Zufallsvariablen



Kumulative Verteilungsfunktionen

Binomial-Zufallsvariablen



Theorem (Kumulative Verteilungsfunktionen von kontinuierlichen ZVen)

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

$$P(x) = \int_{-\infty}^{x} p(s)\,ds \text{ und } p(x) = \frac{d}{dx}P(x). \tag{36}$$

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.
- KVFen von kontinuierlichen ZV heißen auch kumulative Dichtefunktionen (KDFen).

Beispiel (Normalverteilung)

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

- Die WDF von ξ ist

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

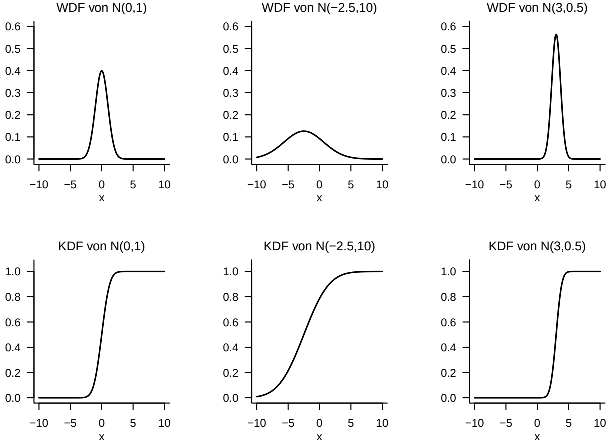
- Die KVF von ξ ist

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

- 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



Definition (Inverse Kumulative Verteilungsfunktion)

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

$$P^{-1} :]0,1[\to \mathbb{R}, q \mapsto P^{-1}(q) := \{x \in \mathbb{R} | P(x) = q\} \tag{37}$$

die *inverse kumulative Verteilungsfunktion von* ξ.

Bemerkungen

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

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

- Die inverse KVF von ξ ist

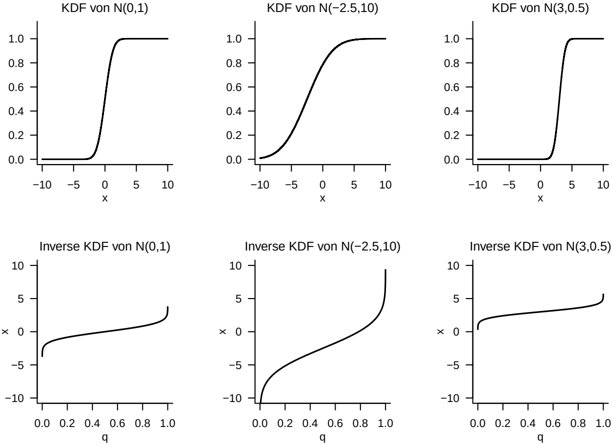
$$P^{-1} :]0,1[\to \mathbb{R}, q \mapsto P^{-1}(q) = \{x \in \mathbb{R} | P(x) = q\}. \qquad (39)$$

- 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
  - $\Phi(1.645) = 0.950$, $\Phi^{-1}(0.950) = \Phi^{-1}(1 - 0.050) = 1.645$.
  - $\Phi(1.960) = 0.975$, $\Phi^{-1}(0.975) = \Phi^{-1}(1 - \frac{0.050}{2})$.

© 2023 Dirk Ostwald CC BY-SA 4.0

Kumulative Verteilungsfunktionen

Normalverteilte Zufallsvariablen



Konstruktion, Definition, Notation, Intuition

Wahrscheinlichkeitsmassefunktionen

Wahrscheinlichkeitsdichtefunktionen

Kumulative Verteilungsfunktionen

Selbstkontrollfragen

Selbstkontrollfragen

1. Definieren Sie den Begriff der Zufallsvariable.
2. Erläutern Sie die Gleichung $\mathbb{P}_\xi(\xi = x) = \mathbb{P}(\{\xi = x\})$.
3. Erläutern Sie die Bedeutung von $\mathbb{P}(\xi = x)$.
4. Definieren Sie den Begriff der Wahrscheinlichkeitsmassefunktion.
5. Definieren Sie die Begriffe der Wahrscheinlichkeitsdichtefunktion.
6. Definieren Sie den Begriff der kumulativen Verteilungsfunktion.
7. Schreiben sie die Intervallwahrscheinlichkeit einer Zufallsvariable mithilfer ihrer KVF.
8. Definieren Sie die WDF und KVF einer normalverteilten Zufallsvariable.
9. Schreiben Sie den Wert $P(x)$ der KVF einer Zufallsvariable mithilfe ihrer WDF.
10. Schreiben Sie den Wert $p(x)$ der WDF einer Zufallsvariable mithilfe ihrer KVF.
11. Definieren Sie den Begriff der inversen Verteilungsfunktion.

Referenzen

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

Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2022/23

Prof. Dr. Dirk Ostwald

(5) Multivariate Verteilungen

Repräsentation zentraler Eigenschaften

Modellierung

Modell

Wahrscheinlichkeitstheorie

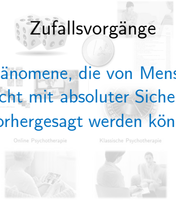
Probabilistisches Modell

$(\Omega, \mathcal{A}, \mathbb{P}), \xi : \Omega \to \mathbb{R}$

Wahrscheinlichkeitsrechnung

$\mathbb{P}_\xi(S) = \mathbb{P}(\xi^{-1}(S))$

Realität



Zufallsvorgänge

Phänomene, die von Menschen nicht mit absoluter Sicherheit vorhergesagt werden können.

Vorhersagen

Quantifizierung von Unsicherheit

Wir nehmen an, dass die BDI Score Erwartungswertabweichungen der Proband:innen Realisierungen **unabhängiger und identisch (u.i.v.)** normalverteilter Zufallsvariablen sind.

Modellierung

Modell

Wahrscheinlichkeitstheorie

$$y_{1j} = \mu_1 + \varepsilon_{1j}\,, \varepsilon_{1j} \sim N(0,\sigma^2), j = 1,\dots,n_1$$

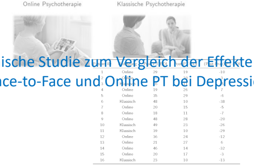
$$y_{2j} = \mu_2 + \varepsilon_{2j}\,, \varepsilon_{2j} \sim N(0,\sigma^2), j = 1,\dots,n_2$$

$\varepsilon_{1j}, j = 1,\dots,n_1$ u.i.v

$\varepsilon_{2j}, j = 1,\dots,n_2$ u.i.v

Realität

Zufallsvorgang



Klinische Studie zum Vergleich der Effekte von Face-to-Face und Online PT bei Depression

Vorhersagen

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

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

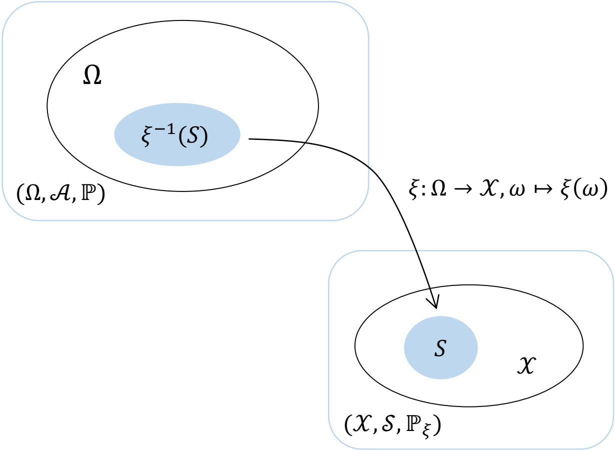
mit der *Messbarkeitseigenschaft*

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

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 $\xi_1, ..., \xi_n$ 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 $\xi := (\xi_1, ..., \xi_n)$.

Definition



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

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

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

$$\xi : \Omega \to \mathcal{X}, \omega \mapsto \xi(\omega) \tag{3}$$

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

$$\mathbb{P}_\xi : \mathcal{S} \to [0,1], S \mapsto \mathbb{P}_\xi(S) := \mathbb{P}(\xi^{-1}(S)) = \mathbb{P}\left(\{\omega \in \Omega | \xi(\omega) \in S\}\right) \tag{4}$$

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.

$$\begin{aligned}
\mathbb{P}_\xi(\xi \in S) &:= \mathbb{P}\left(\{\xi \in S\}\right) = \mathbb{P}\left(\{\omega \in \Omega | \xi(\omega) \in S\}\right) \\
\mathbb{P}_\xi(\xi = x) &:= \mathbb{P}\left(\{\xi = x\}\right) = \mathbb{P}\left(\{\omega \in \Omega | \xi(\omega) = x\}\right) \\
\mathbb{P}_\xi(\xi \le x) &:= \mathbb{P}\left(\{\xi \le x\}\right) = \mathbb{P}\left(\{\omega \in \Omega | \xi(\omega) \le x\}\right) \\
\mathbb{P}_\xi(x_1 \le \xi \le x_2) &:= \mathbb{P}\left(\{x_1 \le \xi \le x_2\}\right) = \mathbb{P}\left(\{\omega \in \Omega | x_1 \le \xi(\omega) \le x_2\}\right)
\end{aligned} \tag{5}$$

- Relationsoperatoren wie $\le$ werden hier *komponentenweise* verstanden.
- Zum Beispiel heißt $x \le y$ für $x, y \in \mathbb{R}^n$, dass $x_i \le y_i$ für alle $i = 1, ..., n$.

Definition (Multivariate kumulative Verteilungsfunktionen)

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

$$P_\xi : \mathcal{X} \to [0,1],\, x \mapsto P_\xi(x) := \mathbb{P}_\xi(\xi \le x) \tag{6}$$

multivariate kumulative Verteilungsfunktion von ξ.

Bemerkung

- Multivariate kumulative Verteilungsfunktionen können zur Definition von multivariaten Verteilungen genutzt werden, häufiger ist allerdings die Definition multivariater Verteilungen durch multivariate Wahrscheinlichkeitsmasse- oder Wahrscheinlichkeitsdichtefunktionen.

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

$$p_\xi : \mathcal{X} \to [0,1], x \mapsto p_\xi(x) \tag{7}$$

existiert, für die gilt

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

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

Ein entsprechende Funktion p heißt *multivariate Wahrscheinlichkeitsmassefunktion (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

$$p_\xi : \{1,2,3\} \times \{1,2,3,4\} \to [0,1], (x_1, x_2) \mapsto p_\xi(x_1, x_2) \tag{8}$$

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

Definition (Kontinuierlicher Zufallsvektor, Multivariate WDF)

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

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

existiert, für die gilt

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

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

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

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

Definition

Multivariate Verteilungen

Marginalverteilungen

Bedingte Verteilungen

Unabhängige Zufallsvariablen

Selbstkontrollfragen

Marginalverteilungen

Definition (Univariate Marginalverteilung)

$(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum, $(\mathcal{X}, \mathcal{S})$ sei ein n-dimensionaler Messraum, $\xi : \Omega \to \mathcal{X}$ sei ein Zufallsvektor, $\mathbb{P}_\xi$ 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 ξ_i. Dann heißt die durch

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

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

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

(2) $\xi = (\xi_1, ..., \xi_n)$ 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

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

Bemerkungen

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

Marginalverteilungen

Beispiel (Marginale Wahrscheinlichkeitsmassefunktionen)

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 $\sum_{x_1=1}^3 p_{\xi_1}(x_1) = 1$ und $\sum_{x_2=1}^3 p_{\xi_2}(x_2) = 1$ gilt.

Beispiel (Marginale Wahrscheinlichkeitsmassefunktionen)

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

$$\frac{0 + 22 + 12}{100} = \frac{0}{100} + \frac{22}{100} + \frac{12}{100} = 0.00 + 0.22 + 0.12 = 0.34 \tag{14}$$

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

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

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \tag{15}$$

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

$$A := \{\xi_1 \in S_1\} \text{ und } B := \{\xi_2 \in S_2\} \tag{16}$$

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

$$\mathbb{P}(\{\xi_1 \in S_1\}|\{\xi_2 \in S_2\}) = \frac{\mathbb{P}(\{\xi_1 \in S_1\} \cap \{\xi_2 \in S_2\})}{\mathbb{P}(\{\xi_2 \in S_2\})}. \tag{17}$$

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

$$p_{\xi_1|\xi_2=x_2} : \mathcal{X}_1 \to [0,1], x_1 \mapsto p_{\xi_1|\xi_2=x_2}(x_1|x_2) := \frac{p_{\xi_1,\xi_2}(x_1, x_2)}{p_{\xi_2}(x_2)} \tag{18}$$

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

$$p_{\xi_2|\xi_1=x_1} : \mathcal{X}_2 \to [0,1], x_2 \mapsto p_{\xi_2|\xi_1=x_2}(x_1|x_2) := \frac{p_{\xi_1,\xi_2}(x_1, x_2)}{p_{\xi_1}(x_1)} \tag{19}$$

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

$$p_{\xi_1|\xi_2}(x_1|x_2) = \frac{p_{\xi_1,\xi_2}(x_1, x_2)}{p_{\xi_2}(x_2)} = \frac{\mathbb{P}(\{\xi_1 = x_1\} \cap \{\xi_2 = x_2\})}{\mathbb{P}(\{\xi_2 = x_2\})}. \tag{20}$$

- Bedingte Verteilungen sind (lediglich) normalisierte gemeinsame Verteilungen.

Beispiel (Bedingte Wahrscheinlichkeitsmassefunktionen)

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

$$p_{\xi_1|\xi_2=x_2} : \mathbb{R} \to \mathbb{R}_{\geq 0}, x_1 \mapsto p_{\xi_1|\xi_2=x_2}(x_1|x_2) := \frac{p_{\xi_1,\xi_2}(x_1, x_2)}{p_{\xi_2}(x_2)} \tag{21}$$

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

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

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

$$\mathbb{P}_\xi(\xi_1 \in S_1, \xi_2 \in S_2) = \mathbb{P}_{\xi_1}(\xi_1 \in S_1)\mathbb{P}_{\xi_2}(\xi_2 \in S_2). \tag{23}$$

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*.

© 2023 Dirk Ostwald CC BY-SA 4.0

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

$$\xi_1 \text{ und } \xi_2 \text{ sind unabhängige Zufallsvariablen} \Leftrightarrow$$
$$p_\xi(x_1, x_2) = p_{\xi_1}(x_1) p_{\xi_2}(x_2) \text{ für alle } (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2. \quad (24)$$

(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

$$\xi_1 \text{ und } \xi_2 \text{ sind unabhängige Zufallsvariablen} \Leftrightarrow$$
$$p_\xi(x_1, x_2) = p_{\xi_1}(x_1) p_{\xi_2}(x_2) \text{ für alle } (x_1, x_2) \in \mathbb{R}^2. \quad (25)$$

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.

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

$$p_\xi(1,1) = 0.10 \neq 0.08 = 0.40 \cdot 0.20 = p_{\xi_1}(1)p_{\xi_2}(1) \tag{26}$$

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

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

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.

Unabhängige Zufallsvariablen

Definition (n unabhängige Zufallsvariablen)

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

$$\mathbb{P}_\xi(\xi_1 \in S_1, ..., \xi_n \in S_n) = \prod_{i=1}^n \mathbb{P}_{\xi_i}(\xi_i \in S_i). \tag{27}$$

Wenn der Zufallsvektor eine n-dimensionale WMF oder WDF p_ξ mit marginalen WMFen oder WDFen $p_{\xi_i}, i = 1, ..., n$ besitzt, dann ist die Unabhängigkeit von $\xi_1, ..., \xi_n$ gleichbedeutend mit der Faktorisierung der gemeinsamen WMF oder WDF, also mit

$$p_\xi(\xi_1, ..., \xi_n) = \prod_{i=1}^n p_{\xi_i}(x_i). \tag{28}$$

Bemerkung

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

Unabhängige Zufallsvariablen

Definition (Unabhängig und identisch verteilte Zufallsvariablen)

n Zufallsvariablen $\xi_1, ..., \xi_n$ heißen *unabhängig und identisch verteilt (u.i.v.)*, wenn

(1) $\xi_1, ..., \xi_n$ unabhängige Zufallsvariablen sind, und

(2) die Marginalverteilungen der ξ_i übereinstimmen, also gilt, dass

$$\mathbb{P}_{\xi_i} = \mathbb{P}_{\xi_j} \text{ für alle } 1 \le i, j \le n. \tag{29}$$

Wenn die Zufallsvariablen $\xi_1, ..., \xi_n$ unabhängig und identisch verteilt sind und die $i$te Marginalverteilung $\mathbb{P}_\xi := \mathbb{P}_{\xi_i}$ ist, so schreibt man auch

$$\xi_1, ..., \xi_n \sim \mathbb{P}_\xi. \tag{30}$$

Bemerkungen

- Man sagt kurz, dass $\xi_1, ..., \xi_n$ 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 $\xi_1, ..., \xi_n \sim N(\mu, \sigma^2)$ geschrieben.

Definition

Multivariate Verteilungen

Marginalverteilungen

Bedingte Verteilungen

Unabhängige Zufallsvariablen

Selbstkontrollfragen

Selbstkontrollfragen

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

Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2022/23

Prof. Dr. Dirk Ostwald

(6) Erwartungswert, Varianz, Kovarianz

Modul B1 Deskriptive Statistik | Wahrscheinlichkeitstheorie und Frequentistische Inferenz

Datum	Einheit	Thema
13.10.2022	Einführung	(1) Einführung
20.10.2022	Wahrscheinlichkeitstheorie	(2) Wahrscheinlichkeitsräume
27.10.2022	Wahrscheinlichkeitstheorie	(3) Elementare Wahrscheinlichkeiten
03.11.2022	Wahrscheinlichkeitstheorie	(4) Zufallsvariablen I
10.11.2022	Wahrscheinlichkeitstheorie	(4) Zufallsvariablen II
17.11.2022	Wahrscheinlichkeitstheorie	(5) Multivariate Verteilungen
24.11.2022	Wahrscheinlichkeitstheorie	(6) Erwartungswert und Kovarianz
01.12.2022	Wahrscheinlichkeitstheorie	(7) Ungleichungen und Grenzwerte
08.12.2022	Wahrscheinlichkeitstheorie	(8) Transformationen der Normalverteilung
15.12.2022	Frequentistische Inferenz	(9) Statistische Modelle, Statistiken, Schätzer
	Weihnachtspause	
05.01.2023	Frequentistische Inferenz	(10) Schätzereigenschaften
12.01.2023	Frequentistische Inferenz	(11) Konfidenzintervalle
19.01.2023	Frequentistische Inferenz	(12) Hypothesentests
26.01.2023	Frequentistische Inferenz	(13) T-Tests
02.02.2023	Klausur	G44-H6, 16:00 - 17:00 Uhr
Jul 2023	Klausurwiederholungstermin	

Repräsentation zentraler Eigenschaften

Modellierung

Modell

Wahrscheinlichkeitstheorie

Probabilistisches Modell

$(\Omega, \mathcal{A}, \mathbb{P}), \xi : \Omega \to \mathbb{R}$

Wahrscheinlichkeitsrechnung

$\mathbb{P}_\xi(S) = \mathbb{P}(\xi^{-1}(S))$

Realität

Zufallsvorgänge

Phänomene, die von Menschen nicht mit absoluter Sicherheit vorhergesagt werden können.

Vorhersagen

Quantifizierung von Unsicherheit

Wir nehmen an, dass die BDI Score Fehler der Proband:innen Realisierungen **unabhängiger und identisch** normalverteilter Zufallsvariablen sind.

Modellierung

Modell

Wahrscheinlichkeitstheorie

$$y_{1j} = \mu_1 + \varepsilon_{1j}\,, \varepsilon_{1j} \sim N(0,\sigma^2), j = 1, \dots, n_1$$

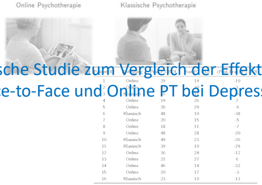
$$y_{2j} = \mu_2 + \varepsilon_{2j}\,, \varepsilon_{2j} \sim N(0,\sigma^2), j = 1, \dots, n_2$$

$$\mathbb{E}(\varepsilon_{ij}) = 0, \mathbb{V}(\varepsilon_{ij}) = \sigma^2 \;\; \forall\, i,j$$

$$\mathbb{C}(\varepsilon_{ij}, \varepsilon_{kl}) = 0 \;\; \forall\, i \neq k, j \neq l$$

Realität

Zufallsvorgang



Klinische Studie zum Vergleich der Effekte von Face-to-Face und Online PT bei Depression

Vorhersagen

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

© 2023 Dirk Ostwald CC BY-SA 4.0

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

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

$\square$

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

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

Mit der Definition des Erwartungswerts für kontinuierliche Zufallsvariablen gilt

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

Mit der Substitutionsregel

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

und der Definition von

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

gilt dann

Erwartungswert

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

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

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

verschwindet der Integralterm und wir erhalten

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

□

Theorem (Eigenschaften des Erwartungswerts)

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

$$\mathbb{E}(a\xi + b) = a\mathbb{E}(\xi) + b. \tag{9}$$

(2) (Linearkombination) Für Zufallsvariablen $\xi_1, ..., \xi_n$ und $a_1, ..., a_n \in \mathbb{R}$ gilt

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

(3) (Faktorisierung bei Unabhängigkeit) Für unabhängige Zufallsvariablen $\xi_1, ..., \xi_n$ gilt

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

Bemerkung

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

Erwartungswert

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 $\upsilon := a\xi + b$. Dann gilt

$$
\begin{aligned}
\mathbb{E}(\upsilon) &= \mathbb{E}(a\xi + b) \\
&= \int_{-\infty}^{\infty} (ax + b) p_\xi(x)\,dx \\
&= \int_{-\infty}^{\infty} axp_\xi(x) + bp_\xi(x)\,dx \\
&= a \int_{-\infty}^{\infty} xp_\xi(x)\,dx + b \int_{-\infty}^{\infty} p_\xi(x)\,dx \\
&= a\mathbb{E}(\xi) + b.
\end{aligned}
\tag{12}
$$

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

$$
\begin{aligned}
&\mathbb{E}\left(\sum_{i=1}^{2} a_i\xi_i\right) \\
&= \mathbb{E}(a_1\xi_1 + a_2\xi_2) \\
&= \iint_{\mathbb{R}^2} (a_1x_1 + a_2x_2)p_{\xi_1,\xi_2}(x_1, x_2)\,dx_1\,dx_2 \\
&= \iint_{\mathbb{R}^2} a_1x_1p_{\xi_1,\xi_2}(x_1, x_2) + a_2x_2p_{\xi_1,\xi_2}(x_1, x_2)\,dx_1\,dx_2 \\
&= a_1 \iint_{\mathbb{R}^2} x_1p_{\xi_1,\xi_2}(x_1, x_2)\,dx_1\,dx_2 + a_2 \iint_{\mathbb{R}^2} x_2p_{\xi_1,\xi_2}(x_1, x_2)\,dx_1\,dx_2
\end{aligned}
\tag{13}
$$

Beweis (fortgeführt)

$$
\begin{aligned}
&= a_1 \int_{-\infty}^{\infty} x_1 \left(\int_{-\infty}^{\infty} p_{\xi_1,\xi_2}(x_1,x_2)\,dx_2\right)\,dx_1 + a_2 \int_{-\infty}^{\infty} x_2 \left(\int_{-\infty}^{\infty} p_{\xi_1,\xi_2}(x_1,x_2)\,dx_1\right)\,dx_2 \\
&= a_1 \int_{-\infty}^{\infty} x_1 p_{\xi_1}(x_1)\,dx_1 + a_2 \int_{-\infty}^{\infty} x_2 p_{\xi_2}(x_2)\,dx_2 \\
&= a_1\mathbb{E}(\xi_1) + a_2\mathbb{E}(\xi_2) \\
&= \sum_{i=1}^{2} a_i\mathbb{E}(\xi_i).
\end{aligned}
\tag{14}
$$

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 $p_{\xi_1,\ldots,\xi_n}$. Weil als $\xi_1, \ldots, \xi_n$ unabhängig vorausgesetzt sind, gilt

$$p_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} p_{\xi_i}(x_i). \tag{15}$$

Weiterhin gilt also

$$\begin{aligned}
\mathbb{E}\left(\prod_{i=1}^{n} x_i\right) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\prod_{i=1}^{n} x_i\right) p_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n)\, dx_1 \ldots dx_n \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{n} x_i \prod_{i=1}^{n} p_{\xi_i}(x_i)\, dx_1 \ldots dx_n \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{n} x_i p_{\xi_i}(x_i)\, dx_1 \ldots dx_n \\
&= \prod_{i=1}^{n} \int_{-\infty}^{\infty} x_i p_{\xi_i}(x_i)\, dx_i = \prod_{i=1}^{n} \mathbb{E}(\xi_i).
\end{aligned} \tag{16}$$

$\square$

Erwartungswert

Varianz und Standardabweichung

Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung

Kovarianz und Korrelation

Stichprobenkovarianz und Stichprobenkorrelation

Selbstkontrollfragen

Definition (Varianz und Standardabweichung)

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

$$\mathbb{V}(\xi) := \mathbb{E}\left((\xi - \mathbb{E}(\xi))^2\right), \tag{17}$$

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

$$\mathbb{S}(\xi) := \sqrt{\mathbb{V}(\xi)}. \tag{18}$$

Bemerkungen

- Die Varianz misst die Streuung (Breite) einer Verteilung.
- Quadration 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 $-\mathbb{E}(\ln p_\xi)$.

Varianz und Standardabweichung

Beispiel (Varianz einer Bernoulli Zufallsvariable)

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

$$\mathbb{V}(\xi) = \mu(1-\mu). \tag{19}$$

<u>Beweis</u>

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

$$\begin{aligned}
\mathbb{V}(\xi) &= \mathbb{E}\left((\xi-\mu)^2\right) \\
&= \sum_{x\in\{0,1\}} (x-\mu)^2 \text{Bern}(x;\mu) \\
&= (0-\mu)^2\mu^0(1-\mu)^{1-0} + (1-\mu)^2\mu^1(1-\mu)^{1-1} \\
&= \mu^2(1-\mu) + (1-\mu)^2\mu \\
&= \left(\mu^2 + (1-\mu)\mu\right)(1-\mu) \\
&= \left(\mu^2 + \mu - \mu^2\right)(1-\mu) \\
&= \mu(1-\mu).
\end{aligned} \tag{20}$$

□

Theorem (Varianzverschiebungssatz)

Es sei ξ eine Zufallsvariable. Dann gilt

$$\mathbb{V}(\xi) = \mathbb{E}\left(\xi^2\right) - \mathbb{E}(\xi)^2. \tag{21}$$

Beweis

Mit der Definition der Varianz und der Linearität des Erwartungswerts gilt

$$\begin{aligned}
\mathbb{V}(\xi) &= \mathbb{E}\left((\xi - \mathbb{E}(\xi))^2\right) \\
&= \mathbb{E}\left(\xi^2 - 2\xi\mathbb{E}(\xi) + \mathbb{E}(\xi)^2\right) \\
&= \mathbb{E}(\xi^2) - 2\mathbb{E}(\xi)\mathbb{E}(\xi) + \mathbb{E}\left(\mathbb{E}(\xi)^2\right) \\
&= \mathbb{E}(\xi^2) - 2\mathbb{E}(\xi)^2 + \mathbb{E}(\xi)^2 \\
&= \mathbb{E}(\xi^2) - \mathbb{E}(\xi)^2.
\end{aligned} \tag{22}$$

$\square$

Bemerkung

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

Varianz und Standardabweichung

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

$$\mathbb{V}(\xi) = \mathbb{E}(\xi^2) - \mathbb{E}(\xi)^2 = \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} x^2 \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \, dx - \mu^2 \qquad (23)$$

Mit der Substitutionsregel

$$\int_a^b f(g(x))g'(x)\, dx = \int_{g(a)}^{g(b)} f(x)\, dx \qquad (24)$$

und der Definition von

$$g : \mathbb{R} \to \mathbb{R}, x \mapsto \sqrt{2\sigma^2}x + \mu, g(-\infty) := -\infty, g(\infty) := \infty, \text{ with } g'(x) = \sqrt{2\sigma^2}, \qquad (25)$$

kann das Integral auf der rechten Seite von Gleichung (23) dann als

Varianz und Standardabweichung

Beweis (fortgeführt)

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

geschrieben werden. Also gilt

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

Beweis (fortgeführt)

Wir halten weiterhin ohne Beweis fest, dass

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

Es ergibt sich dann

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

Mit der partiellen Integrationsregel

$$\int_a^b f'(x)g(x)\, dx = f(x)g(x)|_a^b - \int_a^b f(x)g'(x)\, dx \tag{29}$$

und der Definition von

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

Varianz und Standardabweichung

Beweis (fortgeführt)

und

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

so dass

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

gilt, ergibt sich dann

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

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

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

□

Theorem (Eigenschaften der Varianz)

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

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

(2) (Linearkombination bei Unabhängigkeit) Für unabhängige Zufallsvariablen $\xi_1, ..., \xi_n$ und $a_1, ..., a_n \in \mathbb{R}$ gilt

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

Beweis

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

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

Wurzelziehen ergibt dann das Resultat für die Standardabweichung.

Für Eigenschaft (2) betrachten wir den Fall zweier unabhängiger Zufallsvariablen ξ_1 und ξ_2 genauer. Wir halten zunächst fest, dass in diesem Fall gilt, dass

$$\mathbb{E}\left(a_1\xi_1 + a_2\xi_2\right) = a_1\mathbb{E}(\xi_1) + a_2\mathbb{E}(\xi_2). \tag{38}$$

Varianz und Standardabweichung

Beweis (fortgeführt)

Es ergibt sich also

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

Varianz und Standardabweichung

Beweis (fortgeführt)

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

$$
\begin{aligned}
\mathbb{E}\left((\xi_1 - \mathbb{E}(\xi_1))(\xi_2 - \mathbb{E}(\xi_2))\right) &= \mathbb{E}\left((\xi_1 - \mathbb{E}(\xi_1))\right)\mathbb{E}\left((\xi_2 - \mathbb{E}(\xi_2))\right) \\
&= (\mathbb{E}(\xi_1) - \mathbb{E}(\xi_1))(\mathbb{E}(\xi_2) - \mathbb{E}(\xi_2)) \\
&= 0
\end{aligned} \tag{39}
$$

ist. Damit folgt also

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

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

$\square$

Erwartungswert

Varianz und Standardabweichung

Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung

Kovarianz und Korrelation

Selbstkontrollfragen

Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung

Definition (Stichprobenmittel, -varianz, -standardabweichung)

$\xi_1, ..., \xi_n$ seien Zufallsvariablen. Dann nennt man $\xi_1, ..., \xi_n$ auch eine *Stichprobe*.

- Das *Stichprobenmittel* von $\xi_1, ..., \xi_n$ ist definiert als der arithmetische Mittelwert

$$\bar{\xi}_n := \frac{1}{n}\sum_{i=1}^n \xi_i. \tag{41}$$

- Die *Stichprobenvarianz* von $\xi_1, ..., \xi_n$ ist definiert als

$$S_n^2 := \frac{1}{n-1}\sum_{i=1}^n (\xi_i - \bar{\xi}_n)^2. \tag{42}$$

- Die *Stichprobenstandardabweichung* ist definiert als

$$S_n := \sqrt{S_n^2}. \tag{43}$$

Bemerkungen

- $\mathbb{E}(\xi), \mathbb{V}(\xi)$, und $\mathbb{S}(\xi)$ sind Kennzahlen einer Zufallsvariable ξ.
- $\bar{\xi}_n, S_n^2$, und S_n sind Kennzahlen einer Stichprobe $\xi_1, ..., \xi_n$.
- $\bar{\xi}_n, S_n^2$, und S_n sind Zufallsvariablen, ihre Realisationen werden mit $\bar{x}_n, s_n^2$, und s_n bezeichnet.

Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung

Beispiel (Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung)

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

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

- Die Stichprobenvarianzrealisation ist

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

- Die Stichprobenstandardabweichungrealisation ist

$$s_{10} = \sqrt{s_{10}^2} = \sqrt{2.82} = 1.68. \tag{46}$$

Erwartungswert

Varianz und Standardabweichung

Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung

Kovarianz und Korrelation

Selbstkontrollfragen

Definition (Kovarianz und Korrelation)

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

$$\mathbb{C}(\xi, \upsilon) := \mathbb{E}\left((\xi - \mathbb{E}(\xi))\left(\upsilon - \mathbb{E}(\upsilon)\right)\right). \tag{47}$$

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

$$\rho(\xi, \upsilon) := \frac{\mathbb{C}(\xi, \upsilon)}{\sqrt{\mathbb{V}(\xi)}\sqrt{\mathbb{V}(\upsilon)}} = \frac{\mathbb{C}(\xi, \upsilon)}{\mathbb{S}(\xi)\mathbb{S}(\upsilon)}. \tag{48}$$

Bemerkungen

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

$$\mathbb{C}(\xi, \xi) = \mathbb{E}\left((\xi - \mathbb{E}(\xi))^2\right) = \mathbb{V}(\xi). \tag{49}$$

- $\rho(\xi, \upsilon)$ wird auch *Korrelationskoeffizient* von ξ und υ genannt.
- Wenn $\rho(\xi, \upsilon) = 0$ ist, werden ξ und υ *unkorreliert* genannt.
- Wir zeigen später mit der Cauchy-Schwarz Ungleichung, dass $-1 \le \rho(\xi, \upsilon) \le 1$.

Beispiel (Kovarianz und Korrelation zweier diskreter Zufallsvariablen)

Es sei $\xi := (\xi_1, \xi_2)$ ein Zufallsvektor mit WMF p_{ξ_1,ξ_2} definiert durch

$p_{\xi_1,\xi_2}(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$p_{\xi_1}(x_1)$
$x_1 = 1$	0.10	0.05	0.15	0.30
$x_1 = 2$	0.60	0.05	0.05	0.70
$p_{\xi_2}(x_2)$	0.70	0.10	0.20	

ξ_1, ξ_2 sind also zwei Zufallsvariablen mit einer definierten bivariaten Verteilung. Um $\mathbb{C}(\xi_1, \xi_2)$ und $\rho(\xi_1, \xi_2)$ zu berechnen, halten wir zunächst fest, dass

$$\mathbb{E}(\xi_1) = \sum_{x_1=1}^{2} x_1 p_{\xi_1}(x_1) = 1 \cdot 0.3 + 2 \cdot 0.7 = 1.7 \tag{50}$$

und

$$\mathbb{E}(\xi_2) = \sum_{x_2=1}^{3} x_2 p_{\xi_2}(x_2) = 1 \cdot 0.7 + 2 \cdot 0.1 + 3 \cdot 0.2 = 1.5. \tag{51}$$

Mit der Definition der Kovarianz von ξ_1 und ξ_2, gilt dann

$$
\begin{aligned}
&\mathbb{C}(\xi_1, \xi_2) \\
&= \mathbb{E}((\xi_1 - \mathbb{E}(\xi_1))(\xi_2 - \mathbb{E}(\xi_2))) \\
&= \sum_{x_1=1}^{2} \sum_{x_2=1}^{3} (x_1 - \mathbb{E}(\xi_1))(x_2 - \mathbb{E}(\xi_2)) p_{\xi_1,\xi_2}(x_1, x_2) \\
&= \sum_{x_1=1}^{2} \sum_{x_2=1}^{3} (x_1 - 1.7)(x_2 - 1.5) p_{\xi_1,\xi_2}(x_1, x_2) \\
&= \sum_{x_1=1}^{2} (x_1 - 1.7)(1 - 1.5) p_{\xi_1,\xi_2}(x_1, 1) \\
&\qquad + (x_1 - 1.7)(2 - 1.5) p_{\xi_1,\xi_2}(x_1, 2) \\
&\qquad + (x_1 - 1.7)(3 - 1.5) p_{\xi_1,\xi_2}(x_1, 3) \\
&= \quad (1 - 1.7)(1 - 1.5) p_{\xi_1,\xi_2}(1, 1) + (1 - 1.7)(2 - 1.5) p_{\xi_1,\xi_2}(1, 2) + (1 - 1.7)(3 - 1.5) p_{\xi_1,\xi_2}(1, 3) \\
&\quad + (2 - 1.7)(1 - 1.5) p_{\xi_1,\xi_2}(2, 1) + (2 - 1.7)(2 - 1.5) p_{\xi_1,\xi_2}(2, 2) + (2 - 1.7)(3 - 1.5) p_{\xi_1,\xi_2}(2, 3) \\
&= \quad (-0.7) \cdot (-0.5) \cdot 0.10 \quad + (-0.7) \cdot 0.5 \cdot 0.05 \quad + (-0.7) \cdot 1.5 \cdot 0.15 \\
&\quad + 0.3 \cdot (-0.5) \cdot 0.60 \quad + 0.3 \cdot 0.5 \cdot 0.05 \quad + 0.3 \cdot 1.5 \cdot 0.05 \\
&= 0.035 - 0.0175 - 0.1575 - 0.09 + 0.0075 + 0.0225 \\
&= -0.2.
\end{aligned}
$$

Theorem (Kovarianzverschiebungssatz)

ξ und υ seien Zufallsvariablen. Dann gilt

$$\mathbb{C}(\xi, \upsilon) = \mathbb{E}(\xi\upsilon) - \mathbb{E}(\xi)\mathbb{E}(\upsilon). \tag{52}$$

Beweis

Mit der Definition der Kovarianz gilt

$$\begin{aligned}
\mathbb{C}(\xi, \upsilon) &= \mathbb{E}\left((\xi - \mathbb{E}(\xi))(\upsilon - \mathbb{E}(\upsilon))\right) \\
&= \mathbb{E}\left(\xi\upsilon - \xi\mathbb{E}(\upsilon) - \mathbb{E}(\xi)\upsilon + \mathbb{E}(\xi)\mathbb{E}(\upsilon)\right) \\
&= \mathbb{E}(\xi\upsilon) - \mathbb{E}(\xi)\mathbb{E}(\upsilon) - \mathbb{E}(\xi)\mathbb{E}(\upsilon) + \mathbb{E}(\xi)\mathbb{E}(\upsilon) \\
&= \mathbb{E}(\xi\upsilon) - \mathbb{E}(\xi)\mathbb{E}(\upsilon).
\end{aligned} \tag{53}$$

$\square$

Bemerkungen

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

Theorem (Varianzen von Summen und Differenzen von Zufallsvariablen)

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

$$\mathbb{V}(a\xi + b\upsilon + c) = a^2\mathbb{V}(\xi) + b^2\mathbb{V}(\upsilon) + 2ab\mathbb{C}(\xi, \upsilon). \tag{54}$$

Speziell gelten

$$\mathbb{V}(\xi + \upsilon) = \mathbb{V}(\xi) + \mathbb{V}(\upsilon) + 2\mathbb{C}(\xi, \upsilon) \tag{55}$$

und

$$\mathbb{V}(\xi - \upsilon) = \mathbb{V}(\xi) + \mathbb{V}(\upsilon) - 2\mathbb{C}(\xi, \upsilon) \tag{56}$$

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

$$\mathbb{E}(a\xi + b\upsilon + c) = a\mathbb{E}(\xi) + b\mathbb{E}(\upsilon) + c. \tag{57}$$

Es ergibt sich also

$$\begin{aligned}
&\mathbb{V}(a\xi + b\upsilon + c) \\
&= \mathbb{E}\left((a\xi + b\upsilon + c - a\mathbb{E}(\xi) - b\mathbb{E}(\upsilon) - c)^2\right) \\
&= \mathbb{E}\left((a(\xi - \mathbb{E}(\xi)) + b(\upsilon - \mathbb{E}(\upsilon)))^2\right) \\
&= \mathbb{E}\left(a^2(\xi - \mathbb{E}(\xi))^2 + 2ab(\xi - \mathbb{E}(\xi))(\upsilon - \mathbb{E}(\upsilon))) + b^2(\upsilon - \mathbb{E}(\upsilon))^2\right) \\
&= a^2\mathbb{E}\left((\xi - \mathbb{E}(\xi))^2\right) + b^2\mathbb{E}\left((\upsilon - \mathbb{E}(\upsilon))^2\right) + 2ab\mathbb{E}\left((\xi - \mathbb{E}(\xi))(\upsilon - \mathbb{E}(\upsilon))\right)) \\
&= a^2\mathbb{V}(\xi) + b^2\mathbb{V}(\upsilon) + 2ab\mathbb{C}(\xi, \upsilon)
\end{aligned} \tag{58}$$

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

$\square$

Theorem (Korrelation und Unabhängigkeit)

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

Beweis

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

$$\mathbb{E}(\xi\upsilon) = \mathbb{E}(\xi)\mathbb{E}(\upsilon). \tag{59}$$

Mit dem Kovarianzverschiebungssatz folgt dann

$$\mathbb{C}(\xi, \upsilon) = \mathbb{E}(\xi\upsilon) - \mathbb{E}(\xi)\mathbb{E}(\upsilon) = \mathbb{E}(\xi)\mathbb{E}(\upsilon) - \mathbb{E}(\xi)\mathbb{E}(\upsilon) = 0. \tag{60}$$

Mit der Definition des Korrelationskoeffizienten folgt dann unmittelbar, dass $\rho(\xi, \upsilon) = 0$ und ξ und υ somit unkorreliert sind.

Beweis (fortgeführt)

Wir zeigen nun durch Angabe eines Beispiels, dass die Kovarianz von abhängigen Zufallsvariablen ξ und υ null sein kann.

Zu diesem Zweck betrachten wir den Fall zweier diskreter Zufallsvariablen ξ und υ mit Ergebnisräumen $\mathcal{X} = \{-1, 0, 1\}$ und $\upsilon = \{0, 1\}$, marginaler WMF von ξ gegeben durch $p_\xi(x) := 1/3$ für $x \in \mathcal{X}$ und der Definition $\upsilon := \xi^2$.

Wir halten dann zunächst fest, dass

$$\mathbb{E}(\xi) = \sum_{x \in \mathcal{X}} x p_\xi(x) = -1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = 0 \tag{61}$$

und

$$\mathbb{E}(\xi\upsilon) = \mathbb{E}(\xi\xi^2) = \mathbb{E}(\xi^3) = \sum_{x \in \mathcal{X}} x^3 p_\xi(x) = -1^3 \cdot \frac{1}{3} + 0^3 \cdot \frac{1}{3} + 1^3 \cdot \frac{1}{3} = 0. \tag{62}$$

Mit dem Kovarianzverschiebungssatz ergibt sich dann

$$\mathbb{C}(\xi, \upsilon) = \mathbb{E}(\xi\upsilon) - \mathbb{E}(\xi)\mathbb{E}(\upsilon) = \mathbb{E}(\xi^3) - \mathbb{E}(\xi)\mathbb{E}(\upsilon) = 0 - 0 \cdot \mathbb{E}(\upsilon) = 0. \tag{63}$$

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

© 2023 Dirk Ostwald CC BY-SA 4.0

Beweis (fortgeführt)

Die Definition of $\upsilon := \xi^2$ impliziert die folgende bedingte WMF

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

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

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

$$p_{\xi,\upsilon}(-1,0) = 0 \neq \frac{1}{9} = \frac{1}{3}\cdot\frac{1}{3} = p_\xi(-1)p_\upsilon(0) \tag{64}$$

und damit sind ξ und υ nicht unabhängig.

$\square$

Erwartungswert

Varianz und Standardabweichung

Stichprobenmittel, Stichprobenvarianz, Stichprobenstandardabweichung

Kovarianz und Korrelation

Selbstkontrollfragen

Selbstkontrollfragen

1. Definieren und interpretieren Sie den Erwartungswert einer Zufallsvariable.
2. Berechnen Sie den Erwartungswert einer Bernoulli Zufallsvariable.
3. Nennen Sie drei Eigenschaften des Erwartungswerts.
4. Definieren und interpretieren Sie die Varianz einer Zufallsvariable.
5. Berechnen Sie die Varianz einer Bernoulli Zufallsvariable.
6. Drücken Sie $\mathbb{E}(\xi^2)$ mithilfe der Varianz und des Erwartungswerts von ξ aus.
7. Was ist $\mathbb{V}(a\xi)$ für konstantes $a \in \mathbb{R}$?
8. Definieren Sie die Kovarianz und Korrelation zweier Zufallsvariablen ξ und υ.
9. Geben Sie das Theorem zur Varianz von Linearkombinationen von Zufallsvariablen bei Unabhängigkeit wieder.
10. Definieren Sie den Begriff der Stichprobe.
11. Definieren Sie den Begriff des Stichprobenmittels.
12. Definieren Sie Stichprobenvarianz und Stichprobenstandardabweichung.
13. Erläutern Sie die Unterschiede zwischen dem Erwartungswertparameter, dem Erwartungswert und dem Stichprobenmittel von normalverteilten Zufallsvariablen.
14. Definieren Sie die Kovarianz und die Korrelation zweier Zufallsvariablen.
15. Schreiben Sie die Kovarianz zweier Zufallsvariablen mithilfe von Erwartungswerten.
16. Geben Sie das Theorem zur Korrelation und Unabhängigkeit zweier Zufallsvariablen wieder.
17. Was ist die Varianz der Summe zweier Zufallsvariablen bei Unabhängigkeit?
18. Was ist die Varianz der Summe zweier Zufallsvariablen im Allgemeinen?

Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2022/23

Prof. Dr. Dirk Ostwald

(7) 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

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

Bemerkungen

- Weil $\mathbb{P}(\xi \geq 0) = 1$ gilt, sagt man auch, dass ξ eine *nicht-negative* Zufallvariable 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

$$\mathbb{E}(\xi) = \int_{-\infty}^{\infty} s\, p(s)\, ds = \int_{0}^{\infty} s\, p(s)\, ds = \int_{0}^{x} s\, p(s)\, ds + \int_{x}^{\infty} s\, p(s)\, ds, \tag{2}$$

weil ξ nicht-negativ ist. Es folgt dann

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

Dabei gilt die erste Ungleichung weil

$$\int_{0}^{x} s\, p(s)\, ds \geq 0 \tag{4}$$

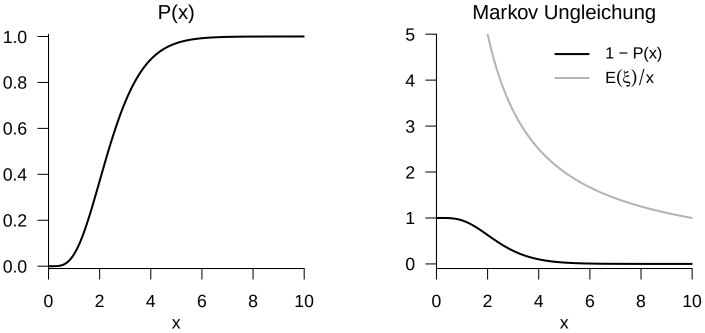
und die zweite Ungleichung gilt, weil $x \leq \xi$ für $\xi \in [x, \infty[$. Es folgt also, dass

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

$\square$

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 $x \in \mathbb{R}$

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

Bemerkungen

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

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

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

$$a^2 \geq b^2 \Rightarrow \sqrt{a^2} \geq \sqrt{b^2} \Rightarrow a \geq b \Rightarrow |a| \geq b. \tag{8}$$

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

$$a^2 \geq b^2 \Rightarrow \sqrt{a^2} \geq \sqrt{b^2} \Rightarrow -a \geq b \Rightarrow |a| \geq b. \tag{9}$$

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

$$a^2 \geq b^2 \Rightarrow \sqrt{a^2} \geq \sqrt{b^2} \Rightarrow a \geq -b \geq b \Rightarrow |a| \geq b. \tag{10}$$

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

$$a^2 \geq b^2 \Rightarrow \sqrt{a^2} \geq \sqrt{b^2} \Rightarrow -a \geq -b \geq b \Rightarrow |a| \geq b. \tag{11}$$

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

$$\mathbb{P}\left(\upsilon \geq x^2\right) \leq \frac{\mathbb{E}(\upsilon)}{x^2} \Leftrightarrow \mathbb{P}\left((\xi - \mathbb{E}(\xi))^2 \geq x^2\right) \leq \frac{\mathbb{E}\left((\xi - \mathbb{E}(\xi))^2\right)}{x^2} \Leftrightarrow \mathbb{P}(|\xi - \mathbb{E}(\xi)| \geq x) \leq \frac{\mathbb{V}(\xi)}{x^2}. \tag{12}$$

$\square$

Wahrscheinlichkeitsungleichungen

Erwartungswertungleichungen

Gesetze der Großen Zahl

Zentrale Grenzwertsätze

Selbstkontrollfragen

Theorem (Cauchy-Schwarz-Ungleichung)

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

$$\mathbb{E}(\xi\upsilon)^2 \le \mathbb{E}\left(\xi^2\right)\mathbb{E}\left(\upsilon^2\right). \tag{13}$$

Bemerkungen

- Analog gilt für Vektoren $x, y \in \mathbb{R}^n$, dass $\langle x, y\rangle^2 \le \|x\| \cdot \|y\|$.
- 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 υ seien Zufallsvariablen mit $\mathbb{V}(\xi), \mathbb{V}(\upsilon) > 0$. Dann gelten

$$\frac{\mathbb{C}(\xi, \upsilon)^2}{\mathbb{V}(\xi)\mathbb{V}(\upsilon)} \le 1 \text{ und } -1 \le \rho(\xi, \upsilon) \le 1. \tag{14}$$

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

$$\mathbb{E}(\alpha\beta)^2 \le \mathbb{E}\left(\alpha^2\right)\mathbb{E}\left(\beta^2\right). \tag{15}$$

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

$$\mathbb{E}\left((\xi - \mathbb{E}(\xi))(\upsilon - \mathbb{E}(\upsilon))\right)^2 \le \mathbb{E}\left((\xi - \mathbb{E}(\xi))^2\right)\mathbb{E}\left((\upsilon - \mathbb{E}(\upsilon))^2\right). \tag{16}$$

Also gilt

$$\mathbb{C}(\xi,\upsilon)^2 \le \mathbb{V}(\xi)\mathbb{V}(\upsilon) \Leftrightarrow \frac{\mathbb{C}(\xi,\upsilon)^2}{\mathbb{V}(\xi)\mathbb{V}(\upsilon)} \le 1. \tag{17}$$

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

$$\rho(\xi,\upsilon)^2 \le 1. \tag{18}$$

Dann gilt aber auch

$$|\rho(\xi,\upsilon)^2| \le 1 \Leftrightarrow -1 \le \rho(\xi,\upsilon) \le 1, \tag{19}$$

denn

$$\rho(\xi,\upsilon)^2 \le 1 \Rightarrow \sqrt{\rho(\xi,\upsilon)^2} \le \sqrt{1} \Rightarrow \quad \rho(\xi,\upsilon) \le 1 \Rightarrow |\rho(\xi,\upsilon)| \le 1 \text{ für } \rho(\xi,\upsilon) \ge 0 \tag{20}$$

und

$$\rho(\xi,\upsilon)^2 \le 1 \Rightarrow \sqrt{\rho(\xi,\upsilon)^2} \le \sqrt{1} \Rightarrow -\rho(\xi,\upsilon) \le 1 \Rightarrow |\rho(\xi,\upsilon)| \le 1 \text{ für } \rho(\xi,\upsilon) \le 0 \tag{21}$$

$\square$

Erwartungswertungleichungen

Theorem (Jensensche Ungleichung)

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

$$g(\lambda x_1 + (1-\lambda)x_2) \le \lambda g(x_1) + (1-\lambda)g(x_2) \tag{22}$$

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

$$\mathbb{E}(g(\xi)) \ge g(\mathbb{E}(\xi)). \tag{23}$$

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

$$g(\lambda x_1 + (1-\lambda)x_2) \ge \lambda g(x_1) + (1-\lambda)g(x_2) \tag{24}$$

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

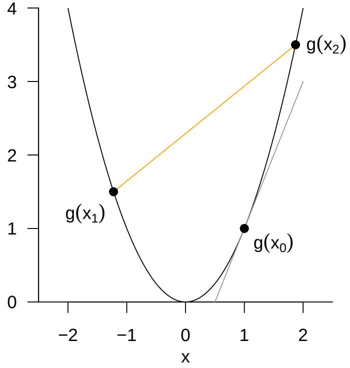
$$\mathbb{E}(g(\xi)) \le g(\mathbb{E}(\xi)). \tag{25}$$

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 $\mathbb{E}(\ln \xi) \le \ln \mathbb{E}(\xi)$.

Visualisierung einer konvexen Funktion

— $g(x) := x^2$



— $\lambda g(x_1) + (1-\lambda)g(x_2)$ für $x_1 := -\sqrt{1.5}, x_2 := \sqrt{3.5}, \lambda \in [0,1]$

— $t(x) := g(x_0) + g'(x_0)(x - x_0)$ für $x_0 := 1$

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

$$g(x) \geq t(x) := g(x_0) + g'(x_0)(x - x_0) \tag{26}$$

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

$$g(\xi) \geq g(\mathbb{E}(\xi)) + g'(\mathbb{E}(\xi))(\xi - \mathbb{E}(\xi)) \tag{27}$$

Erwartungswertbildung ergibt dann

$$\begin{aligned}
& \mathbb{E}(g(\xi)) \geq \mathbb{E}(g(\mathbb{E}(\xi))) + \mathbb{E}(g'(\mathbb{E}(\xi))(\xi - \mathbb{E}(\xi))) \\
\Leftrightarrow\ & \mathbb{E}(g(\xi)) \geq g(\mathbb{E}(\xi)) + g'(\mathbb{E}(\xi))\mathbb{E}((\xi - \mathbb{E}(\xi))) \\
\Leftrightarrow\ & \mathbb{E}(g(\xi)) \geq g(\mathbb{E}(\xi)) + g'(\mathbb{E}(\xi))(\mathbb{E}(\xi) - \mathbb{E}(\xi)) \\
\Leftrightarrow\ & \mathbb{E}(g(\xi)) \geq g(\mathbb{E}(\xi)).
\end{aligned} \tag{28}$$

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

$$\begin{aligned}
& \mathbb{E}(-g(\xi)) \geq -g(\mathbb{E}(\xi)) \\
\Leftrightarrow\ & -\mathbb{E}(g(\xi)) \geq -g(\mathbb{E}(\xi)) \\
\Leftrightarrow\ & \mathbb{E}(g(\xi)) \leq g(\mathbb{E}(\xi)).
\end{aligned} \tag{29}$$

$\square$

Wahrscheinlichkeitsungleichungen

Erwartungswertungleichungen

Gesetze der Großen Zahl

Zentrale Grenzwertsätze

Selbstkontrollfragen

Überblick

- Es gibt ein *Schwaches Gesetz der Großen Zahl* und ein *Starkes Gesetz der Großen Zahl*.
- Intuitiv besagen beide Gesetze, dass sich das Stichprobenmittel von unabhängigen und identisch verteilten Zufallsvariablen für eine große Anzahl an Zufallsvariablen dem Erwartungswert der zugrundeliegenden Verteilung nähert.
- Das Schwache und das Starke Gesetz der Großen Zahl unterscheiden sich in Hinblick auf die zu ihrer Formulierung benutzen Formen der *Konvergenz von Zufallsvariablen*.
  - Das Schwache Gesetz basiert auf der *Konvergenz in Wahrscheinlichkeit*.
  - Das Starke Gesetz basiert auf der *fast sicheren Konvergenz*.
- Wir begnügen uns mit Konvergenz in Wahrscheinlichkeit und dem Schwachen Gesetz.

Definition (Konvergenz in Wahrscheinlichkeit)

Eine Folge von Zufallsvariable $\xi_1, \xi_2, \ldots$ *konvergiert gegen eine Zufallsvariable* ξ *in Wahrscheinlichkeit*, wenn für jedes noch so kleine $\epsilon > 0$ gilt, dass

$$\lim_{n \to \infty} \mathbb{P}(|\xi_n - \xi| < \epsilon) = 1 \Leftrightarrow \lim_{n \to \infty} \mathbb{P}(|\xi_n - \xi| \geq \epsilon) = 0 \tag{30}$$

Die Konvergenz von $\xi_1, \xi_2, ...$ gegen ξ in Wahrscheinlichkeit wird geschrieben als

$$\xi_n \xrightarrow[n \to \infty]{\mathrm{P}} \xi \tag{31}$$

Bemerkungen

- $\xi_n \xrightarrow[n \to \infty]{\mathrm{P}} \xi$ heißt, dass sich die Wahrscheinlichkeit, dass ξ_n in dem zufälligen Intervall

$$]\xi - \epsilon, \xi + \epsilon[\tag{32}$$

  liegt, unabhängig davon, wie klein dieses Intervall sein mag, 1 nähert, wenn n gegen Unendlich strebt.
- Intuitiv heißt das, dass sich für eine konstante Zufallsvariable $\xi := a$ die Verteilung von ξ_n mehr und mehr um a konzentriert, wenn n gegen Unendlich strebt.

Theorem (Schwaches Gesetz der Großen Zahl)

$\xi_1, ..., \xi_n$ seien unabhängig und gleichverteilte Zufallsvariablen mit $\mathbb{E}(\xi_i) = \mu$ für alle $i = 1, ..., n$. Weiterhin bezeichne

$$\bar{\xi}_n := \frac{1}{n}\sum_{i=1}^{n} \xi_i \tag{33}$$

das Stichprobenmittel der $\xi_i, i = 1, ..., n$. Dann konvergiert $\bar{\xi}_n$ in Wahrscheinlichkeit gegen μ,

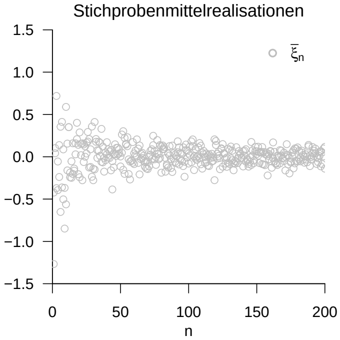
$$\bar{\xi}_n \xrightarrow[n\to\infty]{\mathrm{P}} \mu. \tag{34}$$

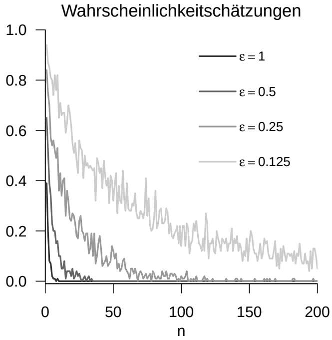
Bemerkungen

- Für einen Beweis siehe zum Beispiel Georgii (2009), Abschnitt 5.1.
- $\bar{\xi}_n \xrightarrow[n\to\infty]{\mathrm{P}} \mu$ heißt, dass die Wahrscheinlichkeit, dass das Stichprobenmittel nahe dem Erwartungswert der zugrundeliegenden Verteilung liegt, sich 1 nähert, wenn $n \to \infty$.

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

- 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.





Wahrscheinlichkeitsungleichungen

Erwartungswertungleichungen

Gesetze der Großen Zahl

Zentrale Grenzwertsätze

Selbstkontrollfragen

Zentrale Grenzwertsätze

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 eine Messgröße y also als Summe eines deterministischen Einflusses μ und der Summe

$$\varepsilon := \sum_{i=1}^{n} \xi_i \tag{35}$$

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

$$y = \mu + \varepsilon \text{ mit } \varepsilon \sim N(0, \sigma^2) \tag{36}$$

  also mathematisch gerechtfertigt. Wie wir später sehen werden, liegt die Annahme in Gleichung (36) vielen statischen Modellen zugrunde.
- In der "Lindenberg und Lévy" Form des Zentralen Grenzwertsatzes werden unabhängig und identische Zufallsvariablen vorausgesetzt. In der "Liapunov" Form werden nur unabhängige Zufallsvariablen voraussetzt. Der Beweis der "Lindenberg und Lévy" Form ist einfacher als der Beweis der "Liapunov" Form. Wir verzichten hier aber auf die Angabe von Beweisen.
- In beiden Formulierungen des Zentralen Grenzwertsatzes ist die betrachtete Konvergenz von Zufallsvariablen die *Konvergenz in Verteilung*, welche wir zunächst einführen.
- Zur mathematik-geschichtlichen Genese der Zentralen Grenzwertsätze siehe z.B. Fischer (2011).

Definition (Konvergenz in Verteilung)

Eine Folge $\xi_1, \xi_2, \ldots$ von Zufallsvariablen *konvergiert in Verteilung gegen eine Zufallsvariable* ξ, wenn

$$\lim_{n \to \infty} P_{\xi_n}(x) = P_\xi(x). \tag{37}$$

für alle ξ an denen P_ξ stetig ist. Die Konvergenz in Verteilung von $\xi_1, \xi_2, \ldots$ gegen ξ wird geschrieben als

$$\xi_n \xrightarrow[n \to \infty]{\text{D}} \xi, \tag{38}$$

Gilt $\xi_n \xrightarrow[n \to \infty]{\text{D}} \xi$, dann heißt die Verteilung von ξ die *asymptotische Verteilung der Folge* $\xi_1, \xi_2, \ldots$.

Bemerkungen

- $\xi \xrightarrow[n \to \infty]{\text{D}} \xi$ ist eine Aussage über die Konvergenz von KVFs.
- Konvergenz in Wahrscheinlichkeit impliziert Konvergenz in Verteilung.

Theorem (Zentraler Grenzwertsatz nach Lindenberg und Lévy)

$\xi_1, ..., \xi_n$ seien unabhängig und identisch verteilte Zufallsvariablen mit

$$\mathbb{E}(\xi_i) := \mu \text{ und } \mathbb{V}(\xi_i) := \sigma^2 > 0 \text{ für alle } i = 1, ..., n. \tag{39}$$

Weiterhin sei ζ_n die Zufallsvariable definiert als

$$\zeta_n := \sqrt{n}\left(\frac{\bar{\xi}_n - \mu}{\sigma}\right). \tag{40}$$

Dann gilt für alle $z \in \mathbb{R}$

$$\lim_{n\to\infty} P_{\zeta_n}(z) = \Phi(z), \tag{41}$$

wobei Φ die kumulative Verteilungsfunktion der Standardnormalverteilung bezeichnet.

Bemerkung

- Wir zeigen später, dass damit für $n \to \infty$ asymptotisch auch gilt, dass

$$\sum_{i=1}^n \xi_i \sim N(n\mu, n\sigma^2) \text{ und } \bar{\xi}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right). \tag{42}$$

© 2023 Dirk Ostwald CC BY-SA 4.0

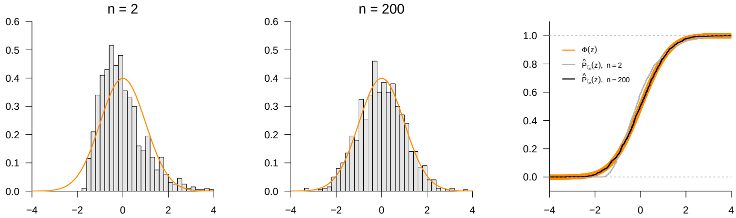
Beispiel $(\xi_1, ..., \xi_n \sim \chi^2(k))$

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

$$\zeta_n := \sqrt{n}\left(\frac{\bar{\xi}_n - \mu}{\sigma}\right) \tag{43}$$

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)

$\xi_1, ..., \xi_n$ seien unabhängige aber nicht notwendigerweise identisch verteilten Zufallsvariablen mit

$$\mathbb{E}(\xi_i) := \mu_i \text{ und } \mathbb{V}(\xi_i) := \sigma_i^2 > 0 \text{ für alle } i = 1, ..., n. \tag{44}$$

Weiterhin sollen für $\xi_1, ..., \xi_n$ folgende Eigenschaften gelten:

$$\mathbb{E}(|\xi_i - \mu_i|^3) < \infty \text{ und } \lim_{n\to\infty} \frac{\sum_{i=1}^n \mathbb{E}\left(|\xi_i - \mu_i|^3\right)}{(\sum_{i=1}^n \sigma_i^2)^{3/2}} = 0. \tag{45}$$

Dann gilt für die Zufallsvariable ζ_n definiert als

$$\zeta_n := \frac{\sum_{i=1}^n \xi_i - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}}, \tag{46}$$

für alle $z \in \mathbb{R}$, dass

$$\lim_{n\to\infty} P_{\zeta_n}(z) = \Phi(z), \tag{47}$$

wobei Φ KVF der Standardnormalverteilung bezeichnet.

Bemerkungen

- 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. Definieren Sie den Begriff der Konvergenz in Wahrscheinlichkeit.
6. Definieren Sie den Begriff der Konvergenz in Verteilung.
7. Geben Sie das Schwache Gesetz der Großen Zahl wieder.
8. Erläutern Sie den Zentralen Grenzwertsatz nach Lindenberg und Lévy.
9. Erläutern Sie den Zentralen Grenzwertsatz nach Liapunov.
10. Warum sind die Zentralen Grenzwertsätze für die statistische Modellbildung wichtig?

References

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.

Georgii, Hans-Otto. 2009. *Stochastik: Einführung in Die Wahrscheinlichkeitstheorie Und Statistik*. 4., überarb. und erw. Aufl. De-Gruyter-Lehrbuch. Berlin: de Gruyter.

Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2022/23

Prof. Dr. Dirk Ostwald

(8) Transformationen der Normalverteilung

Vorbemerkungen

Transformationstheoreme

Standardtransformationen

- Summentransformation
- Mittelwerttransformation
- Z-Transformation
- χ^2-Transformation
- T-Transformation
- F-Transformation

Selbstkontrollfragen

Vorbemerkungen

Transformationstheoreme

Standardtransformationen

- Summentransformation
- Mittelwerttransformation
- Z-Transformation
- χ^2-Transformation
- T-Transformation
- F-Transformation

Selbstkontrollfragen

Realisierungen von Zufallsvariablen

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

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

```
[1] -1.4
```

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

```
[1] 10.3
```

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

```
[1] 1.55 4.99 5.88
```

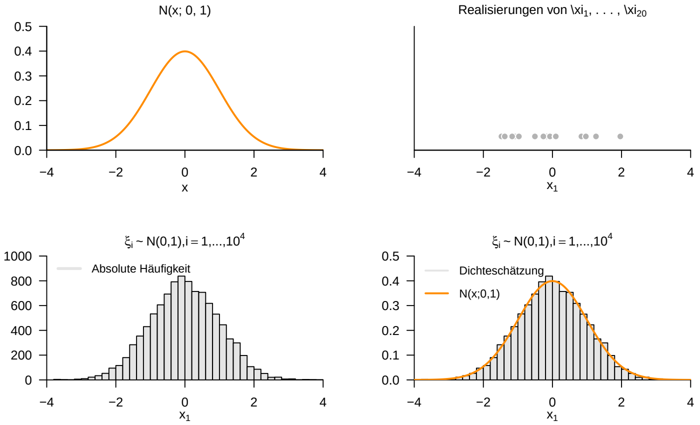
```
rnorm(1e1,5,sqrt(2))                # \xi_i \sim N(5,2), i = 1,...,10 (u.i.v.)
```

```
 [1] 6.62 2.42 4.65 4.65 4.60 4.22 5.89 7.92 2.69 5.72
```

© 2023 Dirk Ostwald CC BY-SA 4.0

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 υ, 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 $\upsilon := \xi^2$ und sind $x_1 = 0.10, x_2 = -0.20, x_3 = 0.80$ drei u.i.v. Realisierungen von ξ, so entspricht dies den u.i.v. Realisierungen $y_1 = x_1^2 = 0.01, y_2 = x_2^2 = 0.04, y_3 = x_3^2 = 0.64$ von υ. In diesem Beispiel fällt auf, dass υ keine negativen Werte annimmt, die Verteilung von υ 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)
```

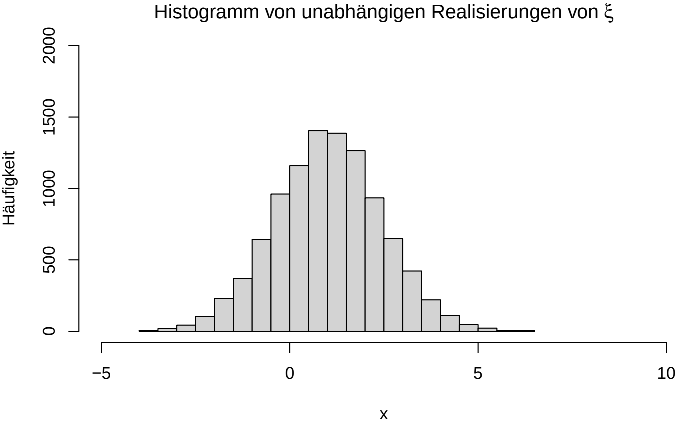
```
[1] -1.63  0.26  0.93  1.77 -0.29  1.66  1.51 -0.84
```

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

```
[1] 2.672 0.069 0.857 3.126 0.086 2.762 2.290 0.714
```

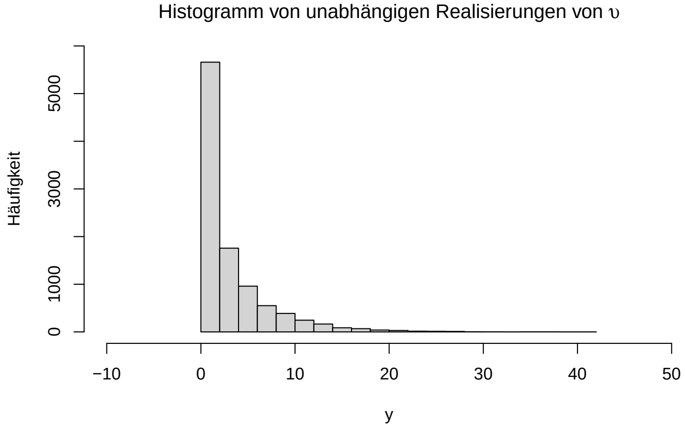
Vorbemerkungen

Simulation der Transformation normalverteilter Zufallsvariablen in R



Vorbemerkungen

Simulation der Transformation normalverteilter Zufallsvariablen in R



Theorem (Transformation eines Zufallsvektors)

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

$$\upsilon : \Omega \to \mathbb{R}, \omega \mapsto \upsilon(\omega) := (f \circ \xi)(\omega) := f(\xi(\omega)) \tag{1}$$

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 υ als Folge der Messbarkeit von ξ nachgewiesen werden.
- Im Folgenden ist oft $\mathcal{X} := \mathbb{R}$ und $f : \mathbb{R} \to \mathbb{R}$.
- Wir schreiben in diesem Fall in der Regel einfach $\upsilon := f(\xi)$ und nennen υ 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 $\upsilon := f(\xi)$ die durch f transformierte Zufallsvariable ist, dann gilt für die WDF von υ die folgende Formel: $p_\upsilon := \{\text{Formel}\}$".

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 $\xi_i, i = 1, ..., n$ unabhängig und identisch normalverteilte Zufallsvariablen sind und $\upsilon := f(\xi_1, ..., \xi_n)$ eine Transformation dieser Zufallsvariablen ist, dann ist die WDF von υ durch die Formel $p_\upsilon := \{\text{Formel}\}$ gegeben und man nennt die Verteilung von υ *Verteilungsname*".

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

(1) die Zentralen Grenzwertsätze die Annahme additiv unabhängig normalverteilter Störvariablen, und damit normalverteilter Daten, rechtfertigt,
(2) wie wir in der nächsten Vorlesungseinheit sehen werden, es sich bei Schätzern und Statistiken um Transformationen von Zufallsvariablen handelt, und
(3) Konfidenzintervalle und Hypothesentests durch die Verteilungen ihrer jeweiligen Statistiken charakterisiert und gerechtfertigt sind.

Vorbemerkungen

Ausblick

Das probabilistische Standardmodell von n Datenpunkten hat die Form

$$y_i := \mu_i + \varepsilon_i, i = 1, ..., n. \tag{2}$$

Die Zufallsvariable y_i dient dabei als das Modell des $i$ten Datenpunktes $\tilde{y}_i \in \mathbb{R}$, d.h. $\tilde{y}_i$ wird als Realisierung von y_i modelliert. Die Normalverteilung $y_i \sim N(\mu_i, \varepsilon)$ der Zufallsvariable y_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 y_i. $\varepsilon_i \sim N(0, \sigma^2)$ dagegen repräsentiert den stochastischen Aspekt des Datenpunktmodells und liefert im Sinne der Zentralen Grenzwertsätzte die theoretischer Erklärung für die Differenz von μ_i und y_i als Resultat der Addition vieler weiterer Einflüsse in der Generation von y_i über μ_i hinaus.

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

Vorbemerkungen

Transformationstheoreme

Standardtransformationen

- Summentransformation
- Mittelwerttransformation
- Z-Transformation
- χ^2-Transformation
- T-Transformation
- F-Transformation

Selbstkontrollfragen

Überblick

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

Das **univariate WDF Transformationstheorem bei linear affinen Abbildungen** gibt eine Formel zur Berechnung der WDF p_υ von $\upsilon := f(\xi)$ 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_υ von $\upsilon := f(\xi)$ 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_υ von $\upsilon := f(\xi)$, wenn ξ ein Zufallsvektor mit WDF p_ξ ist und f eine bijektive multivariate vektorwertigeFunktion ist.

Das **Faltungstheorem** liefert eine Formel zur Berechnung der WDF p_υ von $\upsilon := \xi_1 + \xi_2$, 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

$$\upsilon := f(\xi) \tag{3}$$

wobei die univariate reellwertige Funktion $f :]a,b[\to \mathbb{R}$ 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 Umkehrunktion 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 υ gegeben durch

$$p_\upsilon : \mathbb{R} \to \mathbb{R}_{\geq 0}, y \mapsto p_\upsilon(y) := \begin{cases} \frac{1}{|f'(f^{-1}(y))|} p_\xi \left(f^{-1}(y)\right) & \text{für } y \in f(]a,b[) \\ 0 & \text{für } y \in \mathbb{R} \setminus f(]a,b[). \end{cases} \tag{4}$$

Bemerkungen

- Linear-affine Abbildungen sind ein wichtiger Anwendungsfalls.
- 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

$$P_\upsilon(y) = \mathbb{P}(\upsilon \le y) = \mathbb{P}(f(\xi) \le y) = \mathbb{P}\left(f^{-1}(f(\xi)) \le f^{-1}(y)\right) = \mathbb{P}\left(\xi \le f^{-1}(y)\right) = P_\xi\left(f^{-1}(y)\right).$$

P_υ 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_υ sich ergibt wie folgt:

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

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

$$P_\upsilon(y) = \mathbb{P}(f(\xi) \le y) = \mathbb{P}\left(f^{-1}(f(\xi)) \ge f^{-1}(y)\right) = \mathbb{P}\left(\xi \ge f^{-1}(y)\right) = 1 - P_\xi\left(f^{-1}(y)\right),$$

Mit der Kettenregel und dem Satz von der Umkehrabbildung folgt dann

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

Weil f^{-1} strikt fallend ist, ist $d/dy(f^{-1}(y))$ negativ, so dass $-d/dy(f^{-1}(y))$ gleich $|d/dy(f^{-1}(y))|$ ist und das Theorem trifft zu.

Theorem (Univariate WDF Transformation bei linear-affinen Abbildungen)

ξ sei eine Zufallsvariable mit WDF p_ξ und es sei

$$\upsilon = f(\xi) \text{ mit } f(\xi) := a\xi + b \text{ für } a \neq 0. \tag{5}$$

Dann ist die WDF von υ gegeben durch

$$p_\upsilon : \mathbb{R} \to \mathbb{R}_{\geq 0}, y \mapsto p_\upsilon(y) := \frac{1}{|a|} p_\xi \left(\frac{y-b}{a} \right). \tag{6}$$

Bemerkung

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

Transformationstheoreme

Beweis

Wir halten zunächst fest, dass

$$f^{-1} : \mathbb{R} \to \mathbb{R}, y \mapsto f^{-1}(y) = \frac{y-b}{a} \tag{7}$$

ist, weil dann $f \circ f^{-1} = \mathrm{id}_{\mathbb{R}}$ gilt, wie man anhand von

$$f(f^{-1}(x)) = a\left(\frac{x-b}{a}\right) + b = x - b + b = x \text{ für alle } x \in \mathbb{R} \tag{8}$$

einsieht. Wir halten weiterhin fest, dass

$$f' : \mathbb{R} \to \mathbb{R}, x \mapsto f'(x) = \frac{d}{dx}(ax+b) = a. \tag{9}$$

Also folgt mit dem Theorem zur WDF Transformation bei bijektiven Abbildungen, dass

$$\begin{aligned} p_\upsilon : \mathbb{R} \to \mathbb{R}_{\geq 0}, y \mapsto p_\upsilon(y) &= \frac{1}{|f'\left(f^{-1}(y)\right)|} p_\xi\left(f^{-1}(y)\right) \\ &= \frac{1}{|a|} p_\xi\left(\frac{y-b}{a}\right). \end{aligned} \tag{10}$$

□

Theorem (WDF Transformation bei stückweise bijektiven Abbildungen)

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

$$\upsilon = f(\xi), \tag{11}$$

wobei f so beschaffen sei, dass der Ergebnisraum von ξ in eine endliche Anzahl von Mengen $\mathcal{X}_1, ..., \mathcal{X}_k$ mit einer entsprechenden Anzahl von Mengen $\mathcal{Y}_1 := f(\mathcal{X}_1), ..., \mathcal{Y}_k := f(\mathcal{X}_k)$ im Ergebnisraum $\mathcal{Y}$ von υ partitioniert werden kann (wobei nicht notwendigerweise $\mathcal{Y}_i \cap \mathcal{Y}_j = \emptyset, 1 \le i, j \le k$ gelten muss), so dass die Abbildung f für alle $\mathcal{X}_1, ..., \mathcal{X}_k$ bijektiv ist (d.h. f ist eine *stückweise* bijektive Abbildung). Für $i = 1, ..., k$ bezeichne f_i^{-1} die Umkehrfunktion von f auf $\mathcal{Y}_i$. Schließlich nehmen wir an, dass die Ableitungen f_i' für alle $i = 1, ..., k$ existieren und stetig sind. Dann ist eine WDF von υ durch

$$p_\upsilon : \mathcal{Y} \to \mathbb{R}_{\ge 0}, y \mapsto p_\upsilon(y) := \sum_{i=1}^k 1_{\mathcal{Y}_i}(y) \frac{1}{|f_i'(f_i^{-1}(y))|} p_\xi \left(f_i^{-1}(y)\right). \tag{12}$$

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

$$\upsilon := f(\xi), \tag{13}$$

wobei die multivariate vektorwertige Funktion $f : \mathbb{R}^n \to \mathbb{R}^n$ differenzierbar und bijektiv auf $]a, b[$ sei. Schließlich seien

$$J^f(x) = \left(\frac{\partial}{\partial x_j} f_i(x)\right)_{1 \le i \le n, 1 \le j \le n} \in \mathbb{R}^{n \times n} \tag{14}$$

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 υ durch

$$p_\upsilon : \mathbb{R}^n \to \mathbb{R}_{\ge 0}, y \mapsto p_\upsilon(y) := \begin{cases} \frac{1}{|J^f\left(f^{-1}(y)\right)|} p_\xi\left(f^{-1}(y)\right) & \text{for } y \in f(\mathbb{R}^n) \\ 0 & \text{for } y \in \mathbb{R}^n \setminus f(\mathbb{R}^n) \end{cases} \tag{15}$$

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. $\upsilon := \xi_1 + \xi_2$ sei die Summe von ξ_1 und ξ_2. Dann ergibt sich eine WDF der Verteilung von υ als

$$p_\upsilon(y) = \int_{-\infty}^{\infty} p_{\xi_1}(y - x_2) p_{\xi_2}(x_2)\, dx_2 = \int_{-\infty}^{\infty} p_{\xi_1}(x_1) p_{\xi_2}(y - x_1)\, dx_1 \qquad (16)$$

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

Bemerkung

- Die Summen- und Mittelwerttransformation sind wichtige Anwendungsfälle.

Beweis

Wir nutzen das multivariate WDF Transformationstheorem für bijektive Abbildungen. Dazu definieren wir zunächst

$$f : \mathbb{R}^2 \to \mathbb{R}^2, x \mapsto f(x) := \begin{pmatrix} x_1 + x_2 \\ x_2 \end{pmatrix} := \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \tag{17}$$

Die inverse Funktion von f ist dann gegeben durch

$$f : \mathbb{R}^2 \to \mathbb{R}^2, z \mapsto f(z) := \begin{pmatrix} z_1 - x_2 \\ z_2 \end{pmatrix} \tag{18}$$

weil dann $f \circ f^{-1} = \text{id}_{\mathbb{R}^2}$ gilt, wie man anhand von

$$f^{-1}(f(x)) = f^{-1}\begin{pmatrix} x_1 + x_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 - x_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{19}$$

einsieht. Die Jacobimatrix von f ergibt sich zu

$$J^f(x) = \begin{pmatrix} \frac{\partial}{\partial x_1} f_1(x) & \frac{\partial}{\partial x_2} f_1(x) \\ \frac{\partial}{\partial x_1} f_2(x) & \frac{\partial}{\partial x_2} f_2(x) \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x_1}(x_1 + x_2) & \frac{\partial}{\partial x_2}(x_1 + x_2) \\ \frac{\partial}{\partial x_1} x_2 & \frac{\partial}{\partial x_2} x_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{20}$$

und die Jacobideterminante damit zu $|J^f(x)| = 1$.

Transformationstheoreme

Beweis (fortgeführt)

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

$$p_{\xi_1,\xi_2}(x_1, x_2) = p_{\xi_1}(x_1)p_{\xi_2}(x_2) \tag{21}$$

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

$$\begin{aligned}
p_\zeta(z) &= \frac{1}{|J^f\left(f^{-1}(z)\right)|} p_\xi\left(f^{-1}(z)\right) \\
&= \frac{1}{1} p_{\xi_1,\xi_2}(z_1 - x_2, x_2) \\
&= p_{\xi_1}(z_1 - x_2)p_{\xi_2}(x_2)
\end{aligned} \tag{22}$$

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

$$p_{\zeta_1}(z_1) = \int_{-\infty}^{\infty} p_{\xi_1}(z_1 - x_2)p_{\xi_2}(x_2)\, dx_2 \tag{23}$$

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

$$p_\upsilon(y) = \int_{-\infty}^{\infty} p_{\xi_1}(y - x_2)p_{\xi_2}(x_2)\, dx_2. \tag{24}$$

$\square$

© 2023 Dirk Ostwald CC BY-SA 4.0

Vorbemerkungen

Transformationstheoreme

Standardtransformationen

- Summentransformation
- Mittelwerttransformation
- Z-Transformation
- χ^2-Transformation
- T-Transformation
- F-Transformation

Selbstkontrollfragen

Standardtransformationen

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 Varianzsparameters 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.

Vorbemerkungen

Transformationstheoreme

Standardtransformationen

- **Summentransformation**
- Mittelwerttransformation
- Z-Transformation
- χ^2-Transformation
- T-Transformation
- F-Transformation

Selbstkontrollfragen

Theorem (Summe unabhängig normalverteilter Zufallsvariablen)

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

$$\upsilon \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right) \tag{25}$$

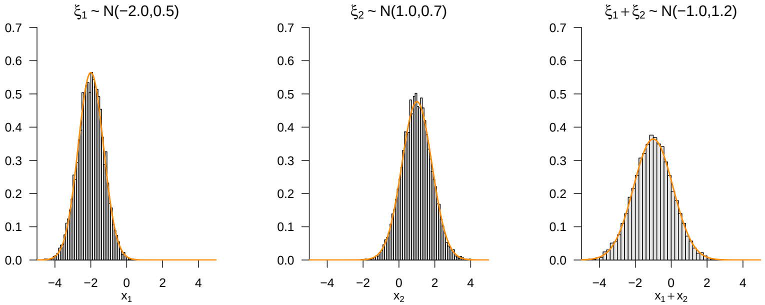
Für unabhängige und identisch normalverteilte Zufallsvariablen $\xi_i \sim N(\mu, \sigma^2)$ gilt folglich

$$\upsilon \sim N(n\mu, n\sigma^2). \tag{26}$$

Bemerkungen

- Die Mittelwerttransformation ist ein wichtiger Anwendungsfall.
- Die Generalisierung der zentralen Grenzwertsätze sind wichtige Anwendungsfälle.

Summentransformation



Summentransformation

Beweis

Wir skizzieren mithilfe der Faltungsformel, dass für $\xi_1 \sim N(\mu_1, \sigma_1^2)$, $\xi_2 \sim N(\mu_2, \sigma_2^2)$, und $\upsilon := \xi_1 + \xi_2$ gilt, dass $\upsilon \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Für $n > 2$ folgt das Theorem dann durch Iteration. Mit der Definition der WDF der Normalverteilung erhalten wir zunächst

$$
\begin{aligned}
p_\upsilon(y) &= \int_{-\infty}^{\infty} p_{\xi_1}(x_1) p_{\xi_2}(y - x_1)\, dx_1 \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2\right) \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2}\left(\frac{y - x_1 - \mu_2}{\sigma_2}\right)^2\right) dx_1 \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2}\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 - \frac{1}{2}\left(\frac{y - x_1 - \mu_2}{\sigma_2}\right)^2\right) dx_1 .
\end{aligned}
\tag{27}
$$

Mit einigem algebraischen Aufwand erhält man die Identität

$$
\begin{aligned}
&-\frac{1}{2}\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 - \frac{1}{2}\left(\frac{y - x_1 - \mu_2}{\sigma_2}\right)^2 \\
&\qquad = -\frac{(y - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)} - \frac{((\sigma_1^2 + \sigma_2^2)x_1 - \sigma_1^2 y + \mu_2\sigma_1^2 - \mu_1\sigma_2^2)^2}{2\sigma_1^2\sigma_2^2(\sigma_1^2 + \sigma_2^2)},
\end{aligned}
\tag{28}
$$

so dass weiterhin gilt, dass

Beweis (fortgeführt)

$$
\begin{aligned}
p_{\upsilon}(y) &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{(y-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)} - \frac{((\sigma_1^2+\sigma_2^2)x_1 - \sigma_1^2 y + \mu_2\sigma_1^2 - \mu_1\sigma_2^2)^2}{2\sigma_1^2\sigma_2^2(\sigma_1^2+\sigma_2^2)}\right) dx_1 \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{(y-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right) \exp\left(-\frac{((\sigma_1^2+\sigma_2^2)x_1 - \sigma_1^2 y + \mu_2\sigma_1^2 - \mu_1\sigma_2^2)^2}{2\sigma_1^2\sigma_2^2(\sigma_1^2+\sigma_2^2)}\right) dx_1 \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{(y-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{((\sigma_1^2+\sigma_2^2)x_1 - \sigma_1^2 y + \mu_2\sigma_1^2 - \mu_1\sigma_2^2)^2}{2\sigma_1^2\sigma_2^2(\sigma_1^2+\sigma_2^2)}\right) dx_1 .
\end{aligned}
\tag{29}
$$

Summentransformation

Beweis (fortgeführt)

Für das verbleibende Integral zeigt man mithilfe der Integration durch Substitution, dass

$$\int_{-\infty}^{\infty} \exp\left(-\frac{((\sigma_1^2+\sigma_2^2)x_1 - \sigma_1^2 y + \mu_2\sigma_1^2 - \mu_1\sigma_2^2)^2}{2\sigma_1^2\sigma_2^2(\sigma_1^2+\sigma_2^2)}\right)\, dx_1 = \frac{\sqrt{2\pi}\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}. \tag{30}$$

Es ergibt sich also

$$\begin{aligned} p_\upsilon(y) &= \frac{1}{2\pi\sigma_1\sigma_2}\frac{\sqrt{2\pi}\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}} \exp\left(-\frac{(y-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right) \\ &= \frac{(2\pi)^{-1}(2\pi)^{\frac{1}{2}}}{\sqrt{\sigma_1^2+\sigma_2^2}} \exp\left(-\frac{(y-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right) \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}} \exp\left(-\frac{(y-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right). \end{aligned} \tag{31}$$

Summentransformation

Beweis (fortgeführt)

Schließlich folgt, dass

$$
\begin{aligned}
p_\upsilon(y) &= \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp\left(-\frac{1}{2(\sigma_1^2 + \sigma_2^2)}\,(y - (\mu_1 + \mu_2))^2\right) \\
&= N(y; \mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)
\end{aligned}
\tag{32}
$$

Ein einfacheres Vorgehen ergibt sich vermutlich nach Fouriertransformation der WDF im Sinne der sogenannten charakteristischen Funktion einer Zufallsvariable. In diesem Fall würde die Faltung der WDFen der Multiplikation der charakteristischen Funktionen entsprechen.

$\square$

Vorbemerkungen

Transformationstheoreme

Standardtransformationen

- Summentransformation
- **Mittelwerttransformation**
- Z-Transformation
- χ^2-Transformation
- T-Transformation
- F-Transformation

Selbstkontrollfragen

Mittelwerttransformation

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

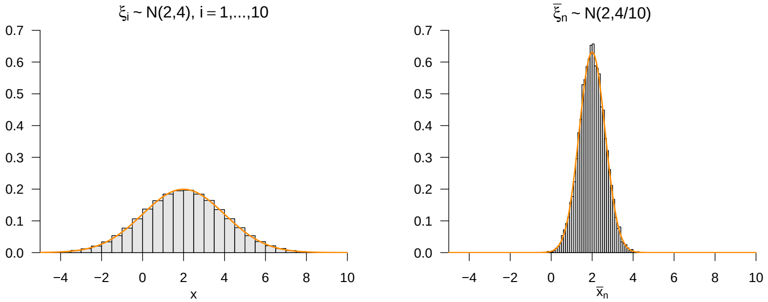
Für $i = 1, ..., n$ seien $\xi_i \sim N(\mu, \sigma^2)$ unabhängig und identisch normalverteilte Zufallsvariablen. Dann gilt für das Stichprobenmittel $\bar{\xi}_n := \frac{1}{n}\sum_{i=1}^n \xi_i$, dass

$$\bar{\xi}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right). \tag{33}$$

Bemerkung

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

Mittelwerttransformation



Mittelwerttransformation

Beweis

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

$$
\begin{aligned}
p_{\bar{\xi}_n}(\bar{x}_n) &= \frac{1}{|1/n|} N\left(n\bar{x}_n; n\mu, n\sigma^2\right) \\
&= \frac{n}{\sqrt{2\pi n\sigma^2}} \exp\left(-\frac{1}{2n\sigma^2}\left(n\bar{x}_n - n\mu\right)^2\right) \\
&= \frac{n}{\sqrt{2\pi n\sigma^2}} \exp\left(-\frac{1}{2n\sigma^2}\left(n\bar{x}_n - n\mu\right)^2\right) \\
&= nn^{-\frac{1}{2}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(n\bar{x}_n)^2}{2n\sigma^2} + \frac{2(n\bar{x}_n)(n\mu)}{2n\sigma^2} - \frac{(n\mu)^2}{2n\sigma^2}\right) \\
&= \sqrt{n}\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{n\bar{x}_n^2}{2\sigma^2} + \frac{2n\bar{x}_n\mu}{2\sigma^2} - \frac{n\mu^2}{2\sigma^2}\right) \\
&= \frac{1}{1/\sqrt{n}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\bar{x}_n^2}{2(\sigma^2/n)} + \frac{2\bar{x}_n\mu}{2(\sigma^2/n)} - \frac{\mu^2}{2(\sigma^2/n)}\right) \\
&= \frac{1}{\sqrt{2\pi(\sigma^2/n)}} \exp\left(-\frac{1}{2(\sigma^2/n)}(\bar{x}_n - \mu)^2\right) \\
&= N\left(\bar{x}_n; \mu, \sigma^2/n\right)
\end{aligned}
\tag{34}
$$

Vorbemerkungen

Transformationstheoreme

Standardtransformationen

- Summentransformation
- Mittelwerttransformation
- **Z-Transformation**
- χ^2-Transformation
- T-Transformation
- F-Transformation

Selbstkontrollfragen

Definition (z-Zufallsvariable)

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

$$p : \mathbb{R} \to \mathbb{R}_{>0}, z \mapsto p(z) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right). \tag{35}$$

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

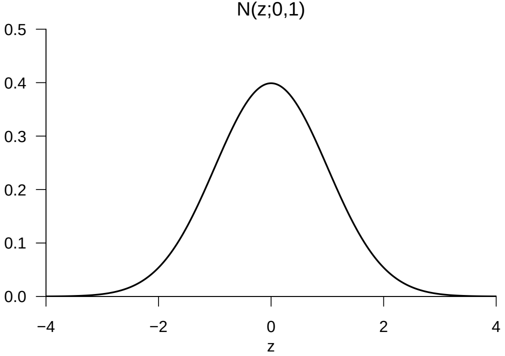
$$N(z; 0, 1) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right). \tag{36}$$

Bemerkung

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

Z-Transformation

Wahrscheinlichkeitsdichtefunktion einer z-Zufallsvariable



Theorem (Z-Transformation)

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

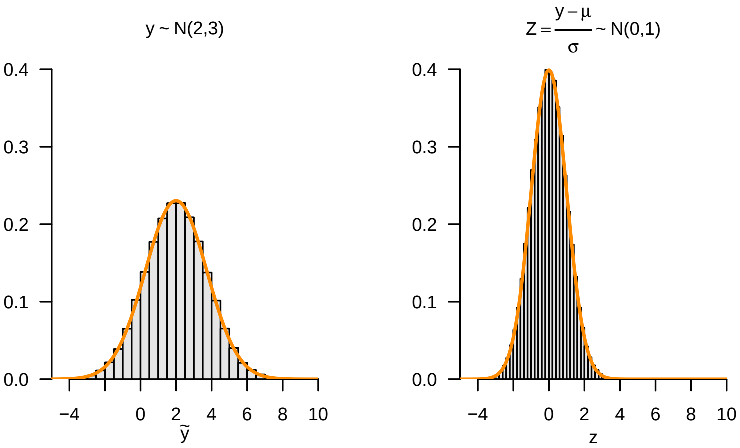
$$Z := \frac{y - \mu}{\sigma} \tag{37}$$

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

Bemerkungen

- Wir benutzen hier den Bezeichner y für eine normalverteilte Zufallsvariable. Werte, die diese Zufallsvariable annehmen kann, bezeichnen wir in der Folge mit $\tilde{y}$.
- Z wird hier als $(y - \mu)/\sigma$ definiert. Dass ein solches Z aber eine z-Zufallsvariable ist, muss bewiesen werden und ergibt sich nicht einfach durch die Wahl des Bezeichners für $(y - \mu)/\sigma$, 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.

Z-Transformation



Beweis

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

$$\zeta : \mathbb{R} \to \mathbb{R}, \tilde{y} \mapsto \zeta(\tilde{y}) := \frac{\tilde{y} - \mu}{\sigma} =: z \tag{38}$$

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

$$\zeta^{-1} : \mathbb{R} \to \mathbb{R}, z \mapsto \zeta^{-1}(z) := \sigma z + \mu \tag{39}$$

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

$$\zeta^{-1}(z) = \zeta^{-1}\left(\frac{\tilde{y} - \mu}{\sigma}\right) = \frac{\sigma(\tilde{y} - \mu)}{\sigma} + \mu = \tilde{y} - \mu + \mu = \tilde{y}. \tag{40}$$

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

$$\zeta'(\tilde{y}) = \frac{d}{d\tilde{y}}\left(\frac{\tilde{y} - \mu}{\sigma}\right) = \frac{d}{d\tilde{y}}\left(\frac{\tilde{y}}{\sigma} - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma}. \tag{41}$$

Z-Transformation

Beweis (fortgeführt)

Einsetzen in das univariate WDF Transformationstheorem für lineare Funktionen ergibt dann

$$
\begin{aligned}
p_\zeta(z) &= \frac{1}{|1/\sigma|} N\left(\sigma z + \mu; \mu, \sigma^2\right) \\
&= \frac{1}{1/\sqrt{\sigma^2}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(\sigma z + \mu - \mu\right)^2\right) \\
&= \frac{\sqrt{\sigma^2}}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\sigma^2 z^2\right) \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2\right) \\
&= N(z; 0, 1)
\end{aligned}
\tag{42}
$$

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

Vorbemerkungen

Transformationstheoreme

Standardtransformationen

- Summentransformation
- Mittelwerttransformation
- Z-Transformation
- χ^2-**Transformation**
- T-Transformation
- F-Transformation

Selbstkontrollfragen

χ^2-Transformation

Definition (χ^2-Zufallsvariable)

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

$$p : \mathbb{R}_{>0} \to \mathbb{R}_{>0}, u \mapsto p(u) := \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} u^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}u\right), \tag{43}$$

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

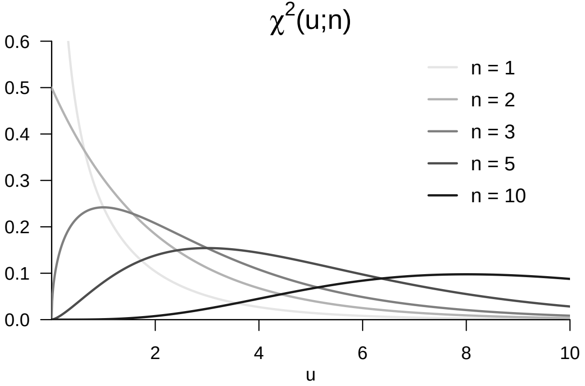
$$\chi^2(u; n) := \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} u^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}u\right). \tag{44}$$

Bemerkung

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

χ^2-Transformation

Wahrscheinlichkeitsdichtefunktionen von ξ^2-Zufallsvariablen



Steigendes n verbreitert $\chi^2(u;n)$ und verschiebt Masse zur größeren Werten.

χ^2-Transformation

Theorem (χ^2-Transformation)

$Z_1, ..., Z_n \sim N(0,1)$ seien unabhängig und identisch verteilte z-Zufallsvariablen. Dann ist die Zufallsvariable

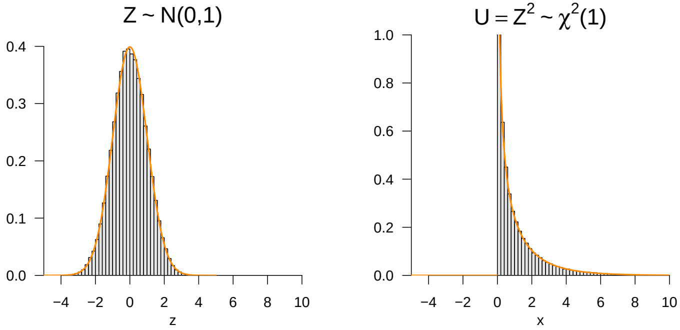
$$U := \sum_{i=1}^{n} Z_i^2 \tag{45}$$

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

Bemerkungen

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

χ^2-Transformation



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

$$p_U(u) = \sum_{i=1}^{k} 1_{\mathcal{U}_i} \frac{1}{|f_i'(f_i^{-1}(u))|} p_\zeta \left(f_i^{-1}(u)\right). \quad (46)$$

Wir definieren

$$\mathcal{U}_1 :=] - \infty, 0[, \mathcal{U}_2 :=]0, \infty[, \text{ und } \mathcal{U}_i := \mathbb{R}_{>0} \text{ für } i = 1, 2, \quad (47)$$

sowie

$$f_i : \mathcal{Z}_i \to \mathcal{U}_i, x \mapsto f_i(z) := z^2 =: u \text{ für } i = 1, 2. \quad (48)$$

Die Ableitung und die Umkehrfunktion der f_i ergeben sich zu

$$f_i' : \mathcal{Z}_i \to \mathcal{Z}_i, x \mapsto f_i'(z) = 2z \text{ für } i = 1, 2, \quad (49)$$

und

$$f_1^{-1} : \mathcal{U}_1 \to \mathcal{U}_1, u \mapsto f_1^{-1}(u) = -\sqrt{u} \text{ und } f_2^{-1} : \mathcal{U}_2 \to \mathcal{U}_2, u \mapsto f_2^{-1}(u) = \sqrt{u}, \quad (50)$$

respektive.

χ^2-Transformation

Beweis (fortgeführt)

Einsetzen in Gleichung (46) ergibt dann

$$
\begin{aligned}
p_U(u) &= 1_{\mathcal{U}_1}(u)\frac{1}{|f_1'(f_1^{-1}(u))|}p_\zeta\left(f_1^{-1}(u)\right) + 1_{\mathcal{U}_2}(u)\frac{1}{|f_2'(f_2^{-1}(u))|}p_\zeta\left(f_2^{-1}(u)\right) \\
&= \frac{1}{|2(-\sqrt{u})|}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(-\sqrt{u})^2\right) + \frac{1}{|2(\sqrt{u})|}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(\sqrt{u})^2\right) \\
&= \frac{1}{2\sqrt{u}}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u\right) + \frac{1}{2\sqrt{u}}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u\right) \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{u}}\exp\left(-\frac{1}{2}u\right).
\end{aligned}
\tag{51}
$$

Andererseits gilt, dass mit $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, die PDF einer χ^2-Zufallsvariable U mit $n = 1$ durch

$$
\frac{1}{\Gamma\left(\frac{1}{2}\right)2^{\frac{1}{2}}}u^{\frac{1}{2}-1}\exp\left(-\frac{1}{2}u\right) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{u}}\exp\left(-\frac{1}{2}u\right) \tag{52}
$$

gegeben ist. Also gilt, dass wenn $Z \sim N(0,1)$ ist, dann ist $U := Z^2 \sim \chi^2(1)$.

Vorbemerkungen

Transformationstheoreme

Standardtransformationen

- Summentransformation
- Mittelwerttransformation
- Z-Transformation
- χ^2-Transformation
- T-**Transformation**
- F-Transformation

Selbstkontrollfragen

Definition (t-Zufallsvariable)

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

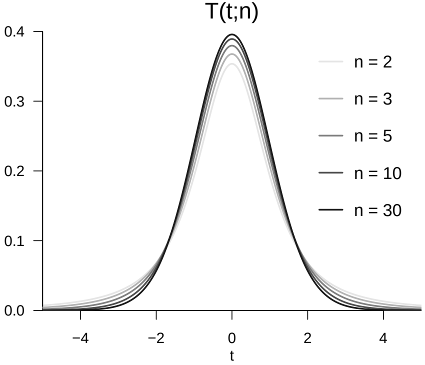
$$p : \mathbb{R} \to \mathbb{R}_{>0}, t \mapsto p(t) := \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \tag{53}$$

wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass T einer t-Verteilung mit n Freiheitsgraden unterliegt und nennen T eine t-Zufallsvariable mit n Freiheitsgraden. Wir kürzen dies mit $T \sim t(n)$ ab. Die WDF einer t-Zufallsvariable bezeichnen wir mit

$$T(t; n) := \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}. \tag{54}$$

T-Transformation

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 n Freiheitsgraden, und Z und U seien unabhängige Zufallsvariablen. Dann ist die Zufallsvariable

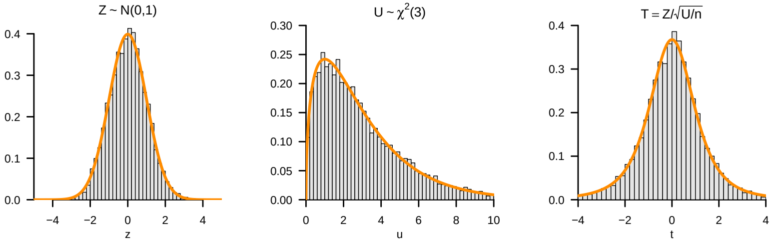
$$T := \frac{Z}{\sqrt{U/n}} \tag{55}$$

eine t-verteilte Zufallsvariable mit n Freiheitsgraden, es gilt also $T \sim t(n)$.

Bemerkungen

- Das Theorem geht auf Student (1908) zurück.
- Das Theorem ist das zentrale Resultat der Frequentistischen Statistik.
- Zabell (2008) gibt hierzu einen historischen Überblick.
- Die T-Konfidenzintervallstatistik und die T-Teststatistik sind wichtige Anwendungsfälle.

T-Transformation



Beweis

Wir halten zunächst fest, dass die zweidimensionale WDF der gemeinsamen (unabhängigen) Verteilung von Z und U durch

$$p_{Z,U}(z,u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} u^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}u\right). \tag{56}$$

gegeben ist. Wir betrachten dann die multivariate vektorwertige Abbildung

$$f : \mathbb{R}^2 \to \mathbb{R}^2, (z,u) \mapsto f(z,u) := \left(\frac{z}{\sqrt{u/n}}, u\right) =: (t,w) \tag{57}$$

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 $\upsilon := f(\xi)$ für eine differenzierbare und bijektive Abbildung $f : \mathbb{R}^n \to \mathbb{R}^n$ ist, die WDF des Zufallsvektors υ durch

$$p_\upsilon : \mathbb{R}^n \to \mathbb{R}_{\geq 0}, y \mapsto p_\upsilon(y) := \frac{1}{|J^f\left(f^{-1}(y)\right)|} p_\xi\left(f^{-1}(y)\right) \tag{58}$$

gegeben ist. Für die im vorliegenden Fall betrachtete Abbildung halten wir zunächst fest, dass

$$f^{-1} : \mathbb{R}^2 \to \mathbb{R}^2, (t,w) \mapsto f^{-1}(t,w) := \left(\sqrt{w/n}t, w\right). \tag{59}$$

Beweis (fortgeführt)

Dies ergibt sich direkt aus

$$f^{-1}(f(z,u)) = f^{-1}\left(\frac{z}{\sqrt{u/n}}, u\right) = \left(\frac{\sqrt{u/n}z}{\sqrt{u/n}}, u\right) = (z,u) \text{ für alle } (z,u) \in \mathbb{R}^2. \qquad (60)$$

Wir halten dann fest, dass die Determinante der Jacobi-Matrix von f an der Stelle (z,u) durch

$$|J^f(z,u)| = \begin{vmatrix} \frac{\partial}{\partial z}\left(\frac{z}{\sqrt{u/n}}\right) & \frac{\partial}{\partial u}\left(\frac{z}{\sqrt{u/n}}\right) \\ \frac{\partial}{\partial z}u & \frac{\partial}{\partial u}u \end{vmatrix} = \left(\frac{v}{n}\right)^{-1/2}, \qquad (61)$$

gegeben ist, sodass folgt, dass

$$\frac{1}{|J^f\left(f^{-1}(z,u)\right)|} = \left(\frac{w}{n}\right)^{1/2}. \qquad (62)$$

Einsetzen in Gleichung (58) ergibt dann

$$p_{T,W}(t,w) = \left(\frac{w}{n}\right)^{1/2} p_{Z,V}\left(\sqrt{w/n}t, w\right), \qquad (63)$$

T-Transformation

Beweis (fortgeführt)

Es folgt also

$$
\begin{aligned}
p_T(t) &= \int_0^\infty p_{T,W}(t,w)\,dw \\
&= \int_0^\infty \left(\frac{w}{n}\right)^{1/2} p_{Z,V}\left(\sqrt{w/n}t, w\right)\,dw \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\sqrt{w/n}t)^2\right) \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} w^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}w\right) \left(\frac{w}{n}\right)^{1/2}\,dw \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}\,n^{\frac{1}{2}}} \int_0^\infty \exp\left(-\frac{1}{2}\frac{w}{n}t^2\right) w^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}w\right) w^{1/2}\,dw \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}\,n^{\frac{1}{2}}} \int_0^\infty \exp\left(-\frac{1}{2}\frac{w}{n}t^2 - \frac{1}{2}w\right) w^{\frac{n}{2}-1} w^{\frac{1}{2}}\,dw \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}\,n^{\frac{1}{2}}} \int_0^\infty \exp\left(-\frac{1}{2}\left(\frac{w}{n}t^2 + w\right)\right) w^{\frac{n+1}{2}-1}\,dw \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}\,n^{\frac{1}{2}}} \int_0^\infty \exp\left(-\frac{1}{2}\left(1 + \frac{t^2}{n}\right)\right) w^{\frac{n+1}{2}-1}\,dw
\end{aligned}
\tag{64}
$$

© 2023 Dirk Ostwald CC BY-SA 4.0

T-Transformation

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:

$$
\begin{aligned}
\Gamma(w;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} w^{\alpha-1}\exp\left(-\frac{w}{\beta}\right) \\
\Rightarrow \Gamma\left(w; \frac{n+1}{2}, \frac{2}{1+\frac{t^2}{n}}\right) &= \frac{1}{\Gamma(\frac{n+1}{2})\left(\frac{2}{1+\frac{t^2}{n}}\right)^{\frac{n+1}{2}}} w^{\frac{n+1}{2}-1}\exp\left(-\frac{w}{\frac{2}{1+\frac{t^2}{n}}}\right) \\
&= \frac{1}{\Gamma(\frac{n+1}{2})\left(\frac{2}{1+\frac{t^2}{n}}\right)^{\frac{n+1}{2}}} \exp\left(-\frac{1}{2}\left(1+\frac{t^2}{n}\right)\right) w^{\frac{n+1}{2}-1}.
\end{aligned}
$$

Beweis (fortgeführt)

Es ergibt sich also

$$p_T(t) = \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}} n^{\frac{1}{2}}} \int_0^\infty \Gamma\left(w; \frac{n+1}{2}, \frac{2}{1+\frac{t^2}{n}}\right) dw. \tag{65}$$

Schließlich stellen wir fest, dass der Integralterm in obiger Gleichung dem Normalisierungsterm einer Gamma WDF entspricht. Abschließend ergibt sich also

$$p_T(t) = \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}} n^{\frac{1}{2}}} \Gamma\left(\frac{n+1}{2}\right) \left(\frac{2}{1+\frac{t^2}{n}}\right)^{\frac{n+1}{2}}. \tag{66}$$

Die Verteilung von $Z/\sqrt{U/n}$ hat also die WDF einer T-Zufallsvariable.

$\square$

Vorbemerkungen

Transformationstheoreme

Standardtransformationen

- Summentransformation
- Mittelwerttransformation
- Z-Transformation
- χ^2-Transformation
- T-Transformation
- F-**Transformation**

Selbstkontrollfragen

Definition (f-Zufallsvariable)

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

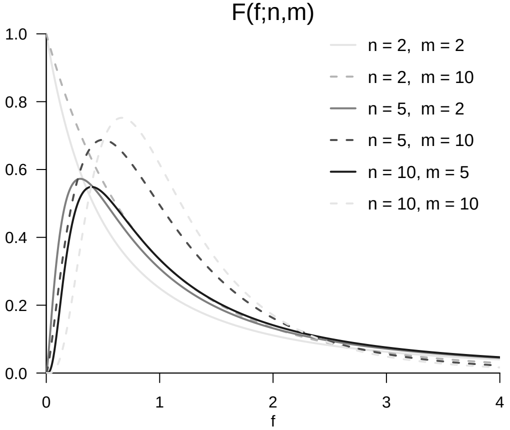
$$p_F : \mathbb{R} \to \mathbb{R}_{>0}, f \mapsto p_F(f) := m^{\frac{m}{2}} n^{\frac{n}{2}} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{f^{\frac{m}{2}-1}}{\left(1+\frac{m}{n}f\right)^{\frac{m+n}{2}}}, \tag{67}$$

wobei Γ die Gammafunktion bezeichne. Dann sagen wir, dass F einer f-Verteilung mit n, m Freiheitsgraden unterliegt und nennen F eine f-Zufallsvariable mit n, m Freiheitsgraden. Wir kürzen dies mit $F \sim f(n, m)$ ab. Die WDF einer f-Zufallsvariable bezeichnen wir mit

$$F(f; n, m) := m^{\frac{m}{2}} n^{\frac{n}{2}} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{f^{\frac{m}{2}-1}}{\left(1+\frac{m}{n}f\right)^{\frac{m+n}{2}}}. \tag{68}$$

F-Transformation

Wahrscheinlichkeitsdichtefunktionen von f-Zufallsvariablen



Theorem (F-Transformation)

$V \sim \chi^2(n)$ und $W \sim \chi^2(m)$ seien zwei unabhängige χ^2-Zufallfsvariablen mit n und m Freiheitsgraden, respektive. Dann ist die Zufallsvariable

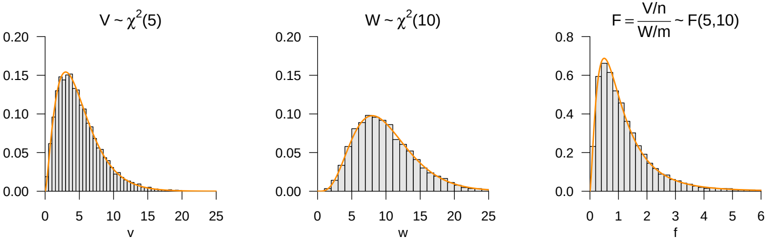
$$F := \frac{V/n}{W/m} \tag{69}$$

eine f-verteilte Zufallsvariable mit n, m Freiheitsgraden, 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.

F-Transformation



Vorbemerkungen

Transformationstheoreme

Standardtransformationen

- Summentransformation
- Mittelwerttransformation
- Z-Transformation
- χ^2-Transformation
- T-Transformation
- F-Transformation

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 Statistik.
4. Geben Sie das Summentransformationstheorem wieder.
5. Geben Sie das Mittelwerttransformationstheorem wieder.
6. Geben Sie das Z-Transformationstheorem wieder.
7. Geben Sie das χ^2-Transformationstheorem wieder.
8. Beschreiben Sie die WDF der t-Verteilung in Abhängigkeit ihrer Freiheitsgrade.
9. Geben Sie das T-Transformationstheorem wieder.
10. Geben Sie das F-Transformationstheorem wieder.

References

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

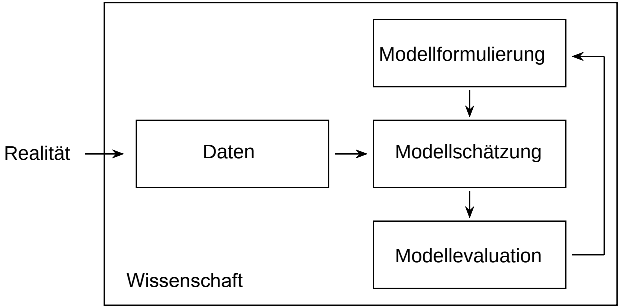
BSc Psychologie WiSe 2022/23

Prof. Dr. Dirk Ostwald

(9) Grundbegriffe Frequentistischer Inferenz

Datum	Einheit	Thema
13.10.2022	Einführung	(1) Einführung
20.10.2022	Wahrscheinlichkeitstheorie	(2) Wahrscheinlichkeitsräume
27.10.2022	Wahrscheinlichkeitstheorie	(3) Elementare Wahrscheinlichkeiten
03.11.2022	Wahrscheinlichkeitstheorie	(4) Zufallsvariablen I
10.11.2022	Wahrscheinlichkeitstheorie	(4) Zufallsvariablen II
17.11.2022	Wahrscheinlichkeitstheorie	(5) Multivariate Verteilungen
24.11.2022	Wahrscheinlichkeitstheorie	(6) Erwartungswert und Kovarianz
01.12.2022	Wahrscheinlichkeitstheorie	(7) Ungleichungen und Grenzwerte
08.12.2022	Wahrscheinlichkeitstheorie	(8) Transformationen der Normalverteilung
15.12.2022	Frequentistische Inferenz	(9) Grundbegriffe Frequentistischer Inferenz
	Weihnachtspause	
05.01.2023	Frequentistische Inferenz	(10) Parameterschätzung
12.01.2023	Frequentistische Inferenz	(11) Konfidenzintervalle
19.01.2023	Frequentistische Inferenz	(12) Hypothesentests I
26.01.2023	Frequentistische Inferenz	(12) Hypothesentests II
02.02.2023	Klausur	G44-H6, 16:00 - 17:00 Uhr
Jul 2023	Klausurwiederholungstermin	

Modellbasierte Datenwissenschaft



Frequentistische Inferenz

Modellformulierung $\Rightarrow$ Statistische Modelle
Modellschätzung $\Rightarrow$ Parameterschätzung und Konfidenzintervalle
Modellevaluation $\Rightarrow$ Hypothesentests (cf. Allgemeines Lineares Modell)

Statistische Modelle

Statistiken und Schätzer

Standardprobleme Frequentistischer Inferenz

Selbstkontrollfragen

Statistische Modelle

Statistiken und Schätzer

Standardprobleme Frequentistischer Inferenz

Selbstkontrollfragen

Definition (Statistisches Modell)

Ein *statistisches Modell* ist ein Tripel

$$\mathcal{M} := (\mathcal{Y}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\}) \tag{1}$$

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 statistisches Modell auch *parametrisches* statistisches Modell und Θ heißt *Parameterraum* des statistischen Modells.

Bemerkungen

- Für Erwartungswerte und (Ko)Varianzen bezüglich $\mathbb{P}_\theta$ schreiben wir $\mathbb{E}_\theta$, $\mathbb{V}_\theta$, $\mathbb{C}_\theta$.
- Ein statistisches Modell $\mathcal{M}$ heißt ein *diskretes Modell*, wenn $\mathcal{Y}$ diskret ist und jedes $\mathbb{P}_\theta$ eine WMF p_θ besitzt, ein statistisches Modell $\mathcal{M}$ heißt ein *stetiges Modell*, wenn $\mathcal{Y} \subset \mathbb{R}^n$ ist und jedes $\mathbb{P}_\theta$ eine WDF p_θ besitzt.
- Für ein statistisches Modell $\mathcal{M}_0 := (\mathcal{Y}_0, \mathcal{A}_0, \{\mathbb{P}_\theta^0 | \theta \in \Theta\})$ heißt das statistische 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 statistischen Modell*. 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 statistischen Modell*.

Statistische Modelle

Bemerkungen (fortgeführt)

- Der Vorgang der Datenbeobachtung wird durch einen Zufallsvektor υ, der Werte in $\mathcal{Y}$ annimmt, beschrieben. Im Kontext statistischer Modelle nennt man diesen Zufallsvektor *Daten*, *Beobachtung*, *Messung* oder *Stichprobe*. Eine Realisierung von υ, 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 $\upsilon := (\upsilon_1, ..., \upsilon_n)$ entspricht dann einer Menge von n unabhängigen Zufallsvariablen.
- Im Gegensatz zum Wahrscheinlichkeitsraummodell betrachtet man bei statistische Modellen zwei oder mehr Wahrscheinlichkeitsmaße, die die Verteilung von υ mutmaßlich bestimmen. Das jeweils zugrundeliegende Wahrscheinlichkeitsmaß ist mit $\theta \in \Theta$ indiziert,
- In einem konkreten Datenanalyseproblem nimmt man an, dass die beobachteten Werte $y = (y_1, ..., y_n)$ von $\upsilon = (\upsilon_1, ..., \upsilon_n)$ durch θ^* generiert wurde, wobei θ^* hier den *wahren, aber unbekannten, Parameterwert* bezeichnet. Der wahre, aber unbekannten, Parameterwert θ^* bleibt auch nach der statistischen Analyse unbekannt. In der mathematischen Analyse von Inferenzmethoden betrachtet man alle möglichen wahren, aber unbekannten, Parameterwerte, schreibt also einfach $\{\mathbb{P}_\theta | \theta \in \Theta\}$.

Definition (Normalverteilungsmodell)

Das univariate parametrische Produktmodell

$$\mathcal{M} := (\mathcal{Y}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\}) \tag{2}$$

mit

$$\mathcal{Y} := \mathbb{R}^n, \mathcal{A} := \mathcal{B}(\mathbb{R}^n), \theta := (\mu, \sigma^2), \Theta := \mathbb{R} \times \mathbb{R}_{>0}, \tag{3}$$

also

$$\{\mathbb{P}_\theta | \theta \in \Theta\} := \left\{ \prod_{i=1}^n N(\mu, \sigma^2) | (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0} \right\}, \tag{4}$$

und damit

$$\upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2) \text{ mit } (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0} \tag{5}$$

heißt *Normalverteilungsmodell*.

Bemerkungen

- Das Normalverteilungsmodell ist Grundlage vieler populärer statistischen Verfahren.
- Diese Verfahren werden im Allgemeinen Linearen Modell integrativ betrachtet.
- Das Modell ist grundlegend durch normalverteilte Fehlerterme motiviert.

Definition (Bernoullimodell)

Das univariate parametrische Produktmodell

$$\mathcal{M} := (\mathcal{Y}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\}) \tag{6}$$

mit

$$\mathcal{Y} := \{0,1\}^n, \mathcal{A} := \mathcal{P}\left(\{0,1\}^n\right), \theta := \mu, \Theta :=]0,1[, \tag{7}$$

also

$$\{\mathbb{P}_\theta | \theta \in \Theta\} := \left\{ \prod_{i=1}^n \text{Bern}(\mu) | \mu \in]0,1[\right\}, \tag{8}$$

und damit

$$\upsilon_1, ..., \upsilon_n \sim \text{Bern}(\mu) \text{ mit } \mu \in]0,1[, \tag{9}$$

heißt *Bernoullimodell*.

Bemerkung

- Das Bernoullimodell spielt in der statistischen Anwendung eine eher untegeordnete Rolle.

Statistische Modelle

Statistiken und Schätzer

Standardprobleme Frequentistischer Inferenz

Selbstkontrollfragen

Definition (Statistik)

$\mathcal{M}$ sei ein statistisches Modell und $(\Sigma, \mathcal{S})$ sei ein Messraum. Dann wird eine Zufallsvariable der Form

$$S : \mathcal{Y} \to \Sigma \tag{10}$$

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.

Statistiken und Schätzer

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*

$$\bar{y}_n : \mathbb{R}^n \to \mathbb{R}, y \mapsto \bar{y}_n(y) := \frac{1}{n}\sum_{i=1}^n y_i, \tag{11}$$

- Die *Stichprobenvarianz*

$$s_n^2 : \mathbb{R}^n \to \mathbb{R}_{\geq 0}, y \mapsto s_n^2(y) := \frac{1}{n-1}\sum_{i=1}^n (y_i - \bar{y}_n(y))^2, \tag{12}$$

- Die *Stichprobenstandardabweichung*

$$s_n : \mathbb{R}^n \to \mathbb{R}_{\geq 0}, y \mapsto s_n(y) := \sqrt{s_n^2(y)}, \tag{13}$$

Definition (Schätzer)

$\mathcal{M}$ sei ein statistisches Modell, $(\Sigma, \mathcal{S})$ sei ein Messraum und $\tau : \Theta \to \Sigma$ sei eine Abbildung, die jedem $\theta \in \Theta$ eine Kenngröße $\tau(\theta) \in \Sigma$ zuordnet. Dann heißt eine Statistik

$$\hat{\tau} : \mathcal{Y} \to \Sigma \tag{14}$$

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 (10) Parameterschätzung.

Beispiele (Schätzer)

$\mathcal{M}$ sei das Normalverteilungsmodell.

- Dann ist zum Beispiel das Stichprobenmittel $\bar{y}_n : \mathbb{R}^n \to \mathbb{R}$ ein Schätzer für

$$\tau : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}, (\mu, \sigma^2) \mapsto \tau(\mu, \sigma^2) := \mu. \tag{15}$$

  Ebenso ist $\bar{y}_n$ ein Schätzer für

$$\tau : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}, (\mu, \sigma_2) \mapsto \tau(\mu, \sigma^2) := \mathbb{E}_{\mu,\sigma^2}(y_1). \tag{16}$$

- Weiterhin ist die konstante Funktion

$$\hat{\tau} : \mathbb{R}^n \to \mathbb{R}, y \mapsto \hat{\tau}(y) := 42 \tag{17}$$

  ein Schätzer für

$$\tau : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}_{>0}, (\mu, \sigma_2) \mapsto \tau(\mu, \sigma^2) := \sigma^2. \tag{18}$$

Dass eine Funktion $\hat{\tau} : \mathcal{Y} \to \Sigma$ ein Schätzer ist, heißt nicht, dass sie ein guter Schätzer ist!

Statistische Modelle

Statistiken und Schätzer

Standardprobleme Frequentistischer Inferenz

Selbstkontrollfragen

Standardprobleme Frequentistischer Inferenz

Mithilfe statistischer Modelle behandelt die Frequentistische Inferenz folgende Standardprobleme:

(1) Parameterschätzung

Ziel der Parameterschätzung ist es, einen möglichst guten Tipp für wahre, aber unbekannte, Parameterwerte oder Funktionen dieser abzugeben, typischerweise mithilfe der Daten.

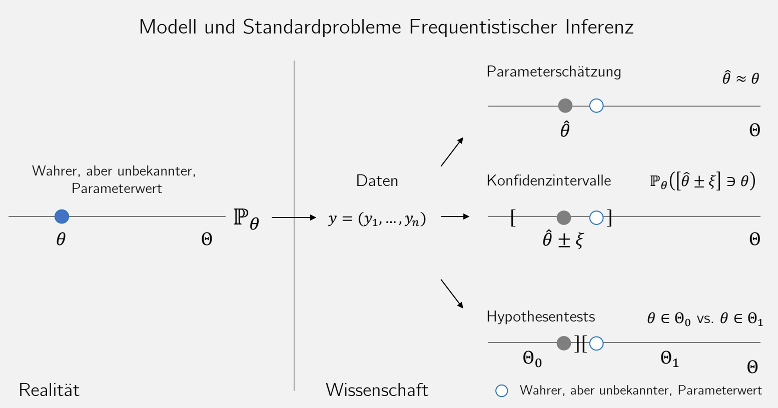
(2) Konfidenzintervalle

Ziel der Bestimmung von Konfidenzintervallen ist es, basierend auf der angenommenen Verteilung der Daten eine quantitative Aussage über die mit Schätzwerten assoziierte Unsicherheit zu treffen.

(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.

Standardprobleme Frequentistischer Inferenz



Standardannahmen Frequentistischer Inferenz

$\mathcal{M}$ sei ein statistisches Modell mit $\upsilon = (\upsilon_1, ..., \upsilon_n) \sim p_\theta$. Es wird angenommen, dass ein konkreter Datensatz $y = (y_1, ..., y_n) \in \mathbb{R}^n$ eine der möglichen Realisierungen von $\upsilon = (\upsilon_1, ..., \upsilon_n) \sim p_\theta$ ist.

Aus Frequentistischer Sicht kann man eine Studie unendlich oft wiederholen und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. das Stichprobenmittel:

Datensatz (1) : $y^{(1)} = \left(y_1^{(1)}, y_2^{(1)}, ..., y_n^{(1)}\right)$ mit $\bar{y}_n^{(1)} = \frac{1}{n}\sum_{i=1}^n y_i^{(1)}$

Datensatz (2) : $y^{(2)} = \left(y_1^{(2)}, y_2^{(2)}, ..., y_n^{(2)}\right)$ mit $\bar{y}_n^{(2)} = \frac{1}{n}\sum_{i=1}^n y_i^{(2)}$

Datensatz (3) : $y^{(3)} = \left(y_1^{(3)}, y_2^{(3)}, ..., y_n^{(3)}\right)$ mit $\bar{y}_n^{(3)} = \frac{1}{n}\sum_{i=1}^n y_i^{(3)}$

Datensatz (4) : $y^{(4)} = \left(y_1^{(4)}, y_2^{(4)}, ..., y_n^{(4)}\right)$ mit $\bar{y}_n^{(4)} = \frac{1}{n}\sum_{i=1}^n y_i^{(4)}$

Datensatz (5) : $y^{(5)} = ...$

Um die Qualität statistischer Methoden zu beurteilen betrachtet die Frequentistische Statistik deshalb die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken unter Annahme von $\upsilon = (\upsilon_1, ..., \upsilon_n) \sim p_\theta$. Was zum Beispiel ist die Verteilung der $\bar{y}_n^{(1)}$, $\bar{y}_n^{(2)}$, $\bar{y}_n^{(3)}$, $\bar{y}_n^{(4)}$, ... also die Verteilung der Zufallsvariable $\bar{\upsilon}_n$?

Wenn eine statistische Methode im Sinne der Frequentistischen Standardannahmen "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.

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Pre-BDI

n = 12

BDI-II Fragebogen

Name | Alter | Geschlecht m / w | Datum

Anleitung: Dieser Fragebogen enthält 21 Gruppen von Aussagen. Bitte lesen Sie jede dieser Gruppen von Aussagen sorgfältig durch und suchen Sie sich dann in jeder Gruppe **eine Aussage** heraus, die am besten beschreibt, wie Sie sich **in den letzten zwei Wochen, einschließlich heute, gefühlt haben.** Kreuzen Sie die Zahl neben der Aussage an, die Sie sich herausgesucht haben (0, 1, 2 oder 3). Falls in einer Gruppe mehrere Aussagen gleichermaßen auf Sie zutreffen, kreuzen Sie die Aussage mit der höheren Zahl an. Achten Sie bitte darauf, dass Sie in jeder Gruppe nicht mehr als eine Aussage ankreuzen, das gilt auch für Gruppe 16 (Veränderungen der Schlafgewohnheiten) oder Gruppe 18 (Veränderungen des Appetits).

1.) Traurigkeit
0 Ich bin nicht traurig.
1 Ich bin oft traurig.
2 Ich bin ständig traurig.
3 Ich bin so traurig oder unglücklich, dass ich es nicht aushalte.

2.) Pessimismus
0 Ich sehe nicht mutlos in die Zukunft.
1 Ich sehe mutloser in die Zukunft als sonst.
2 Ich bin mutlos und erwarte nicht, dass meine Situation besser wird.
3 Ich glaube, dass meine Zukunft hoffnungslos ist und nur noch schlechter wird.

3.) Versagensgefühle
0 Ich fühle mich nicht als Versager.
1 Ich habe häufiger Versagensgefühle.
2 Wenn ich zurückblicke, sehe ich eine Menge Fehlschläge.
3 Ich habe das Gefühl, als Mensch ein völliger Versager zu sein.

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.

5.) Schuldgefühle
0 Ich habe keine besonderen Schuldgefühle.
1 Ich habe oft Schuldgefühle wegen Dingen, die ich getan habe oder hätte tun sollen.
2 Ich habe die meiste Zeit Schuldgefühle.
3 Ich habe ständig Schuldgefühle.

6.) Bestrafungsgefühle
0 Ich habe nicht das Gefühl, für etwas bestraft zu sein.
1 Ich habe das Gefühl, vielleicht bestraft zu werden.
2 Ich erwarte, bestraft zu werden.
3 Ich habe das Gefühl, bestraft zu sein.

7.) Selbstablehnung
0 Ich halte von mir genauso viel wie immer.
1 Ich habe Vertrauen in mich verloren.
2 Ich bin von mir enttäuscht.
3 Ich lehne mich völlig ab.

8.) Selbstvorwürfe
0 Ich kritisiere oder tadle mich nicht mehr als sonst.
1 Ich bin mir gegenüber kritischer als sonst.
2 Ich kritisiere mich für all meine Mängel.
3 Ich gebe mir die Schuld für alles Schlimme, was passiert.

9.) Selbstmordgedanken
0 Ich denke nicht daran, mir etwas anzutun.
1 Ich denke manchmal an Selbstmord, aber ich würde es nicht tun.
2 Ich möchte mich am liebsten umbringen.
3 Ich würde mich umbringen, wenn ich die Gelegenheit dazu hätte.

10.) Weinen
0 Ich weine nicht öfter als früher.
1 Ich weine jetzt mehr als früher.
2 Ich weine beim geringsten Anlass.
3 Ich möchte gern weinen, aber ich kann nicht.

PEARSON PsychCorp
Summe Seite 1:

Post-BDI

⇒ Pre-Post BDI Score Reduktion

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

```
fname = file.path(getwd(), "9_Grundbegriffe_Frequentistischer_Inferenz.csv")
D     = read.table(fname, sep = ",", header = T)
```

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 y_i der $i$ten von n Patient:innen legen wir das Modell

$$\upsilon_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, ..., n \tag{19}$$

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

$$\upsilon = (\upsilon_1, ..., \upsilon_n) \sim N(\mu, \sigma^2). \tag{20}$$

Die Standardprobleme der Frequentistischen Inferenz führen in diesem Szenario auf folgende Fragen:

(1) Was sind sinnvolle Tipps für die wahren, aber unbekannten, Parameterwerte μ und σ^2?
(2) Wie hoch ist im Frequentistischen Sinn die mit diesen Tipps assoziierte Unsicherheit?
(3) Entscheiden wir uns sinnvollerweise für die Hypothese, dass 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., dass

$$\upsilon_i = f(\varepsilon_i) \text{ mit } f : \mathbb{R} \to \mathbb{R}, e_i \mapsto f(e_i) := e_i + \mu. \tag{21}$$

Dann gilt

$$\begin{aligned}
p_{\upsilon_i}(y_i) &= \frac{1}{|1|} p_{\varepsilon_i}\left(\frac{y_i - \mu}{1}\right) \\
&= N\left(y_i - \mu; 0, \sigma^2\right) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu - 0)^2\right) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right) \\
&= N(y_i; \mu, \sigma^2),
\end{aligned} \tag{22}$$

also $\upsilon_i \sim N(\mu, \sigma^2)$ u.i.v. und damit $\upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2)$.

Statistische Modelle

Statistiken und Schätzer

Standardprobleme Frequentistischer Inferenz

Selbstkontrollfragen

Selbstkontrollfragen

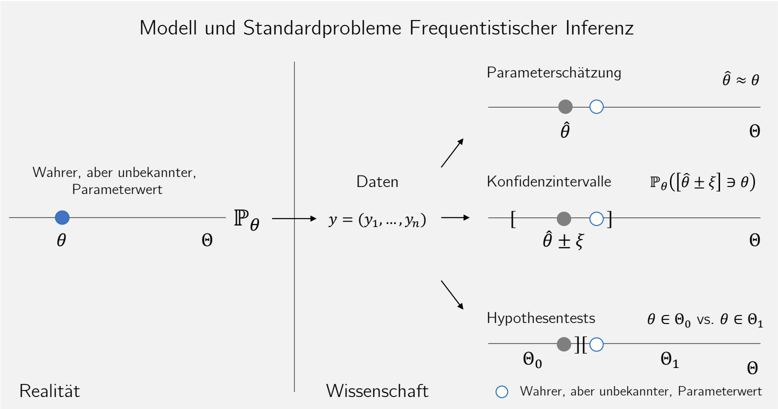
1. Definieren und erläutern Sie den Begriff des parametrischen statistischen Modells.
2. Definieren und erläutern Sie den Begriff eine parametrischen statistischen Produktmodells.
3. Erläutern Sie den Unterschied zwischen univariaten und multivariaten statistischen Modellen.
4. Formulieren und erläutern Sie das Normalverteilungsmodell.
5. Formulieren und erläutern Sie das Bernoullimodell.
6. Definieren und erläutern Sie den Begriff der Statistik.
7. Definieren und erläutern Sie den Begriff des Schätzers.
8. Nennen und erläutern Sie die Standardprobleme der Frequentistischen Inferenz.
9. Erläutern Sie die Standardannahmen der Frequentistischen Inferenz.

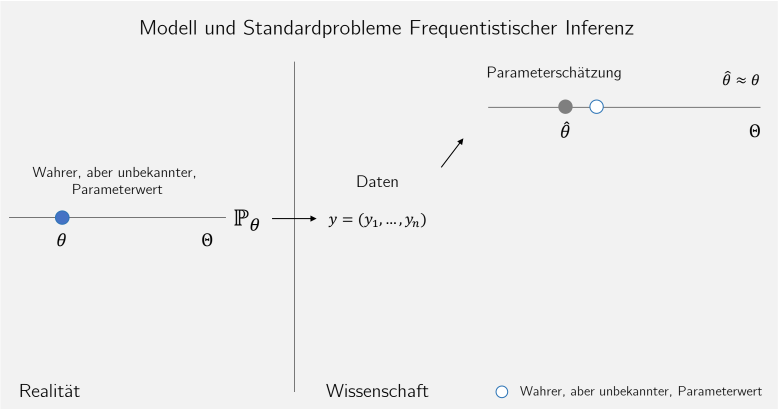
Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2022/23

Prof. Dr. Dirk Ostwald

(10) Parameterschätzung


Modell und Standardprobleme Frequentistischer Inferenz
Wahrer, aber unbekannter, Parameterwert
θ
Θ
$\mathbb{P}_\theta$
Daten
$y = (y_1, \dots, y_n)$
Parameterschätzung
$\hat{\theta} \approx \theta$
$\hat{\theta}$
Θ
Konfidenzintervalle
$\mathbb{P}_\theta([\hat{\theta} \pm \xi] \ni \theta)$
$\hat{\theta} \pm \xi$
Θ
Hypothesentests
$\theta \in \Theta_0$ vs. $\theta \in \Theta_1$
Θ_0
Θ_1
Θ
Realität
Wissenschaft
Wahrer, aber unbekannter, Parameterwert


Modell und Standardprobleme Frequentistischer Inferenz
Parameterschätzung
$\hat{\theta} \approx \theta$
$\hat{\theta}$
Θ
Wahrer, aber unbekannter, Parameterwert
Daten
θ
Θ
$\mathbb{P}_\theta$
$y = (y_1, \ldots, y_n)$
Realität
Wissenschaft
Wahrer, aber unbekannter, Parameterwert

Standardannahmen Frequentistischer Inferenz

$\mathcal{M}$ sei ein statistisches Modell mit Stichprobe $\upsilon_1, ..., \upsilon_n \sim p_\theta$. **Es wird angenommen, dass ein konkret vorliegender Datensatz $y = (y_1, ..., y_n) \in \mathbb{R}^n$ eine der möglichen Realisierungen von $\upsilon_1, ..., \upsilon_n \sim p_\theta$ ist.** Aus Frequentistischer Sicht kann man eine Studie unter gleichen Umständen unendlich oft wiederholen und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. das Stichprobenmittel:

Datensatz (1) : $y^{(1)} = \left(y_1^{(1)}, y_2^{(1)}, ..., y_n^{(1)}\right)$ mit $\bar{y}_n^{(1)} = \frac{1}{n}\sum_{i=1}^n y_i^{(1)}$

Datensatz (2) : $y^{(2)} = \left(y_1^{(2)}, y_2^{(2)}, ..., y_n^{(2)}\right)$ mit $\bar{y}_n^{(2)} = \frac{1}{n}\sum_{i=1}^n y_i^{(2)}$

Datensatz (3) : $y^{(3)} = \left(y_1^{(3)}, y_2^{(3)}, ..., y_n^{(3)}\right)$ mit $\bar{y}_n^{(3)} = \frac{1}{n}\sum_{i=1}^n y_i^{(3)}$

Datensatz (4) : $y^{(4)} = \left(y_1^{(4)}, y_2^{(4)}, ..., y_n^{(4)}\right)$ mit $\bar{y}_n^{(4)} = \frac{1}{n}\sum_{i=1}^n y_i^{(4)}$

Datensatz (5) : $y^{(5)} = ...$

Um die Qualität statistischer Methoden zu beurteilen betrachtet die Frequentistische Statistik deshalb die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken unter Annahme von $\upsilon_1, ..., \upsilon_n \sim p_\theta$. Was zum Beispiel ist die Verteilung der $\bar{y}_n^{(1)}$, $\bar{y}_n^{(2)}$, $\bar{y}_n^{(3)}$, $\bar{y}_n^{(4)}$, ... also die Verteilung der Zufallsvariable $\bar{\upsilon}_n$?

Wenn eine statistische Methode im Sinne der Frequentistischen Standardannahmen "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.

Grundbegriffe

Maximum-Likelihood Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood Schätzern

Selbstkontrollfragen

Grundbegriffe

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 statistisches Modell, $(\Theta, \mathcal{S})$ sei ein Messraum und $\hat{\theta} : \mathcal{Y} \to \Theta$ 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.

© 2023 Dirk Ostwald CC BY 4.0

Prinzipien zur Gewinnung von Parameterschätzern

Die Definition eines Parameterschätzers macht keine Aussage darüber, wie man Parameterschätzer findet. Zur Gewinnung von Parameterschätzern in statistischen Modellen 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*.

Grundbegriffe

Maximum-Likelihood Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood Schätzern

Selbstkontrollfragen

Maximum-Likelihood Schätzer

Definition (Likelihood-Funktion und Log-Likelihood-Funktion)

$\mathcal{M}$ sei ein parametrisches statistisches Produktmodell mit WMF oder WDF p_θ, also $\upsilon_1, ..., \upsilon_n \sim p_\theta$. Dann ist die *Likelihood Funktion* definiert als

$$L_n : \Theta \to [0, \infty[, \theta \mapsto L_n(\theta) := \prod_{i=1}^n p_\theta(y_i) \tag{1}$$

und die *Log-Likelihood-Funktion* ist definiert als

$$\ell_n : \Theta \to \mathbb{R}, \theta \mapsto \ell_n(\theta) := \ln L_n(\theta). \tag{2}$$

Bemerkungen

- L_n ist eine Funktion des Parameters eines statistischen Modells.
- Werte von L_n sind die gemeinsamen Wahrscheinlichkeitsmassen bzw. -dichten von Datenwerten $y_1, ..., y_n$.
- Generell gibt es keinen Grund anzunehmen, dass L_n über Θ zu 1 integriert.
- Die Likelihood-Funktion ist also keine WMF oder WDF.
- Die Log-Likelihood-Funktion ist die logarithmierte Likelihood-Funktion.

Definition (Maximum-Likelihood Schätzer)

$\mathcal{M}$ sei ein parametrisches statistisches Produktmodell mit Parameter $\theta \in \Theta$. Ein *Maximum-Likelihood Schätzer* von θ ist definiert als

$$\hat{\theta}_n^{\mathsf{ML}} : \mathcal{Y} \to \Theta, y \mapsto \hat{\theta}_n^{\mathsf{ML}}(y) := \arg\max_{\theta \in \Theta} L_n(\theta) = \arg\max_{\theta \in \Theta} \ell_n(\theta). \tag{3}$$

Bemerkungen

- $L_n(\theta) := \prod_{i=1}^n p_\theta(y_i)$ hängt von $y := (y_1, ..., y_n)$ ab, also hängt auch $\hat{\theta}_n^{\mathsf{ML}}(y)$ von y ab.
- Weil $\ln$ monoton steigend ist, entspricht eine Maximumstelle von ℓ_n einer Maximumstelle von L_n.
- Das Arbeiten mit der Log-Likelihood-Funktion ist oft einfacher als mit der Likelihood Funktion.
- Multiplikation von L_n mit einer 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

Maximum-Likelihood Schätzer

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.

Maximum-Likelihood Schätzer

Beispiel (Bernoullimodell)

$\mathcal{M}$ sei das Bernoullimodell, also $\upsilon_1, ..., \upsilon_n \sim \text{Bern}(\mu)$.

(1) Formulierung der Log-Likelihood-Funktion

Es gilt

$$L_n :]0,1[\to]0,1[, \mu \mapsto L_n(\mu) := \prod_{i=1}^n \mu^{y_i}(1-\mu)^{1-y_i} = \mu^{\sum_{i=1}^n y_i}(1-\mu)^{n-\sum_{i=1}^n y_i}. \quad (4)$$

Logarithmieren ergibt

$$\ell_n :]0,1[\to \mathbb{R}, \mu \mapsto \ell_n(\mu) = \ln\mu \sum_{i=1}^n y_i + \ln(1-\mu)\left(n - \sum_{i=1}^n y_i\right). \quad (5)$$

Beispiel (Bernoullimodell)

(2) Auswertung der Ableitung der Log-Likelihood-Funktion und Nullsetzen

Es gilt

$$
\begin{aligned}
\frac{d}{d\mu}\ell_n(\mu) &= \frac{d}{d\mu}\left(\ln\mu\sum_{i=1}^n y_i + \ln(1-\mu)\left(n-\sum_{i=1}^n y_i\right)\right) \\
&= \frac{d}{d\mu}\ln\mu\sum_{i=1}^n y_i + \frac{d}{d\mu}\ln(1-\mu)\left(n-\sum_{i=1}^n y_i\right) \\
&= \frac{1}{\mu}\sum_{i=1}^n y_i - \frac{1}{1-\mu}\left(n-\sum_{i=1}^n y_i\right).
\end{aligned} \tag{6}
$$

Die sogenannte *Maximum-Likelihood Gleichung* ergibt sich in diesem Beispiel also zu

$$
\frac{1}{\hat{\mu}_n^{\mathsf{ML}}}\sum_{i=1}^n y_i - \frac{1}{1-\hat{\mu}_n^{\mathsf{ML}}}\left(n-\sum_{i=1}^n y_i\right) = 0. \tag{7}
$$

Beispiel (Bernoullimodell)

(3) Auflösen nach potentiellen Maximumstellen

Es gilt

$$
\begin{aligned}
\frac{1}{\hat{\mu}_n^{\mathsf{ML}}} \sum_{i=1}^n y_i - \frac{1}{1-\hat{\mu}_n^{\mathsf{ML}}} \left(n - \sum_{i=1}^n y_i\right) &= 0 \\
\Leftrightarrow \hat{\mu}_n^{\mathsf{ML}}(1-\hat{\mu}_n^{\mathsf{ML}}) \left(\frac{1}{\hat{\mu}_n^{\mathsf{ML}}} \sum_{i=1}^n y_i - \frac{1}{1-\hat{\mu}_n^{\mathsf{ML}}} \left(n - \sum_{i=1}^n y_i\right)\right) &= 0 \\
\Leftrightarrow \sum_{i=1}^n y_i - \hat{\mu}_n^{\mathsf{ML}} \sum_{i=1}^n y_i - n\hat{\mu}_n^{\mathsf{ML}} + \hat{\mu}_n^{\mathsf{ML}} \sum_{i=1}^n y_i &= 0 \qquad (8) \\
\Leftrightarrow n\hat{\mu}_n^{\mathsf{ML}} &= \sum_{i=1}^n y_i \\
\Leftrightarrow \hat{\mu}_n^{\mathsf{ML}} &= \frac{1}{n} \sum_{i=1}^n y_i .
\end{aligned}
$$

Beispiel (Bernoullimodell)

$\hat{\mu}_n^{\mathsf{ML}} = \frac{1}{n}\sum_{i=1}^n y_i$ ist also ein potentieller Maximum-Likelihood Schätzer von μ. Dies kann durch Betrachten der zweiten Ableitung von ℓ_n verifiziert werden, worauf hier verzichtet werden soll.

$$\hat{\mu}_n^{\mathsf{ML}} : \{0,1\}^n \to [0,1], y \mapsto \hat{\mu}_n^{\mathsf{ML}}(y) := \frac{1}{n}\sum_{i=1}^n y_i \qquad (9)$$

ist also ein Maximum-Likelihood Schätzer von μ im Bernoullimodell.

© 2023 Dirk Ostwald CC BY 4.0

Maximum-Likelihood Schätzer

Beispiel (Normalverteilungsmodell)

$\mathcal{M}$ sei das Normalverteilungsmodell, also $\upsilon_1, ..., \upsilon_n \sim N\left(\mu, \sigma^2\right)$

(1) Formulierung der Log-Likelihood-Funktion

Es gilt

$$
\begin{aligned}
L_n : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}_{>0}, (\mu, \sigma^2) \mapsto L_n(\mu, \sigma^2) &:= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right) \\
&= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right).
\end{aligned} \tag{10}
$$

Logarithmieren ergibt

$$
\ell_n : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}, (\mu, \sigma^2) \mapsto \ell_n(\mu, \sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2. \tag{11}
$$

Beispiel (Normalverteilungsmodell)

(2) Auswertung der Ableitung der Log-Likelihood-Funktion und Nullsetzen

Es ergibt sich

$$\frac{d}{d\mu}\ell_n(\mu) = -\frac{d}{d\mu}\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\mu)^2 = -\frac{1}{2\sigma^2}\sum_{i=1}^n \frac{d}{d\mu}(y_i-\mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^n (y_i-\mu). \tag{12}$$

Weiterhin ergibt sich

$$\frac{d}{d\sigma^2}\ell_n(\sigma^2) = -\frac{n}{2}\frac{d}{d\sigma^2}\ln\sigma^2 - \frac{d}{d\sigma^2}\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\mu)^2 = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n (y_i-\mu)^2. \tag{13}$$

Die Maximum-Likelihood Gleichungen haben also die Form

$$\begin{aligned} \sum_{i=1}^n (y_i - \hat{\mu}_n^{\mathsf{ML}}) &= 0 \\ -\frac{n}{2\hat{\sigma}_n^{2\mathsf{ML}}} + \frac{1}{2\hat{\sigma}_n^{4\mathsf{ML}}}\sum_{i=1}^n (y_i-\mu)^2 &= 0 \end{aligned} \tag{14}$$

Beispiel (Normalverteilungsmodell)

(3) Auflösen nach potentiellen Maximumstellen

Es ergibt sich

$$\sum_{i=1}^{n}(y_i - \hat{\mu}_n^{\mathsf{ML}}) = 0 \Leftrightarrow \sum_{i=1}^{n} y_i = n\hat{\mu}_n^{\mathsf{ML}} \Leftrightarrow \hat{\mu}_n^{\mathsf{ML}} = \frac{1}{n}\sum_{i=1}^{n} y_i. \tag{15}$$

Also ist $\hat{\mu}_n^{\mathsf{ML}} = n^{-1}\sum_{i=1}^{n} y_i$ ein potentieller Maximum-Likelihood Schätzer von μ. Einsetzen ergibt dann

$$\begin{aligned} -\frac{n}{2\hat{\sigma}_n^{2\,\mathsf{ML}}} + \frac{1}{2\hat{\sigma}_n^{4\,\mathsf{ML}}}\sum_{i=1}^{n}(y_i - \hat{\mu}_n^{\mathsf{ML}})^2 &= 0 \\ \Leftrightarrow -n\hat{\sigma}_n^{2\,\mathsf{ML}} + \sum_{i=1}^{n}(y_i - \hat{\mu}_n^{\mathsf{ML}})^2 &= 0 \\ \Leftrightarrow \hat{\sigma}_n^{2\,\mathsf{ML}} &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu}_n^{\mathsf{ML}})^2. \end{aligned} \tag{16}$$

$\hat{\sigma}_n^{2\,\mathsf{ML}} = n^{-1}\sum_{i=1}^{n}(y_i - \hat{\mu}_n^{\mathsf{ML}})^2$ ist also ein potentieller Maximum-Likelihood Schätzer von σ^2.

© 2023 Dirk Ostwald CC BY 4.0

Maximum-Likelihood Schätzer

Beispiel (Normalverteilungsmodell)

Beide potentiellen Maximum-Likelihood Schätzer können durch Betrachten der zweiten Ableitung von ℓ_n verifiziert werden, worauf hier verzichtet werden soll.

Also sind

$$\hat{\mu}_n^{\text{ML}} : \mathbb{R}^n \to \mathbb{R}, y \mapsto \hat{\mu}_n^{\text{ML}}(y) := \frac{1}{n}\sum_{i=1}^n y_i \tag{17}$$

und

$$\hat{\sigma}_n^{2^{\text{ML}}} : \mathbb{R}^n \to \mathbb{R}_{\geq 0}, y \mapsto \hat{\sigma}_n^{2^{\text{ML}}}(y) := \frac{1}{n}\sum_{i=1}^n \left(y_i - \hat{\mu}_n^{\text{ML}}\right)^2 . \tag{18}$$

die Maximum-Likelihood Schätzer von μ und σ^2 im Normalverteilungsmodell. $\hat{\mu}_n^{\text{ML}}$ ist identisch mit dem Stichprobenmittel $\bar{\upsilon}_n$, $\hat{\sigma}_n^{2^{\text{ML}}}$ ist dagegen nicht identisch mit der Stichprobenvarianz S_n^2.

© 2023 Dirk Ostwald CC BY 4.0

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Pre-BDI


BDI-II Fragebogen

Name | Alter | Geschlecht m / w | Datum

Anleitung: Dieser Fragebogen enthält 21 Gruppen von Aussagen. Bitte lesen Sie jede dieser Gruppen von Aussagen sorgfältig durch und suchen Sie sich dann in jeder Gruppe **eine Aussage** heraus, die am besten beschreibt, wie Sie sich **in den letzten zwei Wochen, einschließlich heute, gefühlt haben**. Kreuzen Sie die Zahl neben der Aussage an, die Sie sich herausgesucht haben (0, 1, 2 oder 3). Falls in einer Gruppe mehrere Aussagen gleichermaßen auf Sie zutreffen, kreuzen Sie die Aussage mit der höheren Zahl an. Achten Sie bitte darauf, dass Sie in jeder Gruppe nicht mehr als eine Aussage ankreuzen, das gilt auch für Gruppe 16 (Veränderungen der Schlafgewohnheiten) oder Gruppe 18 (Veränderungen des Appetits).

1.) Traurigkeit
- **0** Ich bin nicht traurig.
- **1** Ich bin oft traurig.
- **2** Ich bin ständig traurig.
- **3** Ich bin so traurig oder unglücklich, dass ich es nicht aushalte.

2.) Pessimismus
- **0** Ich sehe nicht mutlos in die Zukunft.
- **1** Ich sehe mutloser in die Zukunft als sonst.
- **2** Ich bin mutlos und erwarte nicht, dass meine Situation besser wird.
- **3** Ich glaube, dass meine Zukunft hoffnungslos ist und nur noch schlechter wird.

3.) Versagensgefühle
- **0** Ich fühle mich nicht als Versager.
- **1** Ich habe häufiger Versagensgefühle.
- **2** Wenn ich zurückblicke, sehe ich eine Menge Fehlschläge.
- **3** Ich habe das Gefühl, als Mensch ein völliger Versager zu sein.

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.

5.) Schuldgefühle
- **0** Ich habe keine besonderen Schuldgefühle.
- **1** Ich habe oft Schuldgefühle wegen Dingen, die ich getan habe oder hätte tun sollen.
- **2** Ich habe die meiste Zeit Schuldgefühle.
- **3** Ich habe ständig Schuldgefühle.

6.) Bestrafungsgefühle
- **0** Ich habe nicht das Gefühl, für etwas bestraft zu sein.
- **1** Ich habe das Gefühl, vielleicht bestraft zu werden.
- **2** Ich erwarte, bestraft zu werden.
- **3** Ich habe das Gefühl, bestraft zu sein.

7.) Selbstablehnung
- **0** Ich halte von mir genauso viel wie immer.
- **1** Ich habe Vertrauen in mich verloren.
- **2** Ich bin von mir enttäuscht.
- **3** Ich lehne mich völlig ab.

8.) Selbstvorwürfe
- **0** Ich kritisiere oder tadle mich nicht mehr als sonst.
- **1** Ich bin mir gegenüber kritischer als sonst.
- **2** Ich kritisiere mich für all meine Mängel.
- **3** Ich gebe mir die Schuld für alles Schlimme, was passiert.

9.) Selbstmordgedanken
- **0** Ich denke nicht daran, mir etwas anzutun.
- **1** Ich denke manchmal an Selbstmord, aber ich würde es nicht tun.
- **2** Ich möchte mich am liebsten umbringen.
- **3** Ich würde mich umbringen, wenn ich die Gelegenheit dazu hätte.

10.) Weinen
- **0** Ich weine nicht öfter als früher.
- **1** Ich weine jetzt mehr als früher.
- **2** Ich weine beim geringsten Anlass.
- **3** Ich möchte gern weinen, aber ich kann nicht.

PEARSON PsychCorp

Summe Seite 1:

Bitte wenden

Post-BDI

⇒ Pre-Post BDI Score Reduktion

Maximum-Likelihood Schätzer

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

```
fname = file.path(getwd(), "10_Parameterschätzung.csv")
D     = read.table(fname, sep = ",", header = T)
```

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

Maximum-Likelihood Schätzer

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Für die Pre-Post BDI Score Reduktion υ_i der $i$ten von n Patient:innen legen wir das Modell

$$\upsilon_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, ..., n \tag{19}$$

zugrunde. Dabei wird die Pre-Post BDI Reduktion υ_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

$$\upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2). \tag{20}$$

Die Standardprobleme der Frequentistischen Inferenz führen in diesem Szenario auf folgende Fragen:

(1) Was sind sinnvolle Tipps für die wahren, aber unbekannten, Parameterwerte μ und σ^2?
(2) Wie hoch ist im frequentistischen Sinn die mit diesen Tipps assoziierte Unsicherheit?
(3) Entscheiden wir uns sinnvollerweise für die Hypothese, dass gilt $\mu \neq 0$?

Maximum-Likelihood Schätzer

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

```
# Einlesen und Auswahl der Daten
fname = file.path(getwd(), "10_Parameterschätzung.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] 3.17
```

```
# 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] 12.6
```

Basierend auf dem Prinzip der Maximum-Likelihood Schätzung sind also

$$\hat{\mu}_{12}^{\text{ML}} = 3.17, \text{ und } \hat{\sigma}_{12}^{2^{\text{ML}}} = 12.6 \tag{21}$$

sinnvolle Tipps für μ und σ^2 basierend auf den vorliegenden 12 Datenpunkten.

Grundbegriffe

Maximum-Likelihood Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood Schätzern

Selbstkontrollfragen

Vorbemerkungen zu Frequentistischen Schätzereigenschaften

Wir gehen von einem statistischem 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. $\upsilon := \upsilon_1, ..., \upsilon_n \sim p_\theta$ bezeichnet die zu $\mathcal{M}$ gehörende Stichprobe unabhängig und identisch verteilter Zufallsvariablen, es gilt also $\upsilon_1 \sim p_\theta$ und $\upsilon_i \sim p_\theta$ für alle $i = 1, ..., n$.

Für einen Messraum (Σ, S) sei $\hat{\tau} : \mathcal{Y} \to \Sigma$ ein Schätzer von $\tau : \Theta \to \Sigma$. Wir betrachten Erwartungswerts-Varianz-, und Standardabweichungsschätzer, also Schätzer für

$$\tau : \Theta \to \Sigma,\ \theta \mapsto \tau(\theta) \text{ mit } \tau(\theta) := \mathbb{E}_\theta(\upsilon_1), \tau(\theta) := \mathbb{V}_\theta(\upsilon_1),\ \text{und } \tau(\theta) := \mathbb{S}_\theta(\upsilon_1) \tag{22}$$

respektive, sowie Parameterschätzer, also Schätzer für

$$\tau : \Theta \to \Sigma, \tau(\theta) := \theta. \tag{23}$$

In der Folge führen wir *Frequentistische Schätzereigenschaften* ein. Frequentistische Schätzereigenschaften betrachten die Verteilung der Schätzwerte $\hat{\tau}(y_1, ..., y_n)$ in Abhängigkeit von der Verteilung der Datennwerte $\upsilon := \upsilon_1, ..., \upsilon_n$. Weil die Stichprobenwerte zufällig sind, sind auch die Schätzwerte zufällig; ein Schätzer $\hat{\tau}$ ist also wie oben gesehen eine Zufallsvariable. Wir unterscheiden zwischen *Schätzereigenschaften bei endlichen Stichproben*, d.h. Eigenschaften von $\hat{\tau}_n$ für ein fixes $n \in \mathbb{N}$ (z.B. $n = 12$) und *Asymptotischen Schätzereigenschaften*, d.h. Eigenschaften von $\hat{\tau}_n$ für unendlich groß werdende Stichproben mit $n \to \infty$.

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 $\hat{\tau}_n$ heißt *erwartungstreu*, wenn sein Erwartungswert dem wahren, aber unbekannten, Wert $\tau(\theta)$ für alle $\theta \in \Theta$ gleicht. Die *Varianz* eines Schätzers $\hat{\tau}_n$ ist die Varianz der Zufallsvariable $\hat{\tau}_n(\upsilon)$. Der *Standardfehler* eines Schätzers $\hat{\tau}_n$ ist die Standardabweichung der Zufallsvariable $\hat{\tau}_n(\upsilon)$.

Für folgende Schätzereigenschaften bei endlichen Stichproben verweisen wir auf den Appendix:

(1) Mittlerer Quadratischer Fehler
(2) Cramér-Rao Ungleichung

Intuitiv haben diese folgende Bedeutungen: Der *mittlere quadratische Fehler* von $\hat{\tau}_n$ ist der Erwartungswert der quadrierten Abweichung von $\hat{\tau}_n(\upsilon)$ 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)

$\upsilon := \upsilon_1, ..., \upsilon_n \sim p_\theta$ sei eine Stichprobe und $\hat{\tau}_n$ sei ein Schätzer für τ.

- Der *Fehler* von $\hat{\tau}_n$ ist definiert als
$$\hat{\tau}_n(\upsilon) - \tau(\theta). \tag{24}$$
- Der *systematische Fehler (Bias)* von $\hat{\tau}_n$ ist definiert als
$$\mathrm{B}(\hat{\tau}_n) := \mathbb{E}_\theta(\hat{\tau}_n(\upsilon)) - \tau(\theta). \tag{25}$$
- $\hat{\tau}_n$ heißt *erwartungstreu (unbiased)*, wenn
$$\mathrm{B}(\hat{\tau}_n) = 0 \Leftrightarrow \mathbb{E}_\theta(\hat{\tau}_n(\upsilon)) = \tau(\theta) \text{ für alle } \theta \in \Theta, n \in \mathbb{N}. \tag{26}$$
Andernfalls heißt $\hat{\tau}_n$ *verzerrt (biased)*.

Bemerkungen

- Der Fehler hängt von einer Realisation der Stichprobe ab.
- Der systematische Fehler ist der erwartete Fehler über viele Stichprobenrealisationen.
- Ein Parameterschätzer ist erwartungstreu, wenn $\mathbb{E}_\theta(\hat{\theta}_n(\upsilon)) = \theta$.

Theorem (Erwartungstreue von Stichprobenmittel und Stichprobenvarianz)

$\upsilon_1, ..., \upsilon_n \sim p_\theta$ sei die Stichprobe eines statistischen parametrischen Produktmodells $\mathcal{M}$.

- Das Stichprobenmittel

$$\bar{\upsilon}_n := \frac{1}{n}\sum_{i=1}^{n} \upsilon_i \tag{27}$$

  ist ein erwartungstreuer Schätzer des Erwartungswerts $\mathbb{E}_\theta(\upsilon_1)$.

- Die Stichprobenvarianz

$$S_n^2 := \frac{1}{n-1}\sum_{i=1}^{n} (\upsilon_i - \bar{\upsilon}_n)^2 \tag{28}$$

  ist ein erwartungstreuer Schätzer der Varianz $\mathbb{V}_\theta(\upsilon_1)$.

Beweis

Um die Notation zu vereinfachen, definieren wir $\mathbb{E} := \mathbb{E}_\theta$ und $\mathbb{V} := \mathbb{V}_\theta$. Mit der Linearität von Erwartungswerten ergibt sich dann

$$\mathbb{E}(\bar{\upsilon}_n) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n \upsilon_i\right) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(\upsilon_i\right) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(\upsilon_1\right) = \frac{1}{n} n\mathbb{E}\left(\upsilon_1\right) = \mathbb{E}\left(\upsilon_1\right).$$

Dies zeigt die Erwartungstreue des Stichprobenmittels als Schätzer des Erwartungswertes.

Um die Erwartungstreue der Stichprobenvarianz zu zeigen, halten wir zunächst fest, dass

$$\mathbb{V}(\bar{\upsilon}_n) = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^n \upsilon_i\right) = \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}\left(\upsilon_i\right) = \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}\left(\upsilon_1\right) = \frac{1}{n^2} n\mathbb{V}\left(\upsilon_1\right) = \frac{\mathbb{V}\left(\upsilon_1\right)}{n}.$$

Weiterhin gilt wie im Folgenden gezeigt mit $\mu := \mathbb{E}(\upsilon_1)$

$$\sum_{i=1}^n \left(\upsilon_i - \bar{\upsilon}_n\right)^2 = \sum_{i=1}^n (\upsilon_i - \mu)^2 - n(\bar{\upsilon}_n - \mu)^2.$$

$$
\begin{aligned}
\sum_{i=1}^{n} (\upsilon_i - \bar{\upsilon}_n)^2 &= \sum_{i=1}^{n} (\upsilon_i - \mu - \bar{\upsilon}_n + \mu)^2 \\
&= \sum_{i=1}^{n} ((\upsilon_i - \mu) - (\bar{\upsilon}_n - \mu))^2 \\
&= \sum_{i=1}^{n} (\upsilon_i - \mu)^2 - 2(\bar{\upsilon}_n - \mu) \left(\sum_{i=1}^{n} (\upsilon_i - \mu) \right) + \sum_{i=1}^{n} (\bar{\upsilon}_n - \mu)^2 \\
&= \sum_{i=1}^{n} (\upsilon_i - \mu)^2 - 2(\bar{\upsilon}_n - \mu) \left(\sum_{i=1}^{n} \upsilon_i - n\mu \right) + n(\bar{\upsilon}_n - \mu)^2 \qquad (29) \\
&= \sum_{i=1}^{n} (\upsilon_i - \mu)^2 - 2(\bar{\upsilon}_n - \mu) \left(n \left(\frac{1}{n} \sum_{i=1}^{n} \upsilon_i \right) - n\mu \right) + n(\bar{\upsilon}_n - \mu)^2 \\
&= \sum_{i=1}^{n} (\upsilon_i - \mu)^2 - 2n(\bar{\upsilon}_n - \mu)^2 + n(\bar{\upsilon}_n - \mu)^2 \\
&= \sum_{i=1}^{n} (\upsilon_i - \mu)^2 - n(\bar{\upsilon}_n - \mu)^2 .
\end{aligned}
$$

Es ergibt sich also

$$
\begin{aligned}
\mathbb{E}\left((n-1)S_n^2\right) &= \mathbb{E}\left(\sum_{i=1}^n (\upsilon_i - \bar{\upsilon}_n)^2\right) \\
&= \mathbb{E}\left(\sum_{i=1}^n (\upsilon_i - \mu)^2 - n(\bar{\upsilon}_n - \mu)^2\right) \\
&= \sum_{i=1}^n \mathbb{E}\left((\upsilon_i - \mu)^2\right) - n\mathbb{E}\left((\bar{\upsilon}_n - \mu)^2\right) \\
&= n\mathbb{V}(\upsilon_1) - n\mathbb{V}(\bar{\upsilon}_n) \\
&= n\mathbb{V}(\upsilon_1) - n\frac{\mathbb{V}(\upsilon_1)}{n} \\
&= n\mathbb{V}(\upsilon_1) - \mathbb{V}(\upsilon_1) \\
&= (n-1)\mathbb{V}(\upsilon_1)
\end{aligned}
$$

Schließlich ergibt sich

$$
\mathbb{E}(S_n^2) = \mathbb{E}\left(\frac{1}{n-1}(n-1)S_n^2\right) = \frac{1}{n-1}\mathbb{E}\left((n-1)S_n^2\right) = \frac{1}{n-1}(n-1)\mathbb{V}(\upsilon_1) = \mathbb{V}(\upsilon_1)
$$

und damit die Erwartungstreue der Stichprobenvarianz als Schätzer der Varianz.

Theorem (Verzerrtheit der Stichprobenstandardabweichung)

$\upsilon_1, ..., \upsilon_n \sim p_\theta$ sei die Stichprobe eines statistischen parametrischen Produktmodells $\mathcal{M}$. Dann ist die Stichprobenstandardabweichung

$$S_n := \sqrt{S_n^2} \tag{30}$$

ein verzerrter Schätzer der Standardabweichung $\mathbb{S}_\theta(\upsilon_1)$.

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

$$\mathbb{E}(S_n) = \mathbb{E}\left(\sqrt{S_n^2}\right) < \sqrt{\mathbb{E}(S_n^2)} = \sqrt{\mathbb{V}_\theta(\upsilon_1)} = \mathbb{S}_\theta(\upsilon_1). \tag{31}$$

□

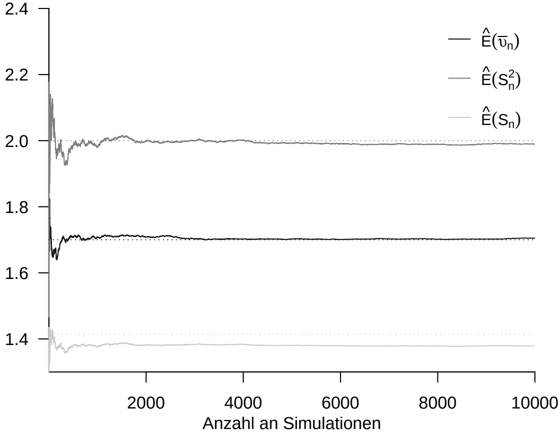
Bemerkung

- Nichtlineare Transformationen von erwartungstreuen Schätzern liefern oft verzerrte Schätzer.

Simulation von $\upsilon_1, ..., \upsilon_{12} \sim N(\mu, \sigma^2)$ 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    = 1e4                    # 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,...,\ups_{12}
    y           = rnorm(n,mu,sqrt(sigsqr))

    # Erwartungswert-, Varianz-, Standardabweichungschaetzer
    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}_n) 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
```

Simulation von $\upsilon_1, ..., \upsilon_{12} \sim N(\mu, \sigma^2)$ mit $n = 12$, $\mu = 1.7$, $\sigma^2 = 2$, $\sigma \approx 1.41$



Definition (Varianz und Standardfehler)

$\upsilon := \upsilon_1, ..., \upsilon_n \sim p_\theta$ sei eine Stichprobe und $\hat{\tau}_n$ sei ein Schätzer von τ.

- Die *Varianz* von $\hat{\tau}_n$ ist definiert als

$$\mathbb{V}_\theta(\hat{\tau}_n) := \mathbb{E}_\theta\left((\hat{\tau}_n(\upsilon) - \mathbb{E}_\theta(\hat{\tau}_n(\upsilon)))^2\right). \tag{32}$$

- Der *Standardfehler* von $\hat{\tau}_n$ ist definiert als

$$\mathsf{SE}(\hat{\tau}_n) := \sqrt{\mathbb{V}_\theta(\hat{\tau}_n)} \tag{33}$$

Bemerkungen

- Die Varianz eines Schätzers $\hat{\tau}_n$ ist die Varianz der Zufallsvariable $\hat{\tau}_n(\upsilon)$.
- Der Standardfehler eines Schätzers $\hat{\tau}_n$ ist die Standardabweichung von $\hat{\tau}_n(\upsilon)$.

© 2023 Dirk Ostwald CC BY 4.0

Theorem (Standardfehler des Stichprobenmittels)

$\upsilon := \upsilon_1, ..., \upsilon_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells. Dann ist der *Standardfehler des Stichprobenmittels* gegeben durch

$$\mathsf{SE}(\bar{\upsilon}_n) = \frac{\mathbb{S}_\theta(\upsilon_1)}{\sqrt{n}}. \tag{34}$$

Der Standardfehler des Stichprobenmittels heißt auch *Standardfehler des Mittelwertes*.

Beweis

Per definitionem und mit $\mathbb{V}_\theta(\bar{\upsilon}_n) = \mathbb{V}_\theta(\upsilon_1)/n$, ergibt sich

$$\mathsf{SE}(\bar{\upsilon}_n) = \sqrt{\mathbb{V}_\theta(\bar{\upsilon}_n)} = \sqrt{\frac{\mathbb{V}_\theta(\upsilon_1)}{n}} = \frac{\mathbb{S}_\theta(\upsilon_1)}{\sqrt{n}}. \tag{35}$$

Bemerkungen

- Der Standardfehler des Mittelwerts beschreibt die Variabilität des Stichprobenmittels.
- Da $\mathbb{S}_\theta(\upsilon_1)$ unbekannt ist, ist auch $\mathsf{SE}(\bar{\upsilon}_n)$ unbekannt.
- Ein verzerrter Schätzer für den Standardfehler des Stichprobenmittels ist gegeben durch $\hat{\mathsf{SE}}(\bar{\upsilon}_n) = \frac{s_n}{\sqrt{n}}$.

Beispiel (Standardfehler des Bernoulli Parameter Maximum-Likelihood Schätzes)

Es sei $\upsilon := \upsilon_1, ..., \upsilon_n \sim \text{Bern}(\mu)$ und $\hat{\mu}_n^{\text{ML}}$ der Maximum-Likelihood Schätzer für μ. Dann ist

$$\text{SE}\left(\hat{\mu}_n^{\text{ML}}\right) = \sqrt{\frac{\mu(1-\mu)}{n}}. \tag{36}$$

Beweis

Es gilt

$$\begin{aligned} \text{SE}\left(\hat{\mu}_n^{\text{ML}}\right) = \sqrt{\mathbb{V}_\mu\left(\hat{\mu}_n^{\text{ML}}\right)} &= \sqrt{\mathbb{V}_\mu\left(\frac{1}{n}\sum_{i=1}^n \upsilon_i\right)} \\ &= \sqrt{\frac{1}{n^2}\sum_{i=1}^n \mathbb{V}_\mu(\upsilon_i)} = \sqrt{\frac{n\mu(1-\mu)}{n^2}} = \sqrt{\frac{\mu(1-\mu)}{n}}, \end{aligned} \tag{37}$$

wobei die dritte Gleichung mit der Unabhängigkeit der $\upsilon_i, i = 1, ..., n$ und die vierte Gleichung mit der Varianz $\mathbb{V}_\mu(\upsilon_1) = \mathbb{V}_\mu(\upsilon_i) = \mu(1-\mu), i = 1, ..., n$ der Bernoulli Stichprobenvariablen folgt.

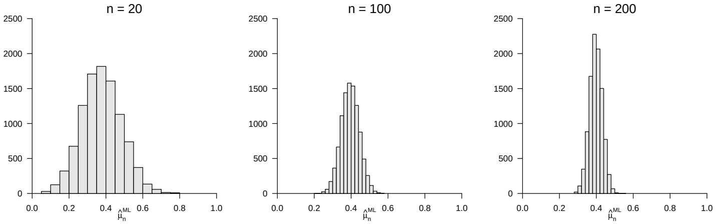
Bemerkung

- Ein Schätzer für den Standardfehler $\text{SE}\left(\hat{\mu}_n^{\text{ML}}\right)$ ist $\hat{\text{SE}}\left(\hat{\mu}_n^{\text{ML}}\right) = \sqrt{\frac{\hat{\mu}_n^{\text{ML}}(1-\hat{\mu}_n^{\text{ML}})}{n}}$

Simulation von $\upsilon_1, ..., \upsilon_n \sim \text{Bern}(\mu)$ mit $\mu = 0.4$

```
# Modellformulierung
mu          = 0.4                                      # wahrer, aber unbekannter, Parameterwert
n_all       = c(20,100,200)                            # Stichprobengrößen n
ns          = 1e4                                      # Anzahl der Simulationen
mu_hat_ML   = matrix(rep(NaN, length(n_all)*ns),       # Maximum-Likelihood Schätzearray
                     nrow = length(n_all))

# Stichprobengroesseniterationen
for(i in seq_along(n_all)){

    # Simulationsiterationen
    for(s in 1:ns){
        y              = rbinom(n_all[i],1,mu)         # Stichprobenrealisation von y_1,...,y_n
        mu_hat_ML[i,s] = mean(y)                       # Stichprobenmittel
    }
}
```

Die Varianz bzw. der Standardfehler von $\hat{\mu}_n^{\text{ML}}$ hängen von n ab.



Grundbegriffe

Maximum-Likelihood Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood Schätzern

Selbstkontrollfragen

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 Stichprobengrößen 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

Intuitiv haben diese folgende Bedeutungen: Ein Schätzer $\hat{\tau}_n$ für τ heißt *asymptotisch erwartungstreu*, wenn der Erwartungswert von $\hat{\tau}_n$ für große Stichprobengrößen $n \to \infty$ gleich dem wahren, aber unbekannten, Wert $\tau(\theta)$ ist. Ein Schätzer $\hat{\tau}_n$ für τ heißt *konsistent*, wenn für große Stichprobengrößen $n \to \infty$ die Wahrscheinlichkeit dafür, dass $\hat{\tau}_n(\upsilon)$ vom wahren, aber unbekannten, Wert $\tau(\theta)$ abweicht beliebig klein wird. Ein Schätzer $\hat{\tau}_n$ für τ heißt *asymptotisch normalverteilt*, wenn für große Stichprobengrößen $n \to \infty$, die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung gegeben ist.

Für folgende asymptotische Schätzereigenschaften verweisen wir auf den Appendix:

(1) Asymptotische Effizienz

Intuitiv hat diese die folgende Bedeutung: Ein Schätzer $\hat{\tau}_n$ für τ heißt *asymptotisch effizient*, wenn für große Stichprobengrößen $n \to \infty$ 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)

$\upsilon := \upsilon_1, ..., \upsilon_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\tau}_n$ sei ein Schätzer für τ. $\hat{\tau}_n$ heißt *asymptotisch erwartungstreu*, wenn

$$\lim_{n \to \infty} \mathbb{E}_\theta(\hat{\tau}_n(\upsilon)) = \tau(\theta) \text{ für alle } \theta \in \Theta. \tag{38}$$

Bemerkungen

- Asymptotisch erwartungstreue Schätzer sind für “unendlich große” Stichproben erwartungstreu.
- Erwartungstreue Schätzer sind immer auch asymptotisch erwartungstreu.

Theorem (Asymptotische Erwartungstreue des Varianzparameterschätzers)

Es sei $\upsilon := \upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2)$. Dann ist der Maximum-Likelihood Schätzer des Varianzparameters,

$$\hat{\sigma}_n^{2^{\mathrm{ML}}} := \frac{1}{n}\sum_{i=1}^n (\upsilon_i - \bar{\upsilon}_n)^2 \tag{39}$$

asymptotisch erwartungstreu.

Beweis

Mit der Erwartungstreue der Stichprobenvarianz ergibt sich

$$\mathbb{E}_{\mu,\sigma^2}\left(\hat{\sigma}_n^{2^{\mathrm{ML}}}\right) = \mathbb{E}_{\mu,\sigma^2}\left(\frac{1}{n}\sum_{i=1}^n \left(\upsilon_i - \bar{\upsilon}_n\right)^2\right) = \frac{1}{n}\mathbb{E}_{\mu,\sigma^2}\left(\sum_{i=1}^n \left(\upsilon_i - \bar{\upsilon}_n\right)^2\right) = \frac{n-1}{n}\sigma^2$$

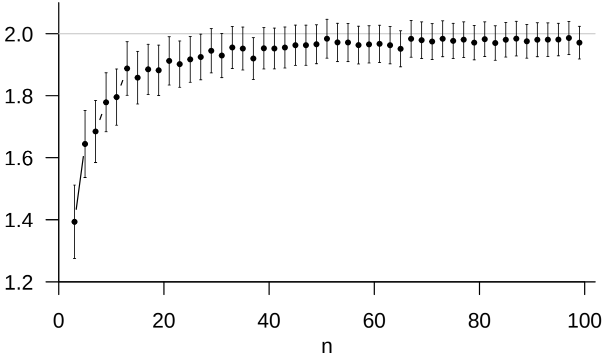
Also gilt $\mathbb{E}_{\mu,\sigma^2}\left(\hat{\sigma}_n^{2^{\mathrm{ML}}}\right) \neq \sigma^2$. $\hat{\sigma}_n^{2^{\mathrm{ML}}}$ ist also ein verzerrter Schätzer von σ^2. Allerdings gilt $(n-1)/n \to 1$ für $n \to \infty$, so dass

$$\lim_{n\to\infty} \mathbb{E}_{\mu,\sigma^2}\left(\hat{\sigma}_n^{2^{\mathrm{ML}}}\right) = \lim_{n\to\infty} \frac{n-1}{n}\sigma^2 = \sigma^2 \lim_{n\to\infty} \frac{n-1}{n} = \sigma^2. \tag{40}$$

Simulation von $\upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2)$ mit $\mu = 1$, $\sigma^2 = 2$

```
# Modellformulierung
mu        = 1                      # wahrer, aber unbekannter, Erwartungswertparameter
sigsqr    = 2                      # wahrer, aber unbekannter, Varianzparameter
n         = seq(1,100, by = 2)     # Stichprobengroessen
ns        = 1e3                    # Anzahl Simulation pro Stichprobengroesse
sigsqr_ml = matrix(                # \hat{\sigma^2}^{ML} Array
            rep(NaN, length(n)*ns),
            ncol = length(n))

# Stichprobengroesseniterationen
for(i in seq_along(n)){

    # Simulationsiterationen
    for(s in 1:ns){

        # Stichprobenrealisation
        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
```

Simulation $\upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2)$ mit $\mu = 1$, $\sigma^2 = 2$



Definition (Konsistenz)

$\upsilon := \upsilon_1, ..., \upsilon_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\tau}_n$ sei ein Schätzer von τ. Eine Folge von Schätzern $\hat{\tau}_1, \hat{\tau}_2, ...$ wird dann eine *konsistente Folge von Schätzern* genannt, wenn für jedes $\epsilon > 0$ und jedes $\theta \in \Theta$ gilt, dass

$$\lim_{n \to \infty} \mathbb{P}_\theta \left(|\hat{\tau}_n(\upsilon) - \tau(\theta)| \geq \epsilon \right) = 0.$$

Wenn $\hat{\tau}_1, \hat{\tau}_2, ...$ eine konsistente Folge von Schätzern ist, dann heißt $\hat{\tau}_n$ *konsistenter Schätzer*.

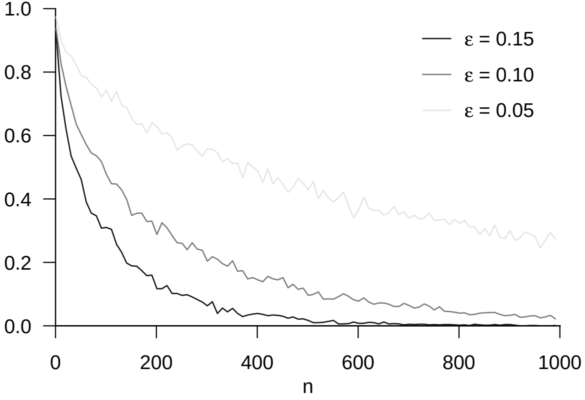
Bemerkungen

- Für $n \to \infty$ wird die Wahrscheinlichkeit, dass $\hat{\tau}_n(\upsilon)$ beliebig nah bei $\tau(\theta)$ liegt, groß.
- Für $n \to \infty$ wird die Wahrscheinlichkeit, dass $\hat{\tau}_n(\upsilon)$ von $\tau(\theta)$ abweicht, klein.
- Diese Eigenschaften gelten für alle möglichen wahren, aber unbekannten, Parameterwerte.
- Die Konvergenz ist *Konvergenz in Wahrscheinlichkeit*.
- Konsistenz von Schätzern kann direkt oder mit Kriterien nachgewiesen werden.

Simulation der Konsistenz von $\bar{\upsilon}_n$ bei $\upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2)$ mit $\mu = 1$, $\sigma^2 = 2$

```
# Modellformulierung
mu      = 1                                             # w.a.u \mu Wert
sigsqr  = 2                                             # w.a.u. \sigma^2 Wert
n       = seq(1,1e3,by = 10)                            # Stichprobengroesse n
eps     = c(0.15, 0.10, 0.05)                           # \epsilon Werte
ne      = length(eps)                                   # Anzahl \epsilon Werte
nn      = length(n)                                     # Anzahl Stichprobengroessen
ns      = 1e3                                           # Anzahl Simulationen
E       = array(rep(NaN,nn,ne,ns),                      # Ereignisindikatorarray
                dim = c(nn,ne,ns))

# Simulation
for(e in seq_along(eps)){                               # \epsilon Iterationen
    for(i in seq_along(n)){                             # n Iterationen
        for(s in 1:ns){                                 # Simulationsiterationen

            # Stichprobenrealisationen
            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)
```

Simulation der Konsistenz von $\bar{\upsilon}_n$ bei $\upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2)$ mit $\mu = 1$, $\sigma^2 = 2$



Definition (Asymptotische Normalität)

$\upsilon_1, ..., \upsilon_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\theta}_n$ sei ein Parameterschätzer für θ. Weiterhin sei $\tilde{\theta} \sim N(\mu, \sigma^2)$ eine normalverteilte Zufallsvariable mit Erwartungswertparameter μ und Varianzparameter σ^2. Wenn $\hat{\theta}_n$ in Verteilung gegen $\tilde{\theta}$ konvergiert, dann heißt $\hat{\theta}_n$ *asymptotisch normalverteilt* und wir schreiben

$$\hat{\theta}_n \overset{a}{\sim} N(\mu, \sigma^2). \tag{41}$$

Bemerkung

- Konvergenz in Verteilung heißt $\lim_{n\to\infty} P_n(\hat{\theta}_n) = P(\tilde{\theta})$.

Grundbegriffe

Maximum-Likelihood Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood Schätzern

Selbstkontrollfragen

Theorem (Eigenschaften von Maximum-Likelihood Schätzern)

$\upsilon_1, ..., \upsilon_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ 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) konsistent,

(3) asymptotisch normalverteilt und

(4) asymptotisch erwartungstreu ist.

Bemerkungen

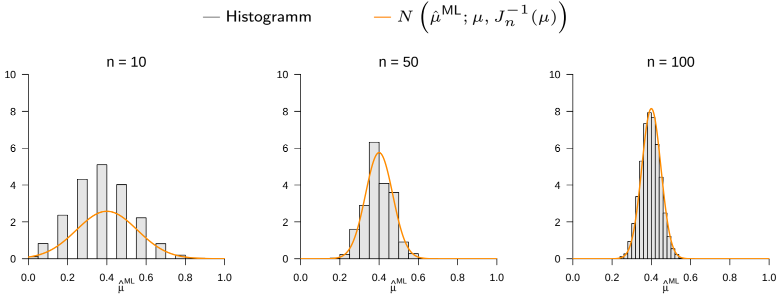
- Maximum-Likelihood Schätzer sind überdies asymptotisch effizient.
- Für einen Beweis verweisen wir auf Held and Sabanés Bové (2014), Abschnitt 3.4.

Eigenschaften von Maximum-Likelihood Schätzern

Simulation der asymptotische Normalverteilung des Maximum-Likelihood 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(NaN,
                        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(NaN, length(n_all)*mu_hat_ML_r),  # ML WDF Array
                     nrow = length(n_all))

# 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]))
}
```

Simulation der asymptotische Normalverteilung des Maximum-Likelihood Bernoulliparameterschätzers



Grundbegriffe

Maximum-Likelihood Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood Schätzern

Selbstkontrollfragen

Selbstkontrollfragen

1. Definieren und erläutern Sie den Begriff des Parameterpunktschätzers.
2. Definieren Sie den Begriff der Likelihood-Funktion.
3. Definieren Sie den Begriff der Log-Likelihood-Funktion.
4. Definieren Sie den Begriff des Maximum-Likelihood Schätzes.
5. Erläutern Sie das Vorgehen zur Maximum-Likelihood Schätzung für ein parametrisches Produktmodell.
6. Geben Sie den Maximum-Likelihood Schätzer für den Parameter μ des Bernoullimodells an.
7. Geben Sie den Maximum-Likelihood Schätzer für den Parameter μ des Normalverteilungmodels an.
8. Geben Sie den Maximum-Likelihood Schätzer für den Parameter σ^2 des Normalverteilungmodels an.
9. Definieren und erläutern Sie den Begriff der Erwartungstreue eines Schätzers.
10. Definieren Sie die Begriffe der Varianz und des Standardfehlers eines Schätzers.
11. Erläutern Sie den Begriff des asymptotischen Erwartungstreue eines Schätzers
12. Erläutern Sie den Begriff der Konsistenz eines Schätzers.
13. Erläutern Sie den Begriff der asymptotischen Normalität eines Schätzers.
14. Nennen Sie vier Eigenschaften eines Maximum-Likelihood Schätzers.

Appendix

Appendix | Mittlerer quadratischer Fehler

Definition (Mittlerer quadratischer Fehler)

$\upsilon := \upsilon_1, ..., \upsilon_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\tau}_n$ ein Schätzer für τ. Dann ist der *mittlere quadratischer Fehler (engl. mean squared error)* von $\hat{\tau}_n$ definiert als

$$\text{MQF}(\hat{\tau}_n) := \mathbb{E}_\theta \left((\hat{\tau}_n(\upsilon) - \tau(\theta))^2 \right). \tag{42}$$

Bemerkungen

- Der MQF von $\hat{\tau}_n$ ist die erwartete quadrierte Abweichung von $\hat{\tau}_n(\upsilon)$ von $\tau(\theta)$.
- Die Varianz von $\hat{\tau}_n$ ist die erwartete quadrierte Abweichung von $\hat{\tau}_n$ von $\mathbb{E}_\theta(\hat{\tau}_n(\upsilon))$.
- $\mathbb{E}_\theta(\hat{\tau}_n(\upsilon))$ kann mit $\tau(\theta)$ übereinstimmen, muss es aber nicht.

Theorem (Zerlegung des mittleren quadratischen Fehlers)

$\upsilon := \upsilon_1, ..., \upsilon_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$, $\hat{\tau}_n$ sei ein Schätzer für τ, und $\mathrm{MQF}(\hat{\tau}_n)$ sei der mittlere quadratische Fehler von $\hat{\tau}_n$. Dann gilt

$$\mathrm{MQF}(\hat{\tau}_n) = \mathrm{B}(\hat{\tau}_n)^2 + \mathbb{V}_\theta(\hat{\tau}_n). \tag{43}$$

Bemerkungen

- $\mathrm{MQF} = \mathrm{Bias}^2 + \mathrm{Varianz}$.
- Der MQF kann als Bias-Varianz Abwägungskriterium benutzt werden.
- Kleine Schätzerverzerrungen können gegenüber einer großen Schätzervarianz präferiert werden.

Beweis

Zur Vereinfachung der Notation seien $\tau := \tau(\theta)$, $\hat{\tau}_n := \hat{\tau}_n(\upsilon)$ und $\bar{\tau}_n := \mathbb{E}_\theta(\hat{\tau}_n(\upsilon))$. Dann gilt:

$$
\begin{aligned}
\mathbb{E}_\theta\left((\hat{\tau}_n - \tau)^2\right) &= \mathbb{E}_\theta\left((\hat{\tau}_n - \bar{\tau}_n + \bar{\tau}_n - \tau)^2\right) \\
&= \mathbb{E}_\theta\left((\hat{\tau}_n - \bar{\tau}_n)^2 + 2(\hat{\tau}_n - \bar{\tau}_n)(\bar{\tau}_n - \tau) + (\bar{\tau}_n - \tau)^2\right) \\
&= \mathbb{E}_\theta\left((\hat{\tau}_n - \bar{\tau}_n)^2\right) + 2\mathbb{E}_\theta\left((\hat{\tau}_n - \bar{\tau}_n)(\bar{\tau}_n - \tau)\right) + \mathbb{E}_\theta\left((\bar{\tau}_n - \tau)^2\right) \\
&= \mathbb{E}_\theta\left((\hat{\tau}_n - \bar{\tau}_n)^2\right) + 2\mathbb{E}_\theta\left(\hat{\tau}_n\bar{\tau}_n - \hat{\tau}_n\tau - \bar{\tau}_n\bar{\tau}_n + \bar{\tau}_n\tau\right) + \mathbb{E}_\theta\left((\bar{\tau}_n - \tau)^2\right) \\
&= \mathbb{E}_\theta\left((\hat{\tau}_n - \bar{\tau}_n)^2\right) + 2\left(\bar{\tau}_n\bar{\tau}_n - \bar{\tau}_n\tau\right) + \mathbb{E}_\theta\left((\bar{\tau}_n - \tau)^2\right) \\
&= \mathbb{E}_\theta\left((\hat{\tau}_n - \bar{\tau}_n)^2\right) + 0 + \mathbb{E}_\theta\left((\bar{\tau}_n - \tau)^2\right) \\
&= \mathbb{E}_\theta\left((\bar{\tau}_n - \tau)^2\right) + \mathbb{E}_\theta\left((\hat{\tau}_n - \bar{\tau}_n)^2\right) \\
&= \mathbb{E}_\theta\left((\mathbb{E}_\theta(\hat{\tau}_n) - \tau)^2\right) + \mathbb{E}_\theta\left((\hat{\tau}_n - \mathbb{E}_\theta(\hat{\tau}_n))^2\right) \\
&= (\mathbb{E}_\theta(\hat{\tau}_n) - \tau)^2 + \mathbb{V}_\theta(\hat{\tau}_n) \\
&= \mathrm{B}(\hat{\tau}_n)^2 + \mathbb{V}_\theta(\hat{\tau}_n).
\end{aligned}
$$

Appendix | Cramér-Rao-Ungleichung

Vorbemerkungen zur Cramér-Rao-Ungleichung

Je kleiner die Varianz eines Schätzers, desto besser. Weil aber Stichproben streuen, kann die Varianz von erwartungstreuen Schätzern nicht beliebig klein sein.

Die **Cramér-Rao-Ungleichung** gibt eine untere Schranke für die Varianz erwartungstreuer Schätzer an. Ein erwartungstreuer Schätzer mit Varianz gleich der in der Cramér-Rao-Ungleichung gegebenen unteren Schranke hat die kleinstmögliche Varianz aller erwartungstreuer Schätzer und ist in diesem Sinne "optimal".

Die Cramér-Rao-Ungleichung basiert auf dem Begriff der **Fisher-Information**. Wir diskutieren deshalb zunächst die Begriffe der **Scorefunktion** und der darauf basierenden **Fisher-Information**.

Die vorgestellten Resultate gelten im Allgemeinen nur unter eine Reihe von Annahmen, den sogenannten **Fisher-Regularitätsbedingungen**:

- Θ ist ein offenes Intervall, d.h. θ liegt nicht an einer Parameterraumgrenze.
- Der Träger von p_θ hängt nicht von θ ab.
- WMFs oder WDF mit unterschiedlichem $\theta \in \Theta$ sind unterschiedlich.
- Die Likelihood-Funktion ist zweimal stetig differenzierbar.
- Integration und Differentiation dürfen vertauscht werden.

Definition (Scorefunktion und Fisher-Information)

$\upsilon := \upsilon_1, ..., \upsilon_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ mit eindimensionalem Parameter θ und ℓ_n sei die zugehörige Log-Likelihood-Funktion.

- Die erste Ableitung der Log-Likelihood-Funktion ℓ_n wird *Scorefunktion der Stichprobe* genannt und mit
$$S_n(\theta) := \frac{d}{d\theta}\ell_n(\theta). \tag{44}$$
bezeichnet. Für $n = 1$ schreiben wir $S(\theta) := S_1(\theta)$ und nennen $S(\theta)$ *Scorefunktion einer Zufallsvariable*.

- Die negative zweite Ableitung der Log-Likelihood-Funktion ℓ_n wird *Fisher-Information der Stichprobe* genannt und mit
$$I_n(\theta) := -\frac{d^2}{d\theta^2}\ell_n(\theta). \tag{45}$$
bezeichnet. Für $n = 1$ schreiben wir $I(\theta) := I_1(\theta)$ und nennen $I(\theta)$ die *Fisher-Information einer Zufallsvariable*.

Definition (Erwartete und beobachtete Fisher-Information)

$\upsilon := \upsilon_1, ..., \upsilon_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ mit eindimensionalem Parameter θ, ℓ_n sei die zugehörige Log-Likelihood-Funktion und $\hat{\theta}_n^{\text{ML}}$ sei ein ML-Schätzer von θ.

- Die *beobachtete Fisher-Information der Stichprobe* ist definiert als

$$I_n\left(\hat{\theta}_n^{\text{ML}}\right) := -\frac{d^2}{d\theta^2}\ell_n\left(\hat{\theta}_n^{\text{ML}}\right), \tag{46}$$

  d.h. die beobachtete Fisher-Information der Stichprobe ist die Fisher-Information an der Stelle des ML-Schätzers $\hat{\theta}_n^{\text{ML}}$.

- Die *erwartete Fisher-Information der Stichprobe* ist definiert als

$$J_n(\theta) := \mathbb{E}_\theta(I_n(\theta)). \tag{47}$$

  Für $n = 1$ schreiben wir $J(\theta) := J_1(\theta)$ und nennen $J(\theta)$ die *erwartete Fisher-Information einer Zufallsvariable.*

Theorem (Additivität der Fisher-Information)

$\upsilon := \upsilon_1, ..., \upsilon_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ mit eindimensionalem Parameter θ, ℓ_n sei die zugehörige Log-Likelihood-Funktion, und $I_n(\theta)$ und $J_n(\theta)$ seien die Fisher-Information und die erwartete Fisher-Information der Stichprobe, respektive. Dann gilt

$$I_n(\theta) = nI_1(\theta) \text{ und } J_n(\theta) = nJ_1(\theta). \tag{48}$$

Bemerkungen

- Um $I_n(\theta)$ oder $J_n(\theta)$ zu berechnen, genügt es also $I(\theta)$ oder $J(\theta)$ zu berechnen.
- Die Additivität der beobachteten Fisher-Information ist in der Additivität von $I_n(\theta)$ implizit.

© 2023 Dirk Ostwald CC BY 4.0

Beweis

Wir zeigen das Resultat für die erwartete Fisher-Information, das Resultat für die Fisher-Information ist dann implizit. Per definitionem und mit der Linearität von Ableitungen und Erwartungswerte gilt

$$
\begin{aligned}
J_n(\theta) &= \mathbb{E}_\theta\left(-\frac{d^2}{d\theta^2}\ell_n(\theta)\right) \\
&= \mathbb{E}_\theta\left(-\frac{d^2}{d\theta^2}\ln\left(\prod_{i=1}^n p_\theta(\upsilon_i)\right)\right) \\
&= \mathbb{E}_\theta\left(-\frac{d^2}{d\theta^2}\sum_{i=1}^n \ln p_\theta(\upsilon_i)\right) \\
&= \mathbb{E}_\theta\left(-\frac{d^2}{d\theta^2}\sum_{i=1}^n \ln p_\theta(y_1)\right) \\
&= \mathbb{E}_\theta\left(-\frac{d^2}{d\theta^2}\ell_1(\theta)n\right) \\
&= n\mathbb{E}_\theta\left(-\frac{d^2}{d\theta^2}\ell_1(\theta))\right) \\
&= nJ_n(\theta).
\end{aligned}
\tag{49}
$$

Theorem (Erwartungswert und Varianz der Scorefunktion)

Der Erwartungswert der Scorefunktion einer Zufallsvariable ist

$$\mathbb{E}_\theta(S(\theta)) = 0 \tag{50}$$

und die Varianz der Scorefunktion einer Zufallsvariable ist

$$\mathbb{V}_\theta(S(\theta)) = J(\theta). \tag{51}$$

Bemerkungen

- Der Erwartungswert der Ableitung der Log-Likelihood-Funktion ist Null.
- Die erwartete Fisher-Information ist gleich der Varianz der Scorefunktion.

Beweis

Wir betrachten nur den Fall, dass p_θ eine WDF ist und zeigen zunächst, dass $\mathbb{E}_\theta(S(\theta)) = 0$ ist:

$$
\begin{aligned}
\mathbb{E}_\theta(S(\theta)) &= \int S(\theta) p_\theta(x)\, dx \\
&= \int \frac{d}{d\theta} \ell(\theta) p_\theta(x)\, dx \\
&= \int \frac{d}{d\theta} \ln L(\theta) p_\theta(x)\, dx \\
&= \int \frac{1}{L(\theta)} \frac{d}{d\theta} L(\theta) p_\theta(x)\, dx \\
&= \int \frac{1}{p_\theta(x)} \frac{d}{d\theta} L(\theta) p_\theta(x)\, dx \\
&= \int \frac{d}{d\theta} L(\theta)\, dx \\
&= \frac{d}{d\theta} \int p_\theta(x)\, dx = \frac{d}{d\theta} 1 = 0.
\end{aligned}
\tag{52}
$$

Beweis (fortgeführt)

Mit der Definition der Varianz folgt dann sofort, dass $\mathbb{V}_\theta(S(\theta)) = \mathbb{E}_\theta(S(\theta)^2)$ ist. Als nächstes zeigen wir, dass $J(\theta) = \mathbb{E}_\theta(S(\theta)^2)$ und deshalb $\mathbb{V}_\theta(S(\theta)) = J(\theta)$ ist:

$$
\begin{aligned}
J(\theta) &= \mathbb{E}_\theta\left(-\frac{d^2}{d\theta^2}\ln L(\theta)\right) \\
&= \mathbb{E}_\theta\left(-\frac{d}{d\theta}\frac{\frac{d}{d\theta}L(\theta)}{L(\theta)}\right) \\
&= \mathbb{E}_\theta\left(-\frac{\frac{d^2}{d\theta^2}L(\theta)L(\theta) - \frac{d}{d\theta}L(\theta)\frac{d}{d\theta}L(\theta)}{L(\theta)L(\theta)}\right) \\
&= -\mathbb{E}_\theta\left(\frac{\frac{d^2}{d\theta^2}L(\theta)}{L(\theta)}\right) + \mathbb{E}_\theta\left(\frac{\left(\frac{d}{d\theta}L(\theta)\right)^2}{(L(\theta))^2}\right) \\
&= -\int \frac{\frac{d^2}{d\theta^2}L(\theta)}{L(\theta)}p_\theta(x)\,dx + \int \frac{\left(\frac{d}{d\theta}L(\theta)\right)^2}{(L(\theta))^2}p_\theta(x)\,dx \\
&= -\frac{d^2}{d\theta^2}\int p_\theta(x)\,dx + \int\left(\frac{1}{L(\theta)}\frac{d}{d\theta}L(\theta)\right)^2 p_\theta(x)\,dx \\
&= -\frac{d^2}{d\theta^2}1 + \int\left(\frac{d}{d\theta}\ln L(\theta)\right)^2 p_\theta(x)\,dx = \mathbb{E}_\theta\left(S(\theta)^2\right).
\end{aligned}
\tag{53}
$$

Theorem (Fisher-Information bei Bernoulli-Verteilung)

Es sei $\upsilon := \upsilon_1, ..., \upsilon_n \sim \text{Bern}(\mu)$ mit $\mu \in]0,1[$. Dann gilt:

- Die Scorefunktion der Stichprobe ist

$$S_n :]0,1[\to \mathbb{R}, \mu \mapsto S_n(\mu) := \frac{1}{\mu}\sum_{i=1}^n y_i - \frac{1}{1-\mu}\left(n - \sum_{i=1}^n y_i\right). \tag{54}$$

- Die Fisher-Information der Stichprobe ist

$$I_n :]0,1[\to \mathbb{R}, \mu \mapsto I_n(\mu) := I_n(\mu) = \frac{ny}{\mu^2} + \frac{n(1-y)^2}{1-\mu}. \tag{55}$$

- Die beobachtete Fisher-Information der Stichprobe ist

$$I_n :]0,1[\to \mathbb{R}, \hat{\mu}_n^{\text{ML}} \mapsto I_n\left(\hat{\mu}_n^{\text{ML}}\right) := \frac{ny}{\hat{\mu}_n^{\text{ML}^2}} + \frac{n(1-y)}{1-\hat{\mu}_n^{\text{ML}}}. \tag{56}$$

- Die erwartete Fisher-Information der Stichprobe ist

$$J_n :]0,1[\to \mathbb{R}, \mu \mapsto J_n(\mu) := \frac{n}{\mu(1-\mu)}. \tag{57}$$

Beweis

Die Scorefunktion wurde bereits im Kontext der Maximum-Likelihood-Schätzung von μ hergeleitet. Wir betrachten die Fisher-Information einer einzelnen Bernoulli Zufallsvariable υ:

$$
\begin{aligned}
I(\mu) &:= -\frac{d^2}{d\mu^2}\ell_1(\mu) \\
&= -\frac{d^2}{d\mu^2}\ln p_\mu(y) \\
&= -\frac{d^2}{d\mu^2}\left(y\ln\mu + (1-y)\ln(1-\mu)\right) \\
&= -\frac{d}{d\mu}\left(\frac{d}{d\mu}\left(y\ln\mu + (1-y)\ln(1-\mu)\right)\right) \\
&= -\frac{d}{d\mu}\left(\frac{y}{\mu} + \frac{(1-y)}{1-\mu}\right) \\
&= -\left(-\frac{y}{\mu^2} - \frac{(1-y)}{1-\mu}^2\right) \\
&= \frac{y}{\mu^2} + \frac{(1-y)}{1-\mu}^2 .
\end{aligned}
\tag{58}
$$

Beweis (fortgeführt)

Damit ergibt sich die erwartete Fisher-Information der Zufallsvariable υ als

$$
\begin{aligned}
J(\mu) &= \mathbb{E}_\mu(I(\mu)) \\
&= \mathbb{E}_\mu\left(\frac{\upsilon}{\mu^2} + \frac{(1-\upsilon)^2}{1-\mu}\right) \\
&= \frac{\mathbb{E}_\mu(\upsilon)}{\mu^2} + \frac{(1-\mathbb{E}_\mu(\upsilon))^2}{1-\mu} \\
&= \frac{\mu}{\mu^2} + \frac{(1-\mu)^2}{1-\mu} \\
&= \frac{1}{\mu(1-\mu)}.
\end{aligned} \tag{59}
$$

Mit der Additivitätseigenschaft der Fisher-Information und der Definition der beobachteten Fisher-Information ergibt sich dann sofort

$$
I_n(\mu) = \frac{ny}{\mu^2} + \frac{n(1-y)^2}{1-\mu} \text{ und } J_n(\mu) = \frac{n}{\mu(1-\mu)}. \tag{60}
$$

Theorem (Fisher-Information bei Normalverteilung I)

Es sei $\upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2)$ und σ^2 als bekannt vorausgesetzt. Dann gilt:

- Die Scorefunktion der Stichprobe ist

$$S_n : \mathbb{R} \to \mathbb{R}, \mu \mapsto S_n(\mu) := \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \mu). \tag{61}$$

- Die Fisher-Information der Stichprobe ist

$$I_n : \mathbb{R} \to \mathbb{R}, \mu \mapsto I_n(\mu) := \frac{n}{\sigma^2}. \tag{62}$$

- Die beobachtete Fisher-Information der Stichprobe ist

$$I_n(\hat{\mu}_n^{\text{ML}}) = \frac{n}{\sigma^2}. \tag{63}$$

- Die erwartete Fisher-Information der Stichprobe ist

$$J_n : \mathbb{R} \to \mathbb{R}, \mu \mapsto J_n(\mu) := \frac{n}{\sigma^2}. \tag{64}$$

Beweis

Wir erinnern uns, dass die Log-Likelihood-Funktion der Stichprobe $v_1, ..., v_n \sim N(\mu, \sigma^2)$ bei bekanntem Varianzparameter σ^2 durch

$$\ell_n : \mathbb{R} \to \mathbb{R}, \mu \mapsto \ell_n(\mu) := -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2. \tag{65}$$

gegeben ist. Damit ergibt sich die Scorefunktion als

$$s_n(\mu) = \frac{d}{d\mu} \ell_n(\mu) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \mu) \tag{66}$$

Die Fisher-Information der Stichprobe ergibt sich als

$$I_n(\mu) = -\frac{d^2}{d\mu^2} \ell_n(\mu) = -\frac{d}{d\mu} s_n(\mu) = -\frac{1}{\sigma^2} \frac{d}{d\mu} \left(\sum_{i=1}^{n} y_i - n\mu \right) = \frac{n}{\sigma^2}. \tag{67}$$

Die beobachtete Fisher-Information ist die Fisher-Information an der Stelle des Maximum-Likelihood Schätzes $\hat{\mu}_n^{\mathrm{ML}}$. Die erwartete Fisher-Information schließlich ergibt sich als

$$J_n(\mu) = \mathbb{E}_\mu(I_n(\mu)) = \mathbb{E}_\mu \left(\frac{n}{\sigma^2} \right) = \frac{n}{\sigma^2}. \tag{68}$$

Theorem (Fisher-Information bei Normalverteilung II)

Es sei $\upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2)$ und μ als bekannt vorausgesetzt. Dann gilt:

- die Scorefunktion ist

$$S_n : \mathbb{R}_{>0} \to \mathbb{R}, \sigma^2 \mapsto s_n(\sigma^2) := -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \mu)^2 \tag{69}$$

- die Fisher-Information der Stichprobe ist

$$I_n : \mathbb{R}_{>0} \to \mathbb{R}, \sigma^2 \mapsto I_n(\sigma^2) := \frac{1}{\sigma^6}\sum_{i=1}^{n}(y_i - \mu)^2 - \frac{n}{2\sigma^4} \tag{70}$$

- die beobachtete Fisher-Information der Stichprobe ist

$$I_n(\hat{\sigma}_n^{2\,\mathsf{ML}}) = \frac{n}{2\hat{\sigma}^4_{\mathsf{ML}}} \tag{71}$$

- die erwartete Fisher-Information der Stichprobe ist

$$J_n : \mathbb{R}_{>0} \to \mathbb{R}, \sigma^2 \mapsto J_n(\sigma^2) := \frac{n}{2\sigma^4}. \tag{72}$$

Beweis

Wir erinnern uns, dass die Log-Likelihood-Funktion der Stichprobe $\upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2)$ bei bekanntem Erwartungswert-Parameter μ durch

$$\ell_n : \mathbb{R}_{>0} \to \mathbb{R}, \sigma^2 \mapsto \ell_n(\sigma^2) := -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \mu)^2. \tag{73}$$

gegeben ist. Die Scorefunktion ergibt sich also als

$$s_n(\sigma^2) = \frac{d}{d\sigma^2}\ell_n(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n (y_i - \mu)^2. \tag{74}$$

Die Fisher-Information der Stichprobe $\upsilon := \upsilon_1, ..., \upsilon_n$ ergibt sich als

$$\begin{aligned} I_n(\sigma^2) = -\frac{d}{d\sigma^2}s_n(\sigma^2) &= -\left(\frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^n (y_i - \mu)^2\right) \\ &= \frac{1}{\sigma^6}\sum_{i=1}^n (y_i - \mu)^2 - \frac{n}{2\sigma^4}. \end{aligned} \tag{75}$$

Beweis (fortgeführt)

Die beobachtete Fisher-Information ist die Fisher-Information an der Stelle des Maximum-Likelihood Schätzes $\hat{\sigma}_n^{2\,\text{ML}}$.

$$
\begin{aligned}
I_n(\hat{\sigma}_n^{2\,\text{ML}}) &= \frac{\sum_{i=1}^n (y_i - \mu)^2}{\left(\hat{\sigma}_n^{2\,\text{ML}}\right)^3} - \frac{n}{2\left(\hat{\sigma}_n^{2\,\text{ML}}\right)^2} \\
&= \frac{\sum_{i=1}^n (y_i - \mu)^2}{\frac{1}{n^3}\left(\sum_{i=1}^n (y_i - \mu)^2\right)^3} - \frac{n}{2\left(\hat{\sigma}_n^{2\,\text{ML}}\right)^2} \\
&= \frac{1}{\frac{1}{n^3}\left(\sum_{i=1}^n (y_i - \mu)^2\right)^2} - \frac{n}{2\left(\hat{\sigma}_n^{2\,\text{ML}}\right)^2} \\
&= \frac{n}{\left(\hat{\sigma}_n^{2\,\text{ML}}\right)^2} - \frac{n}{2\left(\hat{\sigma}_n^{2\,\text{ML}}\right)^2} \\
&= \frac{n}{2\left(\hat{\sigma}_n^{2\,\text{ML}}\right)^2} \\
&= \frac{n}{2\hat{\sigma}_n^{4\,\text{ML}}}.
\end{aligned}
\tag{76}
$$

Beweis (fortgeführt)

Die erwartete Fisher-Information ergibt sich als

$$
\begin{aligned}
J_n(\sigma^2) &= \mathbb{E}_{\sigma^2}(I_n(\sigma^2)) \\
&= \mathbb{E}_{\sigma^2}\left(\frac{1}{\sigma^6}\sum_{i=1}^n (y_i-\mu)^2 - \frac{n}{2\sigma^4}\right) \\
&= \frac{1}{\sigma^6}\sum_{i=1}^n \mathbb{E}_{\sigma^2}\left((y_i-\mu)^2\right) - \frac{n}{2\sigma^4} \\
&= \frac{1}{\sigma^6}\sum_{i=1}^n \sigma^2 - \frac{n}{2\sigma^4} \\
&= \frac{n\sigma^2}{\sigma^6} - \frac{n}{2\sigma^4} \\
&= \frac{n}{\sigma^4} - \frac{n}{2\sigma^4} \\
&= \frac{n}{2\sigma^4}.
\end{aligned}
\tag{77}
$$

Theorem (Cramér-Rao-Ungleichung)

$\mathcal{M}$ sei ein parametrisches statistisches Model mit WMF oder WDF p_θ und $\hat{\tau}_n$ sei ein erwartungstreuer Schätzer von $\tau(\theta)$. Dann gilt

$$\mathbb{V}_\theta(\hat{\tau}_n) \geq \frac{\left(\frac{d}{d\theta}\tau(\theta)\right)^2}{J(\theta)}. \tag{78}$$

Im Speziellen gilt für $\tau(\theta) := \theta$ und somit $\hat{\tau}_n = \hat{\theta}_n$ und $\left(\frac{d}{d\theta}\tau(\theta)\right)^2 = 1$, dass

$$\mathbb{V}_\theta(\hat{\theta}_n) \geq \frac{1}{J(\theta)}. \tag{79}$$

Die rechte Seite obiger Ungleichungen heißt *Cramér-Rao-Schranke*.

Bemerkungen

- Die Varianz eines erwartungstreuen Schätzers $\hat{\theta}$ von θ ist größer oder gleich der reziproken erwarteten Fisher-Information $J(\theta)$.
- Wenn $\mathbb{V}_\theta(\hat{\theta}_n) = \frac{1}{J(\theta)}$ ist, ist die Varianz des Schätzers minimal.

Beweis

Wir halten zunächst fest, dass für die Zufallsvariablen $S(\theta)$ und $\hat{\tau}_n$ mit der Korrelationsungleichung und $\mathbb{V}_\theta(S(\theta)) = J(\theta)$ gilt, dass

$$
\begin{aligned}
\frac{\mathbb{C}_\theta(S_n(\theta), \hat{\tau}_n)^2}{\mathbb{V}_\theta(S(\theta))\mathbb{V}_\theta(\hat{\tau}_n)} \le 1 \\
\Leftrightarrow \mathbb{V}_\theta(\hat{\tau}_n) \ge \frac{\mathbb{C}_\theta(S(\theta), \hat{\tau}_n)^2}{J(\theta)}.
\end{aligned}
\tag{80}
$$

Mit dem Translationstheorem für Kovarianzen, $\mathbb{E}_\theta(S(\theta)) = 0$ und der Erwartungstreue von $\hat{\tau}_n$ ergibt sich dann

$$
\mathbb{C}_\theta(S(\theta), \hat{\tau}_n) = \frac{d}{d\theta}\tau(\theta) \tag{81}
$$

wie unten gezeigt wird. Also gilt

$$
\mathbb{V}_\theta(\hat{\tau}_n) \ge \frac{\left(\frac{d}{d\theta}\tau(\theta)\right)^2}{J(\theta)}. \tag{82}
$$

Es bleibt also zu zeigen, dass $\mathbb{C}_\theta(S(\theta), \hat{\tau}_n) = \frac{d}{d\theta}\tau(\theta)$. Dies ergibt aber ergibt sich mit

$$
\begin{aligned}
\mathbb{C}_\theta(S(\theta), \hat{\tau}_n) = \mathbb{E}_\theta(S(\theta)\hat{\tau}_n) - \mathbb{E}_\theta(S(\theta))\mathbb{E}_\theta(\hat{\tau}_n) & = \mathbb{E}_\theta(S(\theta)\hat{\tau}_n) \\
& = \int S(\theta)\, \hat{\tau}_n \, p_\theta(x)\, dx \\
& = \int \frac{d}{d\theta} \ln L(\theta)\, \hat{\tau}_n \, p_\theta(x)\, dx \\
& = \int \frac{\frac{d}{d\theta} L(\theta)}{L(\theta)}\, \hat{\tau}_n \, p_\theta(x)\, dx \\
& = \int \frac{\frac{d}{d\theta} L(\theta)}{p_\theta(x)}\, \hat{\tau}_n \, p_\theta(x)\, dx \\
& = \int \frac{d}{d\theta} L(\theta)\, \hat{\tau}_n \, dx \\
& = \frac{d}{d\theta} \int L(\theta)\, \hat{\tau}_n \, dx \\
& = \frac{d}{d\theta} \int \hat{\tau}_n \, p_\theta(x)\, dx = \frac{d}{d\theta} \mathbb{E}_\theta(\hat{\tau}_n) = \frac{d}{d\theta} \tau(\theta).
\end{aligned}
\tag{83}
$$

Appendix | Asymptotische Effizienz

Definition (Asymptotische Effizienz)

$\upsilon_1, ..., \upsilon_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\theta}_n$ sei ein Parameterschätzer für θ. Weiterhin sei $J_n(\theta)$ die erwartete Fisher-Information der Stichprobe $\upsilon := \upsilon_1, ..., \upsilon_n$. Wenn gilt, dass

$$\hat{\theta}_n \overset{a}{\sim} N\left(\theta, J_n(\theta)^{-1}\right), \tag{84}$$

dann heißt $\hat{\theta}_n$ *asymptotisch effizient*.

Bemerkungen

- Asymptotische Effizienz impliziert asymptotische Normalität.
- Asymptotische Effizienz impliziert asymptotische Erwartungstreue.
- Die Varianz der asymptotischen Verteilung heißt *asymptotische Varianz*.
- Die Varianz eines asymptotisch effizienten Schätzers ist gleich der Cramér-Rao-Schranke.
- Der Begriff der *Effizienz* wird in der Literatur nicht einheitlich verwendet.

Appendix | Konsistenz des Erwartungswertschätzers

Konsistenz des Erwartungswertschätzers bei Normalverteilung

Theorem (Mittlerer-Quadratischer-Fehler-Kriterium für Konsistenz)

$\upsilon := \upsilon_1, ..., \upsilon_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\tau}_n$ sei ein Schätzer von τ. Wenn $\lim_{n\to\infty} \mathrm{MQF}(\hat{\tau}_n) = 0$ gilt, dann ist $\hat{\tau}_n$ ein konsistenter Schätzer.

Beweis

Mit der Chebychev-Ungleichung gilt, dass

$$\mathbb{P}_\theta\left(|\hat{\tau}_n(\upsilon) - \tau(\theta)| \geq \epsilon\right) \leq \frac{\mathbb{E}_\theta\left((\hat{\tau}_n(\upsilon) - \tau(\theta))^2\right)}{\epsilon^2} \tag{85}$$

Grenzwertbildung ergibt dann

$$\lim_{n\to\infty} \mathbb{P}_\theta\left(|\hat{\tau}_n(\upsilon) - \tau(\theta)| \geq \epsilon\right) \leq \frac{1}{\epsilon^2} \lim_{n\to\infty} \mathbb{E}_\theta\left((\hat{\tau}_n(\upsilon) - \tau(\theta))^2\right). \tag{86}$$

Wenn also $\lim_{n\to\infty} \mathbb{E}_\theta\left((\hat{\tau}_n(\upsilon) - \tau(\theta))^2\right) = 0$ gilt, dann gilt mit $\mathbb{P}_\theta(|\hat{\tau}_n(\upsilon) - \tau(\theta)| \geq \epsilon) \geq 0$, dass

$$\lim_{n\to\infty} \mathbb{P}_\theta\left(|\hat{\tau}_n(\upsilon) - \tau(\theta)| \geq \epsilon\right) = 0. \tag{87}$$

Also ist $\hat{\tau}_n$ ein konsistenter Schätzer.

Konsistenz des Erwartungswertschätzers bei Normalverteilung

Theorem (Bias-Varianz-Kriterium für Konsistenz)

$\upsilon_1, ..., \upsilon_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\tau}_n$ sei ein Schätzer von τ. Wenn

$$\lim_{n\to\infty} \mathrm{B}(\hat{\tau}_n) = 0 \text{ und } \lim_{n\to\infty} \mathbb{V}_\theta(\hat{\tau}_n) = 0 \tag{88}$$

gilt, dann ist $\hat{\tau}_n$ ein konsistenter Schätzer

Beweis

Wenn $n \to \infty$, dann gilt $\mathrm{B}(\hat{\tau}_n) \to 0$, also auch $\mathrm{B}(\hat{\tau}_n)^2 \to 0$. Wenn für $n \to \infty$ sowohl $\mathrm{B}(\hat{\tau}_n)^2 \to 0$ als auch $\mathbb{V}_\theta(\hat{\tau}_n) \to 0$, dann gilt auch $\lim_{n\to\infty} \mathrm{MQF}(\hat{\theta}_n) = 0$. Also gilt mit dem MQF-Kriterium, dass $\hat{\tau}_n$ konsistent ist.

Konsistenz des Erwartungswertschätzers bei Normalverteilung

Theorem (Konsistenz des Erwartungswertschätzers bei Normalverteilung)

Es sei $\upsilon := \upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2)$. Dann ist $\bar{\upsilon}_n$ ein konsistenter Schätzer von $\mathbb{E}(\upsilon_1)$.

Beweis

Mit der Erwartungstreue des Stichprobenmittels als Schätzer für den Erwartungswert gilt zunächst

$$\lim_{n \to \infty} \mathrm{B}(\bar{\upsilon}_n) = 0 \tag{89}$$

Weiterhin gilt mit der Varianz des Stichprobenmittels

$$\lim_{n \to \infty} \mathbb{V}_\theta(\bar{\upsilon}_n) = \lim_{n \to \infty} \frac{1}{n} \mathbb{V}(\upsilon_1) = 0. \tag{90}$$

Mit dem Bias-Varianz-Kriterium folgt dann die Konsistenz von $\bar{\upsilon}_n$ als Schätzer von $\mathbb{E}(\upsilon_1)$

References

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.

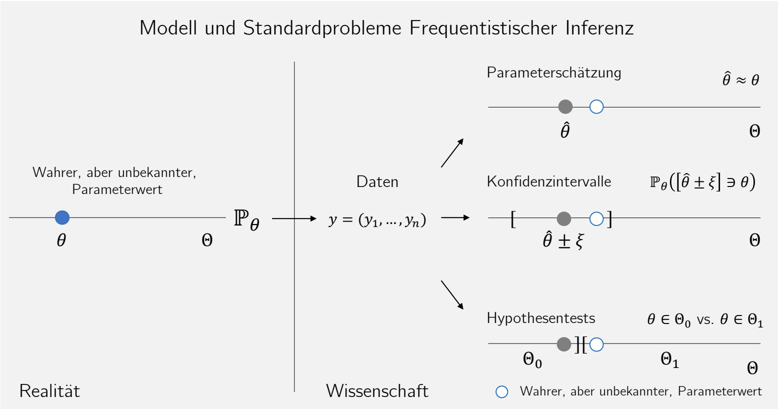
Vaart, A. W. van der. 1998. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge, UK ; New York, NY, USA: Cambridge University Press.

Wahrscheinlichkeitstheorie und Frequentistische Inferenz

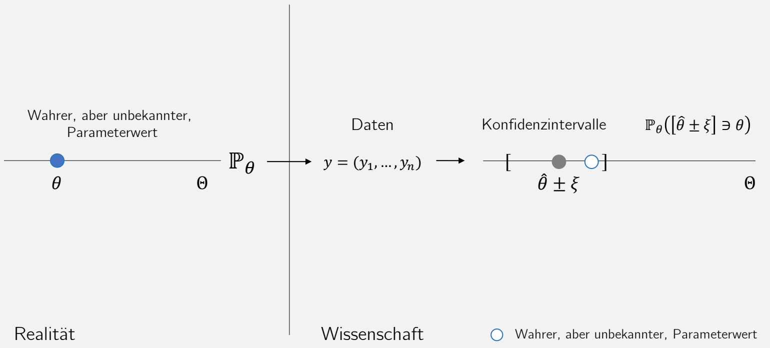
BSc Psychologie WiSe 2022/23

Prof. Dr. Dirk Ostwald

(11) Konfidenzintervalle


Modell und Standardprobleme Frequentistischer Inferenz
Wahrer, aber unbekannter, Parameterwert
θ
Θ
$\mathbb{P}_\theta$
Daten
$y = (y_1, ..., y_n)$
Parameterschätzung
$\hat{\theta} \approx \theta$
$\hat{\theta}$
Θ
Konfidenzintervalle
$\mathbb{P}_\theta([\hat{\theta} \pm \xi] \ni \theta)$
$\hat{\theta} \pm \xi$
Θ
Hypothesentests
$\theta \in \Theta_0$ vs. $\theta \in \Theta_1$
Θ_0
Θ_1
Θ
Realität
Wissenschaft
Wahrer, aber unbekannter, Parameterwert

Modell und Standardprobleme Frequentistischer Inferenz



Standardannahmen Frequentistischer Inferenz

$\mathcal{M}$ sei ein statistisches Modell mit Stichprobe $\upsilon_1, ..., \upsilon_n \sim p_\theta$. **Es wird angenommen, dass ein konkret vorliegender Datensatz $y = (y_1, ..., y_n) \in \mathbb{R}^n$ eine der möglichen Realisierungen von $\upsilon_1, ..., \upsilon_n \sim p_\theta$ ist.** Aus Frequentistischer Sicht kann man eine Studie unter gleichen Umständen unendlich oft wiederholen und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. das Stichprobenmittel:

Datensatz (1) : $y^{(1)} = \left(y_1^{(1)}, y_2^{(1)}, ..., y_n^{(1)}\right)$ mit $\bar{y}^{(1)} = \frac{1}{n}\sum_{i=1}^n y_i^{(1)}$

Datensatz (2) : $y^{(2)} = \left(y_1^{(2)}, y_2^{(2)}, ..., y_n^{(2)}\right)$ mit $\bar{y}^{(2)} = \frac{1}{n}\sum_{i=1}^n y_i^{(2)}$

Datensatz (3) : $y^{(3)} = \left(y_1^{(3)}, y_2^{(3)}, ..., y_n^{(3)}\right)$ mit $\bar{y}^{(3)} = \frac{1}{n}\sum_{i=1}^n y_i^{(3)}$

Datensatz (4) : $y^{(4)} = \left(y_1^{(4)}, y_2^{(4)}, ..., y_n^{(4)}\right)$ mit $\bar{y}^{(4)} = \frac{1}{n}\sum_{i=1}^n y_i^{(4)}$

Datensatz (5) : $y^{(5)} = ...$

Um die Qualität statistischer Methoden zu beurteilen betrachtet die Frequentistische Statistik deshalb die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken unter Annahme von $\upsilon_1, ..., \upsilon_n \sim p_\theta$. Was zum Beispiel ist die Verteilung der $\bar{y}^{(1)}$, $\bar{y}^{(2)}$, $\bar{y}^{(3)}$, $\bar{y}^{(4)}$, ... also die Verteilung der Zufallsvariable $\bar{\upsilon}$?

Wenn eine statistische Methode im Sinne der Frequentistischen Standardannahmen "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

Konfidenzintervall für den Varianzparameter der Normalverteilung

Anwendungsbeispiel

Selbstkontrollfragen

Definition

Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

Konfidenzintervall für den Varianzparameter der Normalverteilung

Anwendungsbeispiel

Selbstkontrollfragen

Definition (δ-Konfidenzintervall)

Es sei $\upsilon = \upsilon_1, ..., \upsilon_n \sim p_\theta$ eine Stichprobe, $\delta \in]0,1[$, und $G_u(\upsilon)$ und $G_o(\upsilon)$ seien zwei Statistiken. Dann ist ein δ-*Konfidenzintervall* ein Intervall der Form $\kappa := [G_u, G_o]$, so dass

$$\mathbb{P}_\theta\left(\kappa \ni \theta\right) = \mathbb{P}_\theta\left(G_u(\upsilon) \le \theta \le G_o(\upsilon)\right) = \delta \text{ für alle } \theta \in \Theta \text{ gilt.} \tag{1}$$

δ heißt das *Konfidenzniveau* oder die *Überdeckungswahrscheinlichkeit* des Konfidenzintervalls. Die Statistiken $G_u(\upsilon)$ und $G_o(\upsilon)$ sind die unteren und oberen Grenzen des Konfidenzintervalls.

Bemerkungen

- θ ist fest, nicht zufällig, und unbekannt.
- κ ist ein zufälliges Intervall, weil $G_u(\upsilon)$ und $G_o(\upsilon)$ 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.

Definition

Zwei Interpretationen von δ-Konfidenzintervallen

(1) Wird ein Zufallsvorgang unter gleichen Umständen häufig wiederholt, so überdeckt das zugehörige δ-Konfidenzintervall den wahren, aber unbekannten Parameterwert im langfristigen Mittel in $\delta \cdot 100\%$ der Fälle. Technischer ausgedrückt, für unabhängig und identisch realisierte Stichproben einer Verteilung mit wahrem, aber unbekannten, 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 unbekannten, Parametern $\theta_1, \theta_2, ...$ und realisierte δ-Konfidenzintervalle für eben jene Menge von wahren, aber unbekannten Parametern $\theta_1, \theta_2, ...$ Dann überdecken im langfristigen Mittel $\delta \cdot 100\%$ der Konfidenzintervalle den wahren, aber unbekannten, Wert θ_i für $i = 1, 2, ...$. Technischer ausgedrückt, für unabhängig realisierte Stichproben von Verteilungen mit wahren, aber unbekannten, Parametern $\theta_1, \theta_2, ...$ überdecken im langfristigen Mittel entsprechende δ-Konfidenzintervalle θ_i für $i = 1, 2, ...$ in $\delta \cdot 100\%$ aller Fälle.

Wir demonstrieren im Folgenden beide Interpretationen mithilfe von Simulationen.

Definition

Allgemeine Konstruktion von Konfidenzintervallen

(1) Definition des statistischen Modells

(2) Definition der Konfidenzintervallstatistik

(3) Analyse der Verteilung der Konfidenzintervallstatistik

(4) Etablierung der Konfidenzbedingung

(5) Definition des Konfidenzintervalls

Beispiele

(1) Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

(2) Konfidenzintervall für den Varianzparameter der Normalverteilung

Definition

Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

Konfidenzintervall für den Varianzparameter der Normalverteilung

Anwendungsbeispiel

Selbstkontrollfragen

Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

(1) Definition des statistischen Modells

Es sei $\upsilon := \upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2)$ eine Stichprobe mit unbekannten Parametern μ und $\sigma^2 > 0$. Wir entwickeln ein δ-Konfidenzintervall für den Erwartungswertparameter μ.

(2) Definition der Statistik

Wir betrachten die T-Konfidenzintervallstatistik

$$T := \frac{\sqrt{n}}{S}(\bar{\upsilon} - \mu) \text{ mit } \bar{\upsilon} := \frac{1}{n}\sum_{i=1}^{n} \upsilon_i \text{ und } S := \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\upsilon_i - \bar{\upsilon})^2}. \tag{2}$$

(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 $\upsilon_1, ..., \upsilon_n$ (via $\bar{\upsilon}$ und S), während ihre Verteilung weder von μ noch von σ^2 abhängt. Wir bezeichnen die die WDF einer t-verteilten Zufallvariable mit t, die KVF einer t-verteilten Zufallvariable mit Ψ und die inverse KVF einer t-verteilten Zufallvariable mit Ψ^{-1}.

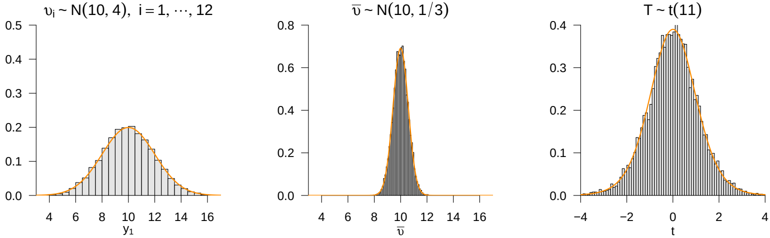
Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

(3) Analyse der Verteilung der Konfidenzintervallstatistik (fortgeführt)

```r
# Modellformulierung
mu      = 10                                          # w.a.u. Erwartungswertparameter
sigsqr  = 4                                           # wahrer bekannter Varianzparameter
n       = 12                                          # Stichprobengroesse
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){                                       # Simulationsiterationen
  y         = rnorm(n,mu,sqrt(sigsqr))                # Stichprobenrealisierung
  y_i[s]    = y[1]                                    # \ups_i
  y_bar[s]  = mean(y)                                 # Stichprobenmittelrealisierung
  Tee[s]    = sqrt(n)*((y_bar[s] - mu)/sqrt(var(y)))  # T-Statistik Realisierung
}
```

Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

(3) Analyse der Verteilung der Konfidenzintervallstatistik (fortgeführt)

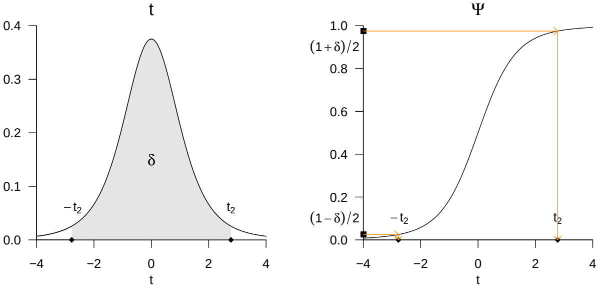


(4) Etablierung der Konfidenzbedingung

Für $\delta \in]0,1[$ seien

$$t_1 := \Psi^{-1}\left(\frac{1-\delta}{2}; n-1\right) \text{ und } t_2 := \Psi^{-1}\left(\frac{1+\delta}{2}; n-1\right) \quad (3)$$

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 $\mathbb{P}(-t_2 \le T \le t_2) = \delta$.



(4) Etablierung der Konfidenzbedingung (fortgeführt)

Mit der Definition von t_2 wie oben folgt dann aber

$$
\begin{aligned}
\delta &= \mathbb{P}\left(-t_2 \le T \le t_2\right) \\
&= \mathbb{P}\left(-t_2 \le \frac{\sqrt{n}}{S}(\bar{\upsilon} - \mu) \le t_2\right) \\
&= \mathbb{P}\left(-\frac{S}{\sqrt{n}}t_2 \le \bar{\upsilon} - \mu \le \frac{S}{\sqrt{n}}t_2\right) \\
&= \mathbb{P}\left(-\bar{\upsilon} - \frac{S}{\sqrt{n}}t_2 \le -\mu \le -\bar{\upsilon} + \frac{S}{\sqrt{n}}t_2\right) \\
&= \mathbb{P}\left(\bar{\upsilon} + \frac{S}{\sqrt{n}}t_2 \ge \mu \ge \bar{\upsilon} - \frac{S}{\sqrt{n}}t_2\right) \\
&= \mathbb{P}\left(\bar{\upsilon} - \frac{S}{\sqrt{n}}t_2 \le \mu \le \bar{\upsilon} + \frac{S}{\sqrt{n}}t_2\right) \\
&= \mathbb{P}\left(\left[\bar{\upsilon} - \frac{S}{\sqrt{n}}t_2, \bar{\upsilon} + \frac{S}{\sqrt{n}}t_2\right] \ni \mu\right).
\end{aligned}
\tag{4}
$$

Wir können nun das gesuchte Konfidenzintervall definieren.

(5) Definition des Konfidenzintervalls

Es sei $\upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2)$ mit unbekannten Parametern μ und σ^2, es sei $\delta \in]0,1[$, und es sei

$$t_\delta := \Psi^{-1}\left(\frac{1+\delta}{2}; n-1\right). \tag{5}$$

Definiere

$$\kappa := \left[\bar{\upsilon} - \frac{S}{\sqrt{n}} t_\delta, \bar{\upsilon} + \frac{S}{\sqrt{n}} t_\delta\right]. \tag{6}$$

Dann gilt wie oben gezeigt

$$\mathbb{P}(\kappa \ni \mu) = \delta. \tag{7}$$

Damit ist κ ein δ-Konfidenzintervall für μ.

Man beachte, dass κ ein zufälliges Intervall ist, weil $\bar{\upsilon}$ und S Zufallsvariablen sind.

Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

(5) Definition des Konfidenzintervalls (fortgeführt)

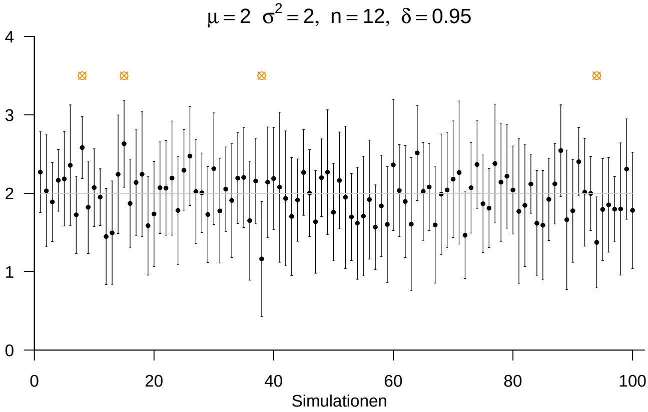
Simulation der ersten Interpretation eines Konfidenzintervalls

```
# 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                                            # Stichprobengroesse
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(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,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 KI Grenze
   kappa[i,2] = y_bar[i] + (S[i]/sqrt(n))*t_delta       # obere KI Grenze
}
```

© 2023 Dirk Ostwald CC BY 4.0

(5) Definition des Konfidenzintervalls (fortgeführt)

Simulation der ersten Interpretation eines Konfidenzintervalls



(5) Definition des Konfidenzintervalls (fortgeführt)

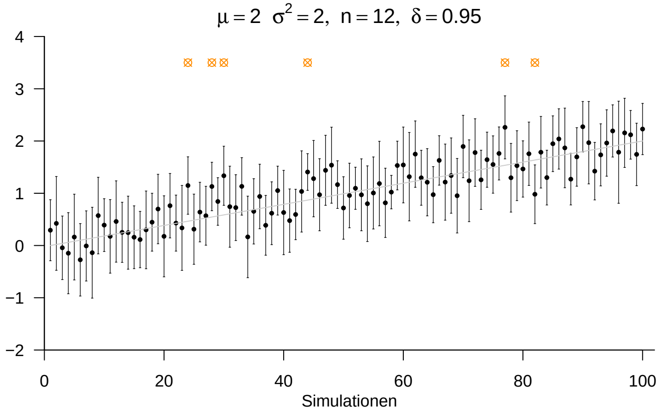
Simulation der zweiten Interpretation eines Konfidenzintervalls

```
# 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                                       # Stichprobengroesse
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 KI Grenze
   kappa[i,2] = y_bar[i] + (S[i]/sqrt(n))*t_delta  # obere KI Grenze
}
```

© 2023 Dirk Ostwald CC BY 4.0

(5) Definition des Konfidenzintervalls (fortgeführt)

Simulation der zweiten Interpretation eines Konfidenzintervalls



Definition

Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

Konfidenzintervall für den Varianzparameter der Normalverteilung

Anwendungsbeispiel

Selbstkontrollfragen

Konfidenzintervall für den Varianzparameter der Normalverteilung

(1) Definition des statistischen Modells

Es sei $\upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2)$ eine Stichprobe mit unbekanntem Varianzparameter $\sigma^2 > 0$ und bekanntem oder unbekanntem Erwartungswertparameter μ. Wir entwickeln ein δ-Konfidenzintervall für den Varianzparameter σ^2.

(2) Definition der Statistik

Wir betrachten die U-Konfidenzintervallstatistik

$$U := \frac{n-1}{\sigma^2} S^2 \text{ mit } S^2 := \frac{1}{n-1} \sum_{i=1}^{n} (\upsilon_i - \bar{\upsilon})^2 . \tag{8}$$

(3) Analyse der Verteilung der Konfidenzintervallstatistik

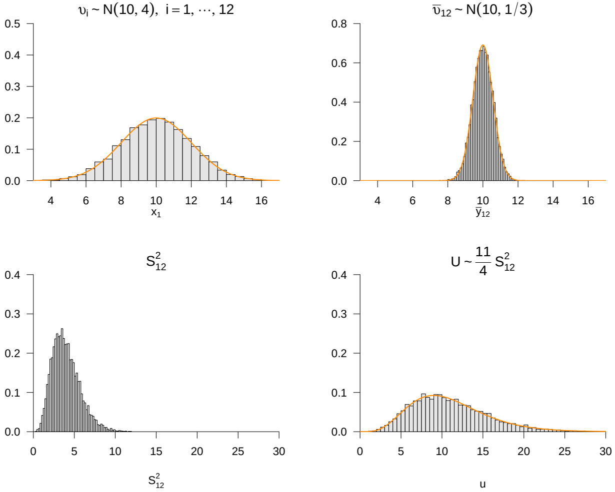
Für die U-Konfidenzintervallstatistik gilt $U \sim \chi^2(n-1)$. Für einen Beweis dieser Tatsache verweisen wir auf Casella and Berger (2012), Abschnitt 5.3. Die U-Konfidenzintervallstatistik ist eine Funktion von $\upsilon_1, ..., \upsilon_n$ (via S^2) und σ^2, während ihre Verteilung nicht von σ^2 abhängt. Wir bezeichnen die WDF einer χ^2-verteilten Zufallvariable mit χ^2, die KVF einer χ^2-verteilten Zufallvariable mit Ξ und die inverse KVF einer χ^2-verteilten Zufallvariable mit Ξ^{-1}.

(2) Konfidenzintervall für den Varianzparameter der Normalverteilung

(3) Analyse der Verteilung der Konfidenzintervallstatistik (fortgeführt)

```
# Modellformulierung
mu      = 10                                    # wahrer Erwartungswertparameter
sigsqr  = 4                                     # wahrer bekannter Varianzparameter
n       = 12                                    # Stichprobengroesse
ns      = 1e4                                   # Anzahl Stichprobenrealisierungen
res     = 1e3                                   # Ausgangsraumaufloesung

# analytische Definitionen und Resultate
y_1     = seq(3,17,len = res)                   # y_1 Raum
y_bar   = seq(3,17,len = res)                   # y_bar Raum
u_1     = seq(0,30,len = res)                   # normalisierte S^2 Raum
p_y_1   = dnorm(y_1,mu,sqrt(sigsqr))            # y_1 WDF
p_y_bar = dnorm(y_bar,mu,sqrt(sigsqr/n))        # y_bar WDF
p_u     = dchisq(u_1,n-1)                       # u_1 WDF

# Simulation
y_i     = rep(NaN,ns)                           # y_1 Array
y_bar   = rep(NaN,ns)                           # \bar{y} Array
S_sqr   = rep(NaN,ns)                           # S^2 Array
U       = rep(NaN,ns)                           # U Array

for(s in 1:ns){                                 # Simulationsiterationen
  y         = rnorm(n,mu,sqrt(sigsqr))          # Stichprobenrealisierung
  y_i[s]    = y[1]                              # \ups_i
  y_bar[s]  = mean(y)                           # Stichprobenmittelrealisierung
  S_sqr[s]  = var(y)                            # Stichprobenvarianzrealisierung

  # U-Statistik Realisiation
  U[s]      = ((n-1)/sigsqr)*S_sqr[s]
}
```

(3) Analyse der Verteilung der Konfidenzintervallstatistik (fortgeführt)

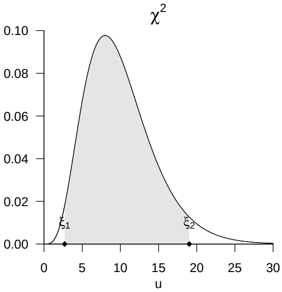


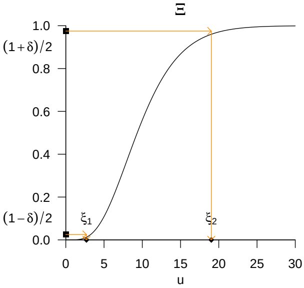
(4) Etablierung der Konfidenzbedingung

Für $\delta \in]0,1[$ seien

$$\xi_1 := \Xi^{-1}\left(\frac{1-\delta}{2}; n-1\right) \text{ und } \xi_2 := \Xi^{-1}\left(\frac{1+\delta}{2}; n-1\right) \tag{9}$$

Es gilt dann $(1+\delta)/2 - (1-\delta)/2 = \delta$ gilt und zum Beispiel gilt für $n = 10$ und $\delta = 0.95$, $\xi_1 := \Xi^{-1}(0.025; 9) = 2.70$ und $\xi_2 := \Xi^{-1}(0.975; 9) = 19.0$. Es gilt hier also per Definition $\mathbb{P}(\xi_1 \le U \le \xi_2) = \delta$.





Konfidenzintervall für den Varianzparameter der Normalverteilung

(4) Etablierung der Konfidenzbedingung (fortgeführt)

Mit der Definition von ξ_1 und ξ_2 wie oben folgt dann aber

$$
\begin{aligned}
\delta &= \mathbb{P}\left(\xi_1 \le U \le \xi_2\right) \\
&= \mathbb{P}\left(\xi_1 \le \frac{n-1}{\sigma^2} S^2 \le \xi_2\right) \\
&= \mathbb{P}\left(\xi_1^{-1} \ge \frac{\sigma^2}{(n-1)S^2} \ge \xi_2^{-1}\right) \\
&= \mathbb{P}\left(\frac{(n-1)S^2}{\xi_1} \ge \sigma^2 \ge \frac{(n-1)S^2}{\xi_2}\right) \\
&= \mathbb{P}\left(\frac{(n-1)S^2}{\xi_2} \le \sigma^2 \le \frac{(n-1)S^2)}{\xi_1}\right) \\
&= \mathbb{P}\left(\left[\frac{(n-1)S^2}{\xi_2}, \frac{(n-1)S^2)}{\xi_1}\right] \ni \sigma^2\right).
\end{aligned}
\tag{10}
$$

Wir können nun das gesuchte Konfidenzintervall definieren.

(5) Definition des Konfidenzintervalls

Es sei $\upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2)$ eine Stichprobe mit unbekanntem Varianzparameter σ^2 und bekanntem oder unbekanntem Erwartungswertparamter μ, und es seien weiterhin $\delta \in]0, 1[$ sowie

$$\xi_1 := \Xi^{-1}\left(\frac{1-\delta}{2}; n-1\right) \text{ und } \xi_2 := \Xi^{-1}\left(\frac{1+\delta}{2}; n-1\right). \quad (11)$$

Definiere

$$\kappa := \left[\frac{(n-1)S^2}{\xi_2}, \frac{(n-1)S^2}{\xi_1}\right]. \quad (12)$$

Dann gilt wie oben gezeigt

$$\mathbb{P}(\kappa \ni \sigma^2) = \delta. \quad (13)$$

Damit ist κ ein δ-Konfidenzintervall für σ^2.

Man beachte, dass κ ein zufälliges Intervall ist, weil S^2 eine Zufallsvariable ist.

© 2023 Dirk Ostwald CC BY 4.0

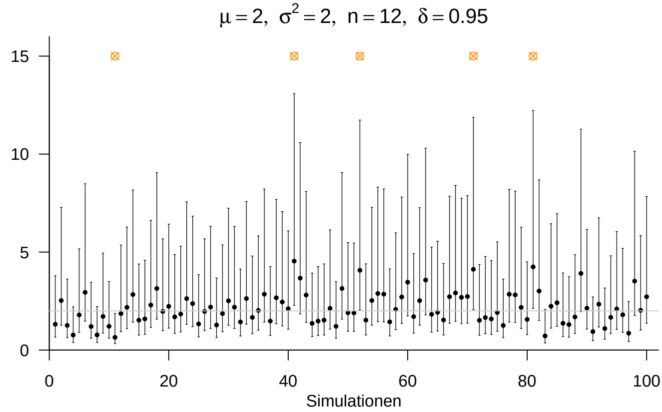
(5) Definition des Konfidenzintervalls (fortgeführt)

Simulation

```
# Modellformulierung
set.seed(1)                                  # random number generator seed
mu      = 2                                  # w.a.u. Erwartungswertparameter
sigsqr  = 2                                  # w.a.u. Varianzparameter
n       = 12                                 # Stichprobengroesse
delta   = 0.95                               # Konfidenzbedingung
xi_1    = qchisq((1-delta)/2, n - 1)         # \Xi^2((1-\delta)/2; n - 1)
xi_2    = qchisq((1+delta)/2, n - 1)         # \Xi^2((1+\delta)/2; n - 1)

# Simulation
ns      = 1e2                                # Anzahl Simulationen
y_bar   = rep(NaN,ns)                        # Stichprobenmittelarray
S2      = rep(NaN,ns)                        # Stichprobenvarianzarray
kappa   = matrix(rep(NaN,2*ns), ncol = 2)    # Konfidenzintervallarray
for(i in 1:ns){                              # Simulationsiterationen
   y        = rnorm(n,mu,sqrt(sigsqr))       # Stichprobenrealisierung
   S2[i]    = var(y)                         # Stichprobenvarianz
   kappa[i,1]   = (n-1)*S2[i]/xi_2           # untere KI Grenze
   kappa[i,2]   = (n-1)*S2[i]/xi_1           # obere KI Grenze
}
```

(5) Definition des Konfidenzintervalls (fortgeführt)

Simulation



Definition

Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

Konfidenzintervall für den Varianzparameter der Normalverteilung

Anwendungsbeispiel

Selbstkontrollfragen

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Pre-BDI

n = 12

Post-BDI

BDI-II Fragebogen

Name | Alter | Geschlecht m / w | Datum

Anleitung: Dieser Fragebogen enthält 21 Gruppen von Aussagen. Bitte lesen Sie jede dieser Gruppen von Aussagen sorgfältig durch und suchen Sie sich dann in jeder Gruppe **eine Aussage** heraus, die am besten beschreibt, wie Sie sich **in den letzten zwei Wochen, einschließlich heute, gefühlt haben**. Kreuzen Sie die Zahl neben der Aussage an, die Sie sich herausgesucht haben (0, 1, 2 oder 3). Falls in einer Gruppe mehrere Aussagen gleichermaßen auf Sie zutreffen, kreuzen Sie die Aussage mit der höheren Zahl an. Achten Sie bitte darauf, dass Sie in jeder Gruppe nicht mehr als eine Aussage ankreuzen, das gilt auch für Gruppe 16 (Veränderungen der Schlafgewohnheiten) oder Gruppe 18 (Veränderungen des Appetits).

1.) Traurigkeit
0 Ich bin nicht traurig.
1 Ich bin oft traurig.
2 Ich bin ständig traurig.
3 Ich bin so traurig oder unglücklich, dass ich es nicht aushalte.

2.) Pessimismus
0 Ich sehe nicht mutlos in die Zukunft.
1 Ich sehe mutloser in die Zukunft als sonst.
2 Ich bin mutlos und erwarte nicht, dass meine Situation besser wird.
3 Ich glaube, dass meine Zukunft hoffnungslos ist und nur noch schlechter wird.

3.) Versagensgefühle
0 Ich fühle mich nicht als Versager.
1 Ich habe häufiger Versagensgefühle.
2 Wenn ich zurückblicke, sehe ich eine Menge Fehlschläge.
3 Ich habe das Gefühl, als Mensch ein völliger Versager zu sein.

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.

5.) Schuldgefühle
0 Ich habe keine besonderen Schuldgefühle.
1 Ich habe oft Schuldgefühle wegen Dingen, die ich getan habe oder hätte tun sollen.
2 Ich habe die meiste Zeit Schuldgefühle.
3 Ich habe ständig Schuldgefühle.

6.) Bestrafungsgefühle
0 Ich habe nicht das Gefühl, für etwas bestraft zu sein.
1 Ich habe das Gefühl, vielleicht bestraft zu werden.
2 Ich erwarte, bestraft zu werden.
3 Ich habe das Gefühl, bestraft zu sein.

7.) Selbstablehnung
0 Ich halte von mir genauso viel wie immer.
1 Ich habe Vertrauen in mich verloren.
2 Ich bin von mir enttäuscht.
3 Ich lehne mich völlig ab.

8.) Selbstvorwürfe
0 Ich kritisiere oder tadle mich nicht mehr als sonst.
1 Ich bin mir gegenüber kritischer als sonst.
2 Ich kritisiere mich für all meine Mängel.
3 Ich gebe mir die Schuld für alles Schlimme, was passiert.

9.) Selbstmordgedanken
0 Ich denke nicht daran, mir etwas anzutun.
1 Ich denke manchmal an Selbstmord, aber ich würde es nicht tun.
2 Ich möchte mich am liebsten umbringen.
3 Ich würde mich umbringen, wenn ich die Gelegenheit dazu hätte.

10.) Weinen
0 Ich weine nicht öfter als früher.
1 Ich weine jetzt mehr als früher.
2 Ich weine beim geringsten Anlass.
3 Ich möchte gern weinen, aber ich kann nicht.

PEARSON PsychCorp

Summe Seite 1:

⇒ Pre-Post BDI Score Reduktion

Anwendungsbeispiel

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

```
fname = file.path(getwd(), "11_Konfidenzintervalle.csv")
D     = read.table(fname, sep = ",", header = T)
```

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

Anwendungsbeispiel

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Für die Pre-Post BDI Score Reduktion υ_i der $i$ten von n Patient:innen legen wir das Modell

$$\upsilon_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, ..., n \tag{14}$$

zugrunde. Dabei wird die Pre-Post BDI Reduktion υ_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

$$\upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2). \tag{15}$$

Die Standardprobleme der Frequentistischen Inferenz führen in diesem Szenario auf folgende Fragen:

(1) Was sind sinnvolle Tipps für die wahren, aber unbekannten, Parameterwerte μ und σ^2?
(2) Wie hoch ist im Frequentistischen Sinn die mit diesen Tipps assoziierte Unsicherheit?
(3) Entscheiden wir uns sinnvollerweise für die Hypothese, dass gilt $\mu \neq 0$?

Anwendungsbeispiel

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

```
# Einlesen und Auswahl der Daten
fname = file.path(getwd(), "11_Konfidenzintervalle.csv")
D     = read.table(fname, sep = ",", header = T)
y     = D$BDI.Reduktion
```

Wir haben in (10) Parameterschätzung gesehen, dass unverzerrte 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 die mit diesen Tipps assoziierte Unsicherheit zu quantifizieren, geht die Frequentistische Inferenz zur Intervallschätzung mithilfe von δ-Konfidenzintervallen über, für die die assoziierte Unsicherheit dann ein ein prädefiniertes Level von 1 - δ hat. Im langfristigen Mittel überdeckt ein angegebenes δ-Konfidenzintervall den wahren, aber unbekannten Parameterwert in (nur) $1 - \delta \cdot$ 100 von 100 Fällen nicht. Für ein großes δ wie $\delta = 0.95$ ist die mit dieser Intervallschätzung assoziierte Unsicherheit also mit $1 - 0.95 = 0.05$ also eher gering.

Anwendungsbeispiel

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 KI Grenze
mu_hat_o = y_bar + (s/sqrt(n))*t_delta           # obere  KI Grenze
```

Das 0.95-Konfidenzintervall für den Erwartungswertparameter ist $[0.80, 5.52]$. Im langfristigen Mittel überdeckt so berechnetes Konfidenzintervall den wahren, aber unbekannten, Erwartungswertparameter in 95 von 100 Fällen.

```
# Konfidenzintervall für den Erwartungswertparameter
delta        = 0.95                              # Konfidenzlevel
n           = length(y)                          # Anzahl Datenpunkte
xi_1        = qchisq((1-delta)/2, n - 1)         # \Xi^2((1-\delta)/2; n - 1)
xi_2        = qchisq((1+delta)/2, n - 1)         # \Xi^2((1+\delta)/2; n - 1)
s2          = var(y)                             # Stichprobenstandardabweichung
sigsqr_hat  = s2                                 # Varianzparameterschätzer
sigsqr_hat_u = (n-1)*s2/xi_2                     # untere KI Grenze
sigsqr_hat_o = (n-1)*s2/xi_1                     # obere KI Grenze
```

Das 0.95-Konfidenzintervall für den Varianzparameter ist $[6.91, 39.74]$. Im langfristigen Mittel überdeckt ein so berechnetes Konfidenzintervall den wahren, aber unbekannten, Varianzparameter in 95 von 100 Fällen.

Definition

Konfidenzintervall für den Erwartungswertparameter der Normalverteilung

Konfidenzintervall für den Varianzparameter der Normalverteilung

Anwendungsbeispiel

Selbstkontrollfragen

Selbstkontrollfragen

1. Definieren Sie den Begriff des δ-Konfidenzintervalls.
2. Geben Sie zwei Interpretationen eines δ-Konfidenzintervalls.
3. Erläutern Sie die typischen Schritte zur Konstruktion eines δ-Konfidenzintervalls.
4. Definieren Sie die T-Konfidenzintervallstatistik und geben Sie ihre Verteilung an.
5. Geben Sie das δ-Konfidenzintervall für den Erwartungswert der Normalverteilung an.
6. Definieren Sie die U-Konfidenzintervallstatistik und geben Sie ihre Verteilung an.
7. Geben Sie das δ-Konfidenzintervall für den Varianzparameter der Normalverteilung an.

References

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

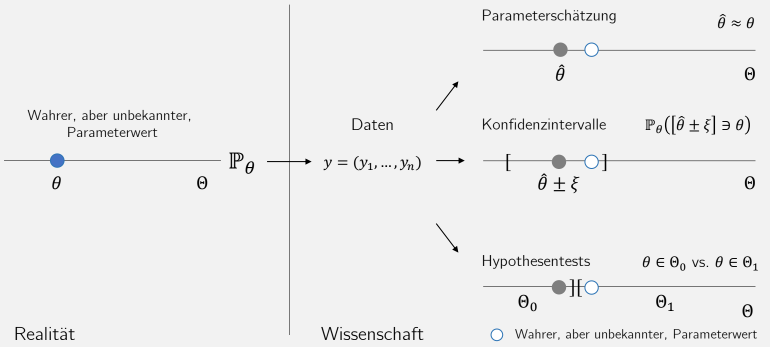
Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2021/22

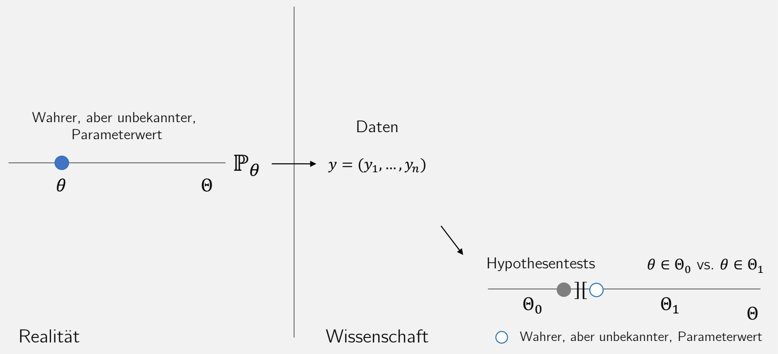
Prof. Dr. Dirk Ostwald

(12) Hypothesentests

Modell und Standardprobleme Frequentistischer Inferenz



Modell und Standardprobleme Frequentistischer Inferenz



Standardannahmen Frequentistischer Inferenz

$\mathcal{M}$ sei ein statistisches Modell mit Stichprobe $\upsilon_1, ..., \upsilon_n \sim p_\theta$. **Es wird angenommen, dass ein konkret vorliegender Datensatz $y = (y_1, ..., y_n) \in \mathbb{R}^n$ eine der möglichen Realisierungen von $\upsilon_1, ..., \upsilon_n \sim p_\theta$ ist.** Aus Frequentistischer Sicht kann man eine Studie unter gleichen Umständen unendlich oft wiederholen und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. das Stichprobenmittel:

$$\text{Datensatz (1)} : y^{(1)} = \left(y_1^{(1)}, y_2^{(1)}, ..., y_n^{(1)}\right) \text{ mit } \bar{y}_n^{(1)} = \frac{1}{n}\sum_{i=1}^n y_i^{(1)}$$

$$\text{Datensatz (2)} : y^{(2)} = \left(y_1^{(2)}, y_2^{(2)}, ..., y_n^{(2)}\right) \text{ mit } \bar{y}_n^{(2)} = \frac{1}{n}\sum_{i=1}^n y_i^{(2)}$$

$$\text{Datensatz (3)} : y^{(3)} = \left(y_1^{(3)}, y_2^{(3)}, ..., y_n^{(3)}\right) \text{ mit } \bar{y}_n^{(3)} = \frac{1}{n}\sum_{i=1}^n y_i^{(3)}$$

$$\text{Datensatz (4)} : y^{(4)} = \left(y_1^{(4)}, y_2^{(4)}, ..., y_n^{(4)}\right) \text{ mit } \bar{y}_n^{(4)} = \frac{1}{n}\sum_{i=1}^n y_i^{(4)}$$

$$\text{Datensatz (5)} : y^{(5)} = ...$$

Um die Qualität statistischer Methoden zu beurteilen betrachtet die Frequentistische Statistik deshalb die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken unter Annahme von $\upsilon_1, ..., \upsilon_n \sim p_\theta$. Was zum Beispiel ist die Verteilung der $\bar{y}_n^{(1)}$, $\bar{y}_n^{(2)}$, $\bar{y}_n^{(3)}$, $\bar{y}_n^{(4)}$, ... also die Verteilung der Zufallsvariable $\bar{\upsilon}$?

Wenn eine statistische Methode im Sinne der Frequentistischen Standardannahmen "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 $y_1, ..., y_n$ vorliegen und nimmt an, dass es sich dabei um die Realisation einer Stichprobe handelt, zum Beispiel von $\upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2)$.

Man berechnet basierend auf dem Datensatz eine *Teststatistik*, zum Beispiel das anhand der Stichprobenvarianz und der Stichprobengröße normalisierte Stichprobenmittel $\sqrt{n}\bar{y}_n/s_n$.

Man fragt sich, wie wahrscheinlich es wäre, den beobachteten oder einen extremeren Wert der Teststatistik unter der Annahme eines *Nullmodels* zu observieren. Dabei meint man mit *Nullmodell* intuitiv ein Wahrscheinlichkeitsverteilungsmodell bei dem kein "interessanter Effekt" vorliegt, also zum Beispiel $\mu = 0$ gilt. Die Wahrscheinlichkeit ist wie immer Frequentistisch zu verstehen, d.h. als idealisierte relative Häufigkeit, wenn man viele Stichprobenrealisationen des Nullmodels generieren würde.

Ist die betrachtete Wahrscheinlichkeit dafür, den beobachteten oder einen extremeren Wert der Teststatistik unter Annahme des Nullmodells zu observieren groß, so sagt man sich "Nunja, dann ist es wohl ganz plausibel, dass das Nullmodel 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 Nullmodel die Daten generiert hat". Im Wissenschaftsjargon spricht man von einem "signifikanten Ergebnis".

Wie immer in der Frequentistischen Statistik weiß man nach Durchführung dieser Prozedur nicht, ob im vorliegenden Fall nun wirklich das Nullmodel oder ein anderes Modell die Daten generiert hat, sondern man weiß nur, wie oft man bei dieser Prozedur im Mittel richtig oder falsch liegen würde, wenn alle Annahmen zuträfen und man diese Prozedur sehr oft wiederholen würde.

Grundlegende Definitionen

Einstichproben-T-Test

p-Werte

Konfidenzintervalle und Hypothesentests

Anwendungsbeispiel

Selbstkontrollfragen

Grundlegende Definitionen

Einstichproben-T-Test

p-Werte

Konfidenzintervalle und Hypothesentests

Anwendungsbeispiel

Selbstkontrollfragen

Definition (Statistische Hypothesen und Testszenario)

$\upsilon_1, ..., \upsilon_n \sim p_\theta$ sei eine Stichprobe mit WMF oder WDF p_θ, $\mathcal{Y}$ sei der Ergebnisraum des Zufallsvektors $\upsilon := (\upsilon_1, ..., \upsilon_n)$, und Θ sei der Parameterraum des zugrundeliegenden statistischen Modells. Weiterhin sei $\{\Theta_0, \Theta_1\}$ eine Partition des Parameterraumes, so dass $\Theta = \Theta_0 \cup \Theta_1$ und $\Theta_0 \cap \Theta_1 = \emptyset$ gelten. Eine *statistische Hypothese* ist dann eine Aussage über den wahren, aber unbekannten, Parameterwert θ in Hinblick auf die Untermengen Θ_0 und Θ_1 des Parameterraums. Speziell werden die Aussagen

- $\theta \in \Theta_0$ als *Nullhypothese* H_0
- $\theta \in \Theta_1$ als *Alternativhypothese* H_1

bezeichnet. Die Einheit aus Stichprobe, Ergebnisraum, Parameterraum und Hypothesen wird im Folgenden als *Testszenario* bezeichnet.

Definition (Einfache und zusammengesetzte Hypothesen)

Für statistische Hypothesen $\Theta_i, i = 0, 1$ gilt:

- Enthält Θ_i nur ein einziges Element, so heißt Θ_i *einfach*.
- Enthält Θ_i mehr als ein Element, so heißt Θ_i *zusammengesetzt*.

Bemerkungen

- Die Nullhypothese $\Theta_0 = \{0\}$ ist ein Beispiel für eine einfache Hypothese.
- Bei einer einfachen Hypothese ist die Wahrscheinlichkeitsverteilung von υ genau festgelegt.
- Bei einer zusammengesetzten Hypothese ist nur die Verteilungsklasse von υ festgelegt.

Definition (Einseitige und zweiseitige Hypothesen)

$\Theta := \mathbb{R}$ sei ein eindimensionaler Parameterraum und θ_0 sei ein Element von Θ. Dann werden zusammengesetzte Nullhypothesen der Form

$$\Theta_0 :=]-\infty, \theta_0] \text{ oder } \Theta_0 := [\theta_0, \infty[\tag{1}$$

einseitige Nullhypothesen genannt und auch in der Form

$$H_0 : \theta \le \theta_0 \text{ oder } H_0 : \theta \ge \theta_0 \tag{2}$$

geschrieben. Die entsprechenden Alternativhypothesen haben dabei die Form

$$\Theta_1 :=]\theta_0, \infty[\text{ oder } \Theta_1 :=]-\infty, \theta_0[\text{ bzw. } H_1 : \theta > \theta_0 \text{ oder } H_1 : \theta < \theta_0. \tag{3}$$

Bei einer einfachen Nullhypothese der Form

$$\Theta_0 := \{\theta_0\} \text{ bzw. } H_0 : \theta = \theta_0 \tag{4}$$

wird die Alternativhypothese

$$\Theta_1 := \Theta \setminus \{\theta_0\} \text{ bzw. } H_1 : \theta \neq \theta_0 \Leftrightarrow \Theta_1 :=]-\infty, \theta_0[\,\cup\,]\theta_0, \infty[\tag{5}$$

zweiseitige Alternativhypothese genannt.

Definition (Test)

In einem Testszenario ist ein *Test* ϕ eine Abbildung aus dem Ergebnisraum $\mathcal{Y}$ nach $\{0, 1\}$,

$$\phi : \mathcal{Y} \to \{0, 1\}, y \mapsto \phi(y). \tag{6}$$

Dabei repräsentiert

- $\phi(y) = 0$ den Vorgang des Nichtablehnens der Nullhypothese.
- $\phi(y) = 1$ den Vorgang des Ablehnens der Nullhypothese.

Bemerkung

- Weil y eine Realisation von υ ist, ist $\phi(y)$ eine Realisation von $\phi(\upsilon)$.

Definition (Standardtest)

Ein *Standardtest* ist definiert durch die Verkettung einer *Teststatistik*

$$\gamma : \mathcal{Y} \to \mathbb{R} \tag{7}$$

und einer *Entscheidungsregel*

$$\delta : \mathbb{R} \to \{0, 1\}. \tag{8}$$

Ein Standardtest kann also geschrieben werden als

$$\phi := \delta \circ \gamma : \mathcal{Y} \to \{0, 1\}. \tag{9}$$

Bemerkungen

- Weil y eine Realisation von υ ist, ist $\gamma(y) \in \mathbb{R}$ eine Realisation von $\gamma(\upsilon)$.
- Weil $\gamma(y)$ eine Realisation von $\gamma(\upsilon)$ ist, ist $(\delta \circ \gamma)(y)$ eine Realisation von $(\delta \circ \gamma)(\upsilon)$.
- Wir betrachten in der Folge nur Standardtests.

Definition (Kritischer Bereich)

Die Untermenge K des Ergebnisraums des Zufallsvektors $\upsilon := (\upsilon_1, ..., \upsilon_n)$, für die ein Test den Wert 1 annimmt, heißt *kritischer Bereich* des Tests,

$$K := \{y \in \mathcal{Y} | \phi(y) = 1\} \subset \mathcal{Y}. \tag{10}$$

Bemerkungen

- Die Ereignisse $\{\phi(\upsilon) = 1\}$ und $\{\upsilon \in K\}$ sind äquivalent.
- Die Ereignisse $\{\phi(\upsilon) = 1\}$ und $\{\upsilon \in K\}$ haben die gleiche Wahrscheinlichkeit.

Definition (Ablehnungsbereich)

Die Untermenge A des Ergebnisraums einer Teststatistik, für die der Test den Wert 1 annimmt, heißt *Ablehnungsbereich* des Tests,

$$A := \{\gamma(y) \in \mathbb{R} | \phi(y) = 1\} \subset \mathbb{R}. \tag{11}$$

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 $k \in \mathbb{R}$ abhängt. Speziell ist

- ein *einseitiger kritischer Wert-basierter Test* von der Form

$$\phi : \mathcal{Y} \to \{0,1\}, y \mapsto \phi(y) := 1_{\{\gamma(y) \geq k\}} = \begin{cases} 1 & \gamma(y) \geq k \\ 0 & \gamma(y) < k \end{cases} \tag{12}$$

- ein *zweiseitiger kritischer Wert-basierter Test* von der Form

$$\phi : \mathcal{Y} \to \{0,1\}, y \mapsto \phi(y) := 1_{\{|\gamma(y)| \geq k\}} = \begin{cases} 1 & |\gamma(y)| \geq k \\ 0 & |\gamma(y)| < k \end{cases} \tag{13}$$

Bemerkung

- Wir betrachten in der Folge nur kritischer Wert-basierte Tests.

Definition (Richtige Testentscheidungen und Testfehler)

Das Nichtablehnen der Nullhypothese, wenn die Nullhypothese zutrifft werden, sowie das Ablehnen der Nullhypothese, wenn die Nullhypothese nicht zutrifft, werden *richtige Testentscheidungen* genannt. Es können weiterhin zwei Arten von Testfehlern auftreten: das Ablehnen der Nullhypothese, wenn die Nullhypothese zutrifft, heißt *Typ I Fehler*, das Nichtablehnen der Nullhypothese, wenn die Alternativhypothese zutrifft, heißt *Typ II Fehler*.

Die untenstehende Graphik gibt eine Übersicht.

Wahrer Parameterwert	Testentscheidung	
	$\phi(y) = 0$	$\phi(y) = 1$
$\theta \in \Theta_0$	Richtige Entscheidung	Typ I Fehler
$\theta \in \Theta_1$	Typ II Fehler	Richtige Entscheidung

Definition (Testgütefunktion)

Für einen Test ϕ ist die *Testgütefunktion* definiert als

$$q_\phi : \Theta \to [0,1], \theta \mapsto q_\phi(\theta) := \mathbb{P}_\theta(\phi = 1). \quad (14)$$

Für $\theta \in \Theta_1$ heißt q_ϕ auch *Powerfunktion* oder *Trennschärfefunktion*.

Bemerkungen

- Wir verzichten hier und im Folgenden auf die explizite Notation der Abhängigkeit von ϕ von υ.
- $\mathbb{P}_\theta$ bezeichnet die Verteilung von ϕ unter der Annahme $\upsilon_1, ..., \upsilon_n \sim p_\theta$.
- Es gilt $\mathbb{P}_\theta(\phi = 1) = \mathbb{P}_\theta(\upsilon \in K) = \mathbb{P}_\theta(\gamma \in A)$
- Für jedes $\theta \in \Theta$ liefert q_ϕ die Wahrscheinlichkeit, dass H_0 durch ϕ abgelehnt wird.
- Bei Poweranalysen betrachtet man q_ϕ als Funktion aller Testszenario und Testparameter.
- Ändert sich ϕ, z.B. weil sich der kritische Wert von ϕ ändert, dann ändert sich $q_\phi(\theta)$.
- Im Idealfall hätte man einen Test ϕ mit

$$q_\phi(\theta) = \mathbb{P}_\theta(\phi = 1) = 0 \text{ für } \theta \in \Theta_0 \text{ und } q_\phi(\theta) = \mathbb{P}_\theta(\phi = 1) = 1 \text{ für } \theta \in \Theta_1. \quad (15)$$

- Die Testentscheidung eines solchen ϕ wäre mit Wahrscheinlichkeit 1 richtig.

© 2023 Dirk Ostwald CC BY 4.0

Intuition zur Testkonstruktion

Im Idealfall hätte man einen Test ϕ mit

$$q_\phi(\theta) = \mathbb{P}_\theta(\phi = 1) = 0 \text{ für } \theta \in \Theta_0 \text{ und } q_\phi(\theta) = \mathbb{P}_\theta(\phi = 1) = 1 \text{ für } \theta \in \Theta_1. \tag{16}$$

$\Rightarrow$ 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

$$q_\phi(\theta) \le \alpha_0 \text{ für alle } \theta \in \Theta_0. \tag{17}$$

Eine konventionelle Wahl für sein solches α_0 ist zum Beispiel $\alpha_0 := 0.05$.

Unter allen Tests und statistischen Modellen, die Ungleichung (17) erfüllen, wird man dann einen Test oder ein statistisches Modell auswählen, so dass $q_\phi(\theta)$ für $\theta \in \Theta_1$ so groß wie möglich ist.

Dieses Vorgehen ist nicht alternativlos, man kann zum Beispiel auch lineare Kombinationen verschiedener Fehlerwahrscheinlichkeiten minimieren. Es ist aber das in der Anwendung populärste Vorgehen. Wir werden uns deshalb in der Folge auf dieses Vorgehen beschränken.

Das beschriebene Vorgehen motiviert die folgenden Definitionen der Begriffe des Level-α_0-Tests, des Signifikanzlevels α_0 (oft auch als *Signifikanzlevel* bezeichnet) und des Testumfangs α (auch als *effektives Niveau* bezeichnet).

Definition (Level-α_0-Test, Signifikanzlevel α_0, Testumfang α)

q_ϕ sei die Testgütefunktion eines Tests ϕ und es sei $\alpha_0 \in [0,1]$. Dann heißt ein Test ϕ, für den gilt, dass

$$q_\phi(\theta) \le \alpha_0 \text{ für alle } \theta \in \Theta_0 \tag{18}$$

ein *Level-α_0-Test* und man sagt, dass der Test das *Signifikanzlevel* α_0 hat. Die Zahl

$$\alpha := \max_{\theta \in \Theta_0} q_\phi(\theta) \in [0,1] \tag{19}$$

heißt der *Testumfang* von ϕ.

Bemerkungen

- α ist die größtmögliche Wahrscheinlichkeit für einen Typ I Fehler.
- Ein Test ist dann, und nur dann, ein Level-α_0-Test, wenn $\alpha \le \alpha_0$ gilt.
- Bei einer einfachen Nullhypothese gilt für den Testumfang, dass $\alpha = q_\phi(\theta_0) = \mathbb{P}_{\theta_0}(\phi = 1)$.

Typ I Fehlerwahrscheinlichkeit vs. Testumfang vs. Signifikanzlevel

Bei einfacher Θ_0 ist der Testumfang gleich der Wahrscheinlichkeit eines Typ I Fehlers

$$\alpha := \max_{\theta \in \Theta_0} q_\phi(\theta) = \max_{\theta \in \{\theta_0\}} q_\phi(\theta) = q_\phi(\theta_0) = \mathbb{P}_{\theta_0}(\phi = 1). \tag{20}$$

Bei zusammengesetzter Θ_0 gibt es je nach Wert von $\theta \in \Theta_0$ verschiedene Wahrscheinlichkeiten für einen Typ I Fehler. Die größte dieser Wahrscheinlichkeiten ist der Testumfang

$$\alpha := \max_{\theta \in \Theta_0} q_\phi(\theta) = \max_{\theta \in \Theta_0} \mathbb{P}_\theta(\phi = 1). \tag{21}$$

Ein Test hat Signifikanzlevel α_0, wenn der Testumfang kleiner oder gleich α_0 ist.

$$\alpha = \max_{\theta \in \Theta_0} q_\phi(\theta) \le \alpha_0 \tag{22}$$

Ein Test, bei dem das Signifikanzlevel größer als der Testumfang ist, heißt *konservativ*.

Ein Test, bei dem das Signifikanzlevel gleich dem Testumfang ist, heißt *exakt*.

Zur Wahl von Nullhypothese und Alternativhypothese

Das Vorgehen in der Testkonstruktion zunächst durch die Wahl eines Signifikanzlevels den Testumfang zu begrenzen und erst in einem zweiten Schritt dafür zu sorgen, dass die Wahrscheinlichkeit von $\phi = 1$ bei $\theta \in \Theta_1$ bei diesem Signifikanzlevel möglichst groß ist, induziert eine Asymmetrie in der Behandlung von Null- und Alternativhypothese. Implizit wichtet man mit diesem Vorgehen Typ I Fehler als schwerwiegender als Typ II Fehler.

Dies wiederum impliziert eine mögliche Strategie zur Festlegung von Null- und Alternativhypothese: Die Nullhypothese ist die Hypothese, hinsichtlich deren assoziierter Testentscheidung man eher keinen Fehler machen möchte bzw. deren Fehlerwahrscheinlichkeit man primär kontrollieren möchte.

In der wissenschaftlichen Anwendung ist es Standard, die falsche Konfirmation der eigenen Theorie als einen schwerwiegenderen Fehler als die falsche Ablehnung der eigenen Theorie zu werten.

Die falsche Konfirmation der eigenen Theorie sollte also ein Typ I Fehler, das falsche Ablehnen der eigenen Theorie ein Typ II Fehler sein.

Damit die falsche Konfirmation der eigenen Theorie einen Typ I Fehler, also das Ablehnen von H_0 bei Zutreffen von H_0, darstellt, muss die eigene Theorie als Alternativhypothese aufgestellt werden. Die Alternativhypothese fälschlichweise Abzulehnen wird damit ein Typ II Fehler.

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), Horvath et al. (2021))

Oberflächlich betrachtet liefern Hypothesentests einfache binäre Aussagen der Form "Die Hypothese (Theorie) ist gegeben die Evidenz abzulehnen oder zu akzeptieren". Solche Aussagen sind im Entscheidungskontext hilfreich, denn es muss etwas passieren, also eine Entscheidung getroffen werden. In der Wissenschaft, also der menschlichen Kommunikationsstruktur über die Beschaffenheit der Welt, muss aber nichts final entschieden, sondern nur das Maß an Unsicherheit über den gerade vorherrschenden Theoriestand quantifiziert und kommuniziert werden. Generell sollten Fragestellungen der Grundlagenwissenschaften deshalb gerade nicht als Entscheidungsprobleme formuliert werden.

Trotz landläufiger Meinung das Bayesianische Herangehensweisen wie Positive Predictive Values oder Bayes Factors hier irgendwie besser sind, ist dem nicht so, so lange die mit einer gewissen Modellpräferenz assoziierte Unsicherheit nicht klar mitkommuniziert wird.

Und trotz alledem ist Frequentistisches Hypothesentesten in der Wissenschaftscommunity weiterhin sehr populär (wenn auch vermutlich nicht immer ganz verstanden) und sollte deshalb im Rahmen eines wissenschaftlichen Studiums wie der Psychologie intellektuell durchdrungen werden.

Grundlegende Definitionen

Einstichproben-T-Test

p-Werte

Konfidenzintervalle und Hypothesentests

Anwendungsbeispiel

Selbstkontrollfragen

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Pre-BDI

n = 12

Post-BDI

BDI-II Fragebogen

Name | Alter | Geschlecht m / w | Datum

Anleitung: Dieser Fragebogen enthält 21 Gruppen von Aussagen. Bitte lesen Sie jede dieser Gruppen von Aussagen sorgfältig durch und suchen Sie sich dann in jeder Gruppe **eine Aussage** heraus, die am besten beschreibt, wie Sie sich **in den letzten zwei Wochen, einschließlich heute, gefühlt haben**. Kreuzen Sie die Zahl neben der Aussage an, die Sie sich herausgesucht haben (0, 1, 2 oder 3). Falls in einer Gruppe mehrere Aussagen gleichermaßen auf Sie zutreffen, kreuzen Sie die Aussage mit der höheren Zahl an. Achten Sie bitte darauf, dass Sie in jeder Gruppe nicht mehr als eine Aussage ankreuzen, das gilt auch für Gruppe 16 (Veränderungen der Schlafgewohnheiten) oder Gruppe 18 (Veränderungen des Appetits).

1.) Traurigkeit
0 Ich bin nicht traurig.
1 Ich bin oft traurig.
2 Ich bin ständig traurig.
3 Ich bin so traurig oder unglücklich, dass ich es nicht aushalte.

2.) Pessimismus
0 Ich sehe nicht mutlos in die Zukunft.
1 Ich sehe mutloser in die Zukunft als sonst.
2 Ich bin mutlos und erwarte nicht, dass meine Situation besser wird.
3 Ich glaube, dass meine Zukunft hoffnungslos ist und nur noch schlechter wird.

3.) Versagensgefühle
0 Ich fühle mich nicht als Versager.
1 Ich habe häufiger Versagensgefühle.
2 Wenn ich zurückblicke, sehe ich eine Menge Fehlschläge.
3 Ich habe das Gefühl, als Mensch ein völliger Versager zu sein.

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.

5.) Schuldgefühle
0 Ich habe keine besonderen Schuldgefühle.
1 Ich habe oft Schuldgefühle wegen Dingen, die ich getan habe oder hätte tun sollen.
2 Ich habe die meiste Zeit Schuldgefühle.
3 Ich habe ständig Schuldgefühle.

6.) Bestrafungsgefühle
0 Ich habe nicht das Gefühl, für etwas bestraft zu sein.
1 Ich habe das Gefühl, vielleicht bestraft zu werden.
2 Ich erwarte, bestraft zu werden.
3 Ich habe das Gefühl, bestraft zu sein.

7.) Selbstablehnung
0 Ich halte von mir genauso viel wie immer.
1 Ich habe Vertrauen in mich verloren.
2 Ich bin von mir enttäuscht.
3 Ich lehne mich völlig ab.

8.) Selbstvorwürfe
0 Ich kritisiere oder tadle mich nicht mehr als sonst.
1 Ich bin mir gegenüber kritischer als sonst.
2 Ich kritisiere mich für all meine Mängel.
3 Ich gebe mir die Schuld für alles Schlimme, was passiert.

9.) Selbstmordgedanken
0 Ich denke nicht daran, mir etwas anzutun.
1 Ich denke manchmal an Selbstmord, aber ich würde es nicht tun.
2 Ich möchte mich am liebsten umbringen.
3 Ich würde mich umbringen, wenn ich die Gelegenheit dazu hätte.

10.) Weinen
0 Ich weine nicht öfter als früher.
1 Ich weine jetzt mehr als früher.
2 Ich weine beim geringsten Anlass.
3 Ich möchte gern weinen, aber ich kann nicht.

PEARSON PsychCorp

Summe Seite 1:

⇒ Pre-Post BDI Score Reduktion

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Für die Pre-Post BDI Score Reduktion υ_i der $i$ten von n Patient:innen legen wir das Modell

$$\upsilon_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, ..., n \tag{23}$$

zugrunde. Dabei wird die Pre-Post BDI Reduktion υ_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

$$\upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2). \tag{24}$$

$\Rightarrow$ Wir sind nun am Testen einer statistischen Hypothese hinsichtlich μ interessiert.

$\Rightarrow$ Das betrachtete Anwendungsszenario ist dann ein Beispiel für einen **Einstichproben-T-Test**.

Einstichproben-T-Test

Mögliche Hypothesen im Einstichproben-T-Test-Szenario

Einfache Nullhypothese, einfache Alternativhypothese $H_0 : \mu = \mu_0, H_1 : \mu = \mu_1$

- Theoretisch wichtiges Szenario (Neymann-Pearson Lemma)
- Praktische Relevanz eher gering

Einfache Nullhypothese, zusammengesetzte Alternativhypothese $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$

- Zweiseitiger Einstichproben-T-Test mit ungerichteter Hypothese
- Ungerichtete Fragestellung nach einem Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese $H_0 : \mu \leq \mu_0, H_1 : \mu > \mu_0$

- Einseitiger Einstichproben-T-Test mit gerichteter Hypothese
- Gerichtete Fragestellung nach einem positiven Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese $H_0 : \mu \geq \mu_0, H_1 : \mu < \mu_0$

- Gerichtete Fragestellung nach einem negativen Unterschied
- Qualitativ äquivalente Theorie zum umgekehrten Fall

Einstichproben-T-Test

Hier betrachtete Hypothese im Einstichproben-T-Test-Szenario

Einfache Nullhypothese, einfache Alternativhypothese $H_0 : \mu = \mu_0, H_1 : \mu = \mu_1$

- Theoretisch wichtiges Szenario (Neymann-Pearson Lemma)
- Praktische Relevanz eher gering

Einfache Nullhypothese, zusammengesetzte Alternativhypothese $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$

- Zweiseitiger Einstichproben-T-Test mit ungerichteter Hypothese
- Ungerichtete Fragestellung nach einem Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese $H_0 : \mu \leq \mu_0, H_1 : \mu > \mu_0$

- Einseitiger Einstichproben-T-Test mit gerichteter Hypothese
- Gerichtete Fragestellung nach einem positiven Unterschied

Zusammengesetzte Nullhypothese/Alternativhypothese $H_0 : \mu \geq \mu_0, H_1 : \mu < \mu_0$

- Gerichtete Fragestellung nach einem negativen Unterschied
- Qualitativ äquivalente Theorie zum umgekehrten Fall

Gliederung

(1) Statistisches Modell und Testhypothesen

(2) Definition und Analyse der Teststatistik

(3) Definition des Tests

(4) Analyse der Testgütefunktion

(5) Testumfangkontrolle

(6) Analyse der Powerfunktion

Einstichproben-T-Test | (1) Statistischem Modell und Testhypothesen

Das **Statistische Modell des Einstichproben-T-Tests** ist definiert als

$$\upsilon_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ für } i = 1, ..., n \tag{25}$$

wobei

- $\upsilon_i, i = 1, ..., n$ beobachtbare Zufallsvariablen,
- μ den wahren, aber unbekannten, Erwartungswertparameter der Stichprobenvariablen,
- $\varepsilon_i, i = 1, ..., n$ unabhängige normalverteilte nicht-beobachtbare Zufallsvariablen und
- $\sigma^2 > 0$ den Varianzparameter der ε_i

bezeichnen. Wie unten (erneut) gezeigt, ist dieses Modell äquivalent zum Normalverteilungsmodell

$$\upsilon = \upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2), \tag{26}$$

also der Annahme unabhängig und identisch normalverteilter Stichprobenvariablen mit Erwartungswertparameter μ und Varianzparameter σ^2. Für ein μ_0 betrachten wir die einfache **Nullhypothese** und die zusammengesetzte **Alternativhypothese**

$$H_0 : \mu = \mu_0 \Leftrightarrow \Theta_0 := \{\mu_0\} \text{ und } H_1 : \mu \neq \mu_0 \Leftrightarrow \Theta_1 := \mathbb{R} \setminus \{\mu_0\}, \tag{27}$$

respektive. Bezogen auf das Anwendungsbeispiel ist hier $\mu_0 := 0$ von Interesse:

- $H_0 : \mu = 0$ entspricht der Hypothese keines Effekts der Therapie auf die BDI Score Reduktion.
- $H_1 : \mu \neq 0$ entspricht der Hypothese eines Effekts der Therapie auf die BDI Score Reduktion.

Beweis der Äquivalenz der Modellformen

Die Äquivalenz beider Modellformen folgt direkt aus der Transformation normalverteilter Zufallsvariablen durch linear-affine Funktionen (cf. (8) Transformationen der Normalverteilung). Speziell gilt im vorliegenden Fall für $\varepsilon_i \sim N(0, \sigma^2)$, dass

$$\upsilon_i = f(\varepsilon_i) \text{ mit } f : \mathbb{R} \to \mathbb{R}, \epsilon_i \mapsto f(\epsilon_i) := \epsilon_i + \mu. \tag{28}$$

Mit den WDF Transformationstheorem bei linear-affinen Abbildungen folgt dann

$$\begin{aligned}
p_{\upsilon_i}(y_i) &= \frac{1}{|1|} p_{\varepsilon_i}\left(\frac{y_i - \mu}{1}\right) \\
&= N\left(y_i - \mu; 0, \sigma^2\right) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu - 0)^2\right) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right) \\
&= N(y_i; \mu, \sigma^2),
\end{aligned} \tag{29}$$

also $\upsilon_i \sim N(\mu, \sigma^2)$.

Definition (Einstichproben-T-Teststatistik)

$\upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2)$ sei die Stichprobe eines Normalverteilungmodells, $\bar{\upsilon}$ bezeichne das Stichprobenmittel, S bezeichne die Stichprobenstandardabweichung und es sei $\mu_0 \in \mathbb{R}$. Dann ist die *T-Teststatistik* definiert als

$$T := \sqrt{n}\left(\frac{\bar{\upsilon} - \mu_0}{S}\right). \tag{30}$$

Bemerkungen

- Im Gegensatz zur T-Konfidenintervallstatistik muss bei der T-Teststatistik nicht $\mu_0 = \mu$ gelten.
- Intuitiv kann die T-Teststatistik als mit der Stichprobengröße (Evidenz) gewichtetes Verhältnis von Signal (sytematischer Variabilität) zu Rauschen (unsystematischer Variabilität) verstanden werden:

$$\sqrt{\text{Stichprobengröße}}\left(\frac{\text{Signal}}{\text{Rauschen}}\right) = \sqrt{n}\left(\frac{\bar{\upsilon} - \mu_0}{S}\right) \tag{31}$$

- Die T-Teststastitik ist eine skalare Deskription des Effekt vs. Variabilität Verhältnisses eines Datensatzes.
- In der T-Teststatistik wird die Effektgröße in Einheiten der Stichprobenstandardabweichung gemessen:
  - $T = 1 \Leftrightarrow \sqrt{n}(\bar{\upsilon} - \mu_0) = 1S$
  - $T = 2 \Leftrightarrow \sqrt{n}(\bar{\upsilon} - \mu_0) = 2S$

Definition (Nichtzentrale t-Zufallsvariable)

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

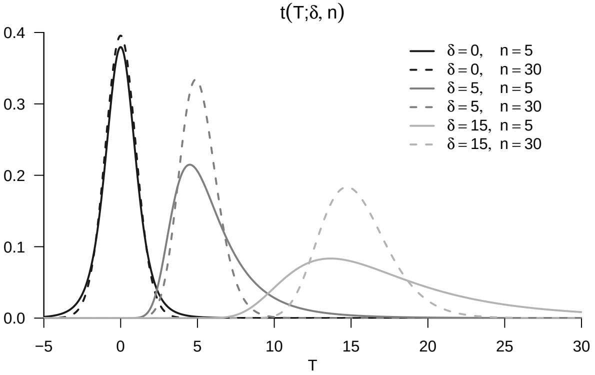
$$p : \mathbb{R} \to \mathbb{R}_{>0}, t \mapsto p(t) := \frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n}{2}\right)(n\pi)^{\frac{1}{2}}} \times \int_0^\infty \tau^{\frac{n-1}{2}} \exp\left(-\frac{\tau}{2}\right) \exp\left(-\frac{1}{2}\left(t\left(\frac{\tau}{n}\right)^{\frac{1}{2}} - \delta\right)^2\right) d\tau. \quad (32)$$

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)

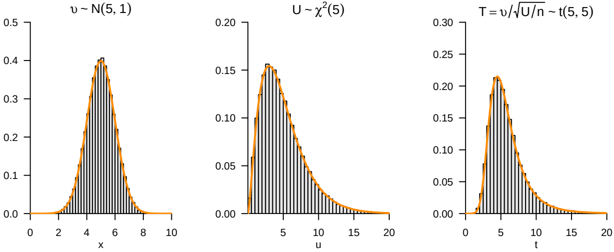
$\upsilon \sim N(\mu, 1)$ sei eine normalverteilte Zufallsvariable, $U \sim \chi^2(n)$ sei eine χ^2 Zufallsvariable mit Freiheitsgradparameter n, und υ und U seien unabhängige Zufallsvariablen. Dann ist die Zufallsvariable

$$T := \frac{\upsilon}{\sqrt{U/n}} \tag{33}$$

eine nichtzentrale t-Zufallsvariable mit Nichtzentralitätsparameter μ und Freiheitsgradparameter n, also $T \sim t(\mu, n)$.

Bemerkung

- Wir verzichten auf einen Beweis.



Theorem (Verteilung der T-Teststatistik)

$\upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2)$ sei die Stichprobe eines Normalverteilungmodells, $\bar{\upsilon}$ sei das Stichprobenmittel, S sei die Stichprobenstandardabweichung, und es sei $\mu_0 \in \mathbb{R}$. Dann ist die T-Teststatistik

$$T := \sqrt{n}\left(\frac{\bar{\upsilon} - \mu_0}{S}\right) \tag{34}$$

eine nichtzentrale t-Zufallsvariable mit Nichtzentralitätsparameter

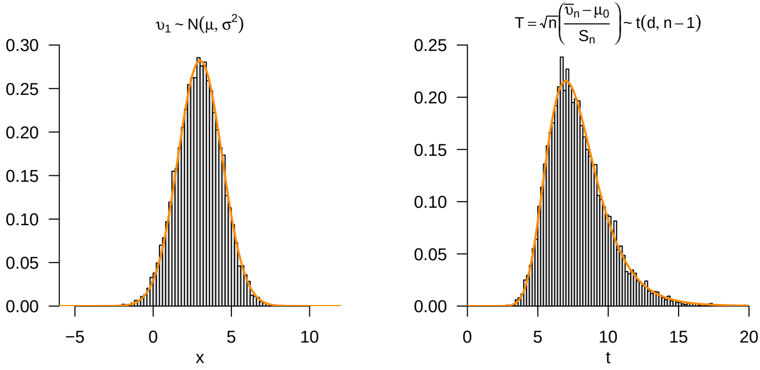
$$d = \sqrt{n}\left(\frac{\mu - \mu_0}{\sigma}\right) \tag{35}$$

und Freiheitsgradparameter $n - 1$, es gilt also $T \sim t(d, n-1)$

Bemerkung

- Wir verzichten auf einen Beweis.

T-Teststatistik bei $n = 12, \mu = 3, \sigma^2 = 2, \mu_0 = 0$



Vor dem Hintergrund des statistischen Modells des Einstichproben-T-Tests betrachten wir die einfache Nullhypothese und die zusammengesetzte Alternativhypothese

$$H_0 : \mu = \mu_0 \Leftrightarrow \Theta_0 := \{\mu_0\} \text{ und } H_1 : \mu \neq \mu_0 \Leftrightarrow \Theta_1 := \mathbb{R} \setminus \{\mu_0\}, \tag{36}$$

respektive, sowie die oben definierte Einstichproben-T-Teststatistik

$$T := \sqrt{n}\left(\frac{\bar{\upsilon} - \mu_0}{S}\right). \tag{37}$$

Wir definieren nun den zweiseitigen Einstichproben-T-Test

$$\phi(\upsilon) := 1_{\{|T| \geq k\}} = \begin{cases} 1 & |T| \geq k \\ 0 & |T| < k \end{cases}. \tag{38}$$

Theorem (Testgütefunktion)

ϕ sei der im obend definierte Test. Dann ist die Testgütefunktion von ϕ gegeben durch

$$q_\phi : \mathbb{R} \to [0,1], \mu \mapsto q_\phi(\mu) := 1 - \Psi(k; d_\mu, n-1) + \Psi(-k; d_\mu, n-1) \tag{39}$$

wobei $\Psi(\cdot; d_\mu, n-1)$ die KVF der nichtzentralen t-Verteilung mit Nichtzentralitätsparameter

$$d_\mu := \sqrt{n}\left(\frac{\mu - \mu_0}{\sigma}\right) \tag{40}$$

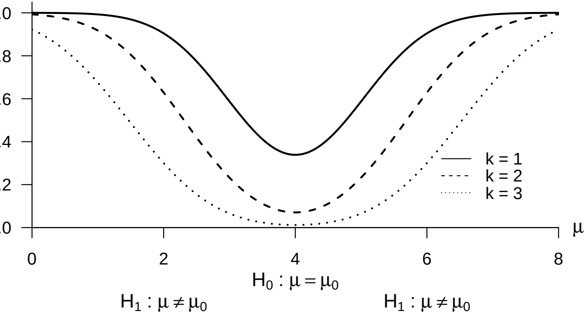
und Freiheitsgradparameter $n-1$ bezeichnet.

Bemerkungen

- Wir visualisieren die Testgütefunktion unten in Abhängigkeit von k.

Testgütefunktion q_ϕ für $\sigma^2 = 9, \mu_0 = 4, n = 12$ und $k = 1, 2, 3$.

$$q_\phi(\mu) = \mathbb{P}_\mu(\phi = 1)$$



Beweis

Die Testgütefunktion des betrachteten Test im vorliegenden Testszenario ist definiert als

$$q_\phi : \mathbb{R} \to [0,1], \mu \mapsto q_\phi(\mu) := \mathbb{P}_\mu(\phi = 1). \tag{41}$$

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 T-Teststatistik

$$T := \sqrt{n}\left(\frac{\bar{\upsilon} - \mu_0}{S}\right) \tag{42}$$

unter der Annahme $\upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2)$ nach einer nichtzentralen t-Verteilung $t(d_\mu, n-1)$ mit Nichtzentralitätsparameter

$$d_\mu := \sqrt{n}\left(\frac{\mu - \mu_0}{\sigma}\right) \tag{43}$$

verteilt ist. Der Ablehnungsbereich des zweiseitigen T-Tests ergibt sich zu

$$A =]-\infty, -k] \cup]k, \infty[. \tag{44}$$

Beweis (fortgeführt)

Mit diesem Ablehungsbereich ergibt sich dann

$$
\begin{aligned}
q_\phi(\mu) &= \mathbb{P}_\mu(\phi = 1) \\
&= \mathbb{P}_\mu\left(T \in]-\infty, -k] \cup]k, \infty[\right) \\
&= \mathbb{P}_\mu\left(T \in]-\infty, -k]\right) + \mathbb{P}_\mu\left(T \in [k, \infty[\right) \\
&= \mathbb{P}_\mu(T \le -k) + \mathbb{P}_\mu(T \ge k) \\
&= \mathbb{P}_\mu(T \le -k) + (1 - \mathbb{P}_\mu(T \le k)) \\
&= 1 - \mathbb{P}_\mu(T \le k) + \mathbb{P}_\mu(T \le -k) \\
&= 1 - \Psi(k; d_\mu, n-1) + \Psi(-k; d_\mu, n-1),
\end{aligned}
\tag{45}
$$

wobei $\Psi(\cdot; d_\mu, n-1)$ die KVF der nichtzentralen T-Verteilung mit Nichtzentralitätsparameter d_μ und Freiheitsgradparameter $n-1$ bezeichnet.

□

Theorem (Testumfangkontrolle)

ϕ sei der oben definierte Test. Dann ist ϕ ein Level-α_0-Test mit Testumfang α_0, wenn der kritische Wert definiert ist durch

$$k_{\alpha_0} := \Psi^{-1}\left(1 - \frac{\alpha_0}{2}; n-1\right), \tag{46}$$

wobei $\Psi^{-1}(\cdot; n-1)$ die inverse KVF der t-Verteilung mit Freiheitsgradparameter $n-1$ ist.

Beweis

Damit der betrachtete Test ein Level-α_0-Test ist, muss bekanntlich $q_\phi(\mu) \le \alpha_0$ für alle $\mu \in \{\mu_0\}$, also hier $q_\phi(\mu_0) \le \alpha_0$, 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

$$\begin{aligned} q_\phi(\mu_0) &= 1 - \Psi(k; d_{\mu_0}, n-1) + \Psi(-k; d_{\mu_0}, n-1) \\ &= 1 - \Psi(k; 0, n-1) + \Psi(-k; 0, n-1) \\ &= 1 - \Psi(k; n-1) + \Psi(-k; n-1), \end{aligned} \tag{47}$$

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.

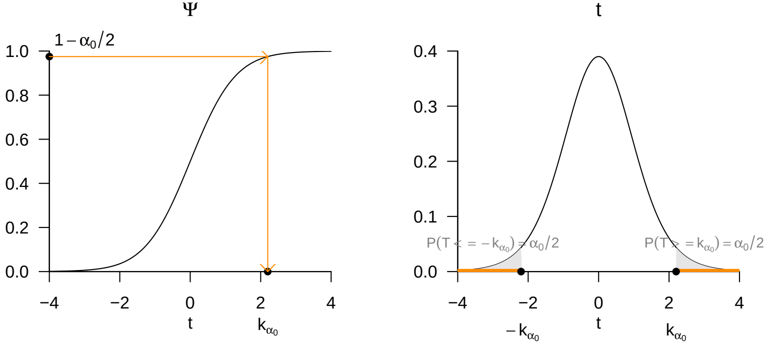
Beweis (fortgeführt)

Sei nun also $k := k_{\alpha_0}$. Dann gilt

$$
\begin{aligned}
q_\phi(\mu_0) &= 1 - \Psi(k_{\alpha_0}; n-1) + \Psi(-k_{\alpha_0}; n-1) \\
&= 1 - \Psi(k_{\alpha_0}; n-1) + (1 - \Psi(k_{\alpha_0}; n-1) \\
&= 2(1 - \Psi(k_{\alpha_0}; n-1)) \\
&= 2\left(1 - \Psi\left(\Psi^{-1}\left(1 - \frac{\alpha_0}{2}, n-1\right), n-1\right)\right) \\
&= 2\left(1 - 1 + \alpha_0/2\right) \\
&= \alpha_0,
\end{aligned}
\tag{48}
$$

wobei die zweite Gleichung mit der Symmetrie der t-Verteilung folgt. Es folgt also direkt, dass bei der Wahl von $k = k_{\alpha_0}$, $q_\phi(\mu_0) \le \alpha_0$ 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.

Wahl von $k_{\alpha_0} := \Psi^{-1}(1 - \frac{\alpha_0}{2}; n-1)$ mit $n = 12$, $\alpha_0 := 0.05$ und Ablehnungsbereich



Praktisches Vorgehen

- Man nimmt an, dass ein vorliegender Datensatz $y_1, ..., y_n$ eine Realisation von $\upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2)$ mit unbekannten Parametern μ und $\sigma^2 > 0$ ist.
- Man möchte entscheiden ob für ein $\mu_0 \in \mathbb{R}$ eher $H_0 : \mu = \mu_0$ oder $H_1 : \mu \neq \mu_0$ zutrifft.
- Man wählt ein Signifikanzlevel α_0 und bestimmt den zugehörigen Freiheitsgradparameter-abhängigen kritischen Wert k_{α_0}. Zum Beispiel gilt bei Wahl von $\alpha_0 := 0.05$ und $n = 12$, also Freiheitsgradparameter 11, dass $k_{0.05} = \Psi^{-1}(1 - 0.05/2; 11) \approx 2.20$ ist.
- Anhand von $n, \mu_0, \bar{\upsilon}$ und s_n berechnet man die Realisierung der T-Teststatistik

$$t := \sqrt{n}\left(\frac{\bar{y} - \mu_0}{s}\right) \tag{49}$$

- Wenn t größer-gleich k_{α_0} ist oder wenn t kleiner- gleich $-k_{\alpha_0}$ ist, lehnt man die Nullhypothese ab, andernfalls lehnt man sie nicht ab. Die oben entwickelte Theorie garantiert dann, dass man in höchstens $\alpha_0 \cdot 100$ von 100 Fällen die Nullhypothese fälschlicherweise ablehnt.

Simulation des praktischen Vorgehens

```
# Modellparameter
n         = 12                                        # Anzahl der Datenpunkte
mu        = 0                                         # wahrer, aber unbekannter, Erwartungswertparameter
sigsqr    = 2                                         # wahrer, aber unbekannter, Varianzmatrixparameter

# Testparameter
mu_0      = 0                                         # H_0 Hypothesenparameter, hier \mu = \mu_0
alpha_0   = 0.05                                      # Signifikanzlevel
k_alpha_0 = qt(1-alpha_0/2,n-1)                       # kritischer Wert

# Simulation der Testumfangkontrolle
set.seed(1)                                           # Random number generator seed
nsim      = 1e6                                       # 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)               # T-Teststatistik
    if(abs(Tee) > k_alpha_0){                         # Test 1_{|t| >= k_alpha_0}
        phi[j] = 1                                    # Ablehnen von H_0
    } else {
        phi[j] = 0                                    # Nicht Ablehnen von H_0
    }
}

# Ausgabe
cat("Kritischer Wert              =", k_alpha_0,
    "\nGeschätzter Testumfang alpha =", mean(phi))
```

```
> Kritischer Wert              = 2.2
> Geschätzter Testumfang alpha = 0.0498
```

Wir betrachten die Testgütefunktion

$$q_\phi : \mathbb{R} \to [0,1], \mu \mapsto q_\phi(\mu) := 1 - \Psi(k_{\alpha_0}; d_\mu, n-1) + \Psi(-k_{\alpha_0}; d_\mu, n-1) \qquad (50)$$

bei kontrolliertem Testumfang, also für $k_{\alpha_0} := \Psi^{-1}(1 - \alpha_0/2; n-1)$ mit festem α_0 als Funktion des Nichtzentralitätsparameters und des Stichprobenumfangs. Namentlich hängt hier k_{α_0} auch von n ab.

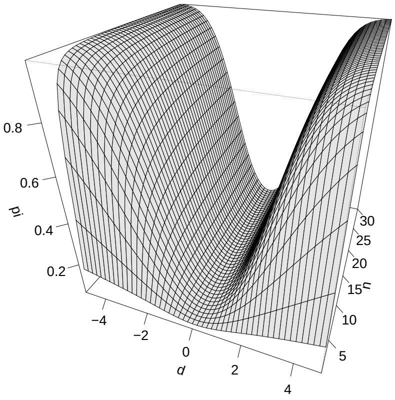
Es ergibt sich die bivariate reellwertige Funktion

$$\pi : \mathbb{R} \times \mathbb{N} \to [0,1], (d,n) \mapsto \pi(d,n) := 1 - \Psi(k_{\alpha_0}; d, n-1) + \Psi(-k_{\alpha_0}; d, n-1) \qquad (51)$$

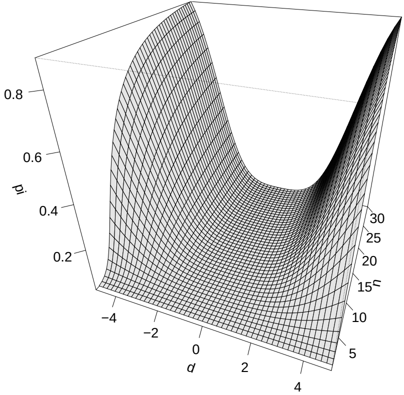
Bei festgelegten α_0 hängt die Powerfunktion des zweiseitigen T-Tests mit einfacher Nullhypothese also vom unbekannten Wert d und von der Stichprobengröße n ab. Wir visualisieren diese Abhängigkeiten untenstehend.

© 2023 Dirk Ostwald CC BY 4.0

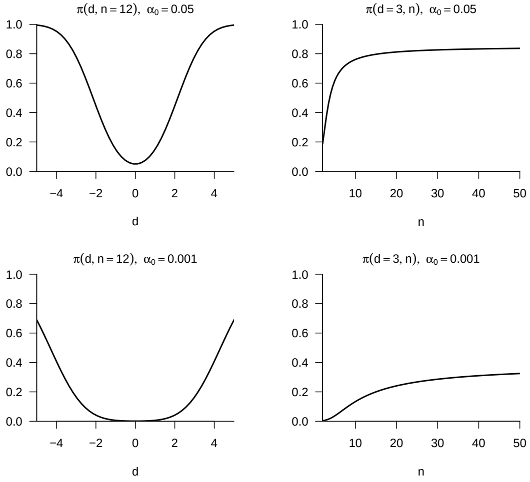
Powerfunktion für $\alpha_0 = 0.05$



Powerfunktion für $\alpha_0 = 0.001$



Powerfunktionen für $\mu_0 = 0$



Einstichproben-T-Test | (6) Analyse der Powerfunktion

Praktisches Vorgehen

Mit größerem n steigt die Powerfunktion des Tests an

- Ein großer Stichprobenumfang ist besser als ein kleiner Stichprobenumfang.
- Kosten für die Erhöhung des Stichprobenumfangs werden aber nicht berücksichtigt.

$\Rightarrow$ Die Theorie statistischer Hypothesentests ist nicht besonders lebensnah.

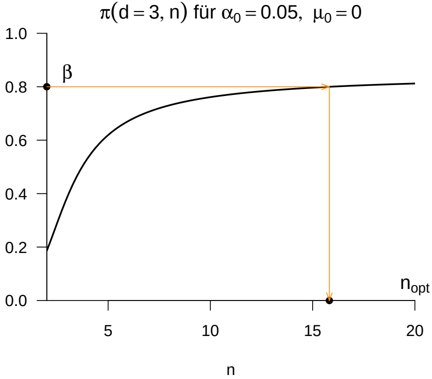
Die Powerfunktion hängt vom wahren, aber unbekannten, Parameterwert $d = \sqrt{n}(\mu - \mu_0)/\sigma$ ab.

$\Rightarrow$ Wenn man d schon kennen würde, würde man den Test nicht durchführen.

Generell wird folgendes Vorgehen favorisiert

- Man legt das Signifikanzlevel α_0 fest und evaluiert die Powerfunktion.
- Man wählt einen Mindestparameterwert d^*, den man mit $\pi(d, n) = \beta$ detektieren möchte.
- Ein konventioneller Wert ist $\beta = 0.8$.
- Man liest die für $\pi(d = d^*, n) = \beta$ nötige Stichprobengröße n ab.

Praktisches Vorgehen



Grundlegende Definitionen

Einstichproben-T-Test

p-Werte

Konfidenzintervalle und Hypothesentests

Anwendungsbeispiel

Selbstkontrollfragen

p-Werte

Motivation

- Es werde ein zweiseitiger Einstichproben-T-Test mit $n = 12$ und $\alpha_0 = 0.05$ durchgeführt.
    - H_0 wird abgelehnt, wenn $|T| \geq 2.20$.
- Nehmen wir an, es werde $t = 2.26$ beobachtet.
    - Das Testergebnis lautet "H_0 Ablehnen".
- Nehmen wir an, es werde $t = 3.81$ beobachtet.
    - Das Testergebnis lautet "H_0 Ablehnen".
- Der alleinige Bericht des Testergebnis supprimiert interessante Information.

$\Rightarrow$ Neben der Testumfangkontrolle durch z.B. $\alpha_0 = 0.05$ ist es daher üblich, alle Werte von α_0 anzugeben, für die ein Level-α_0-Test zum Ablehnen von H_0 führen würde.
    - Bei $t = 2.26$ würde H_0 für jedes α_0 mit $2.26 \geq \Psi^{-1}\left(1 - \frac{\alpha_0}{2}; 11\right)$ abgelehnt werden.
    - Bei $t = 3.81$ würde H_0 für jedes α_0 mit $3.81 \geq \Psi^{-1}\left(1 - \frac{\alpha_0}{2}; 11\right)$ abgelehnt werden.
- Das kleinste Signifikanzlevel α_0, bei dem man H_0 basierend auf einem Wert der Teststatistik ablehnen würde, wird *p-Wert* des Wertes der Teststatistik genannt.

Definition (p-Wert)

ϕ sei ein kritischer Wert-basierter Test. Der *p-Wert* ist das kleinste Signifikanzlevel α_0, bei welchem man die Nullhypothese basierend auf einem vorliegendem Wert der Teststatistik ablehnen würde.

Beispiel (Zweiseitiger Einstichproben-T-Test mit einfacher Nullhypothese)

- Bei $T = t$ würde H_0 für jedes α_0 mit $|t| \geq \Psi^{-1}\left(1 - \frac{\alpha_0}{2}; n-1\right)$ abgelehnt werden. Für diese α_0 gilt, wie unten gezeigt,
$$\alpha_0 \geq 2\mathbb{P}(T \geq |t|). \tag{52}$$
- Das kleinste $\alpha_0 \in [0,1]$ mit $\alpha_0 \geq 2\mathbb{P}(T \geq |t|)$ ist dann $\alpha_0 = 2\mathbb{P}(T \geq |t|)$, also folgt
$$\text{p-Wert} = 2\mathbb{P}(T \geq |t|) = 2(1 - \Psi(|t|; n-1)). \tag{53}$$
- Zum Beispiel ist bei $n = 12$ für $T = 2.26$ der p-Wert 0.045, für $T = -2.26$ ist der p-Wert auch 0.045, für $T = 3.81$ ist der p-Wert 0.003 und für $T = -3.81$ ist der p-Wert auch 0.003.

Beispiel (Zweiseitiger Einstichproben-T-Test mit einfacher Nullhypothese)

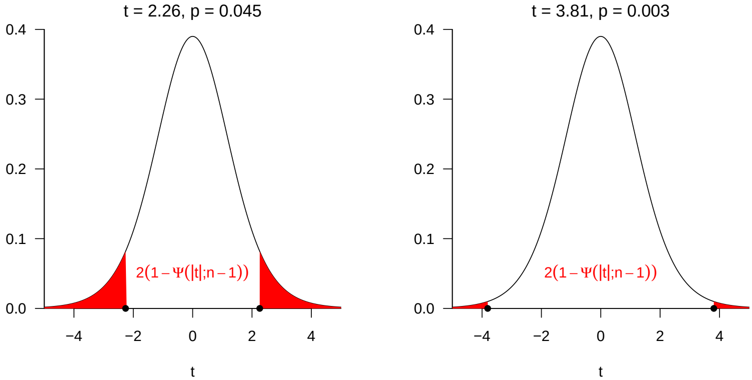
- Es bleibt zu zeigen, dass gilt

$$|t| \geq \Psi^{-1}(1 - \frac{\alpha_0}{2}; n-1) \Leftrightarrow \alpha_0 \geq 2\mathbb{P}(T \geq |t|) \tag{54}$$

- Dies aber folgt aus

$$\begin{aligned}
|t| &\geq \Psi^{-1}\left(1 - \frac{\alpha_0}{2}; n-1\right) \\
\Leftrightarrow \Psi(|t|; n-1) &\geq \Psi\left(\Psi^{-1}\left(1 - \frac{\alpha_0}{2}; n-1\right); n-1\right) \\
\Leftrightarrow \Psi(|t|; n-1) &\geq 1 - \frac{\alpha_0}{2} \\
\Leftrightarrow \mathbb{P}(T \leq |t|) &\geq 1 - \frac{\alpha_0}{2} \\
\Leftrightarrow \frac{\alpha_0}{2} &\geq 1 - \mathbb{P}(T \leq |t|) \\
\Leftrightarrow \frac{\alpha_0}{2} &\geq \mathbb{P}(T \geq |t|) \\
\Leftrightarrow \alpha_0 &\geq 2\mathbb{P}(T \geq |t|).
\end{aligned} \tag{55}$$

Beispiel (Zweiseitiger Einstichproben-T-Test mit einfacher Nullhypothese)



p-Werte

Bemerkungen

- p-Werte spiegeln die Antwort auf die intuitive Frage wie wahrscheinlich es im Frequentistischen Sinne wäre, den beobachteten oder einen extremeren Wert der Teststatistik unter der Annahme eines Nullmodels zu observieren.
- p-Werte sind ist extrem populär, ihre uninformierte Benutzung ist aber auch sehr umstritten.

$\rightarrow$ The American Statistician (2019) Statistical Inference in the 21st Century: A World Beyond p < 0.05

- p-Werte werden, wie Hypothesentestergebnisse generell, leider oft überinterpretiert.
- Es gibt basierend auf dem Gesagten keinen Grund dies anzunehmen, trotzdem vorsorglich:
  - p-Werte quantifizieren nicht die Wahrscheinlichkeit, dass die Nullhypothese wahr ist.
  - Aufgrund von $p < 0.05$ sollte man nicht glauben, dass ein Effekt existiert.
  - Aufgrund von $p > 0.05$ sollte man nicht glauben, dass ein Effekt nicht existiert.
- p-Werte sind eine Möglichkeit ein Signal-zu-Rauschen Verhältnis zu quantifizieren.
- p-Werte sind eine Möglichkeit Unsicherheit zu quantifizieren.

Grundlegende Definitionen

Einstichproben-T-Test

p-Werte

Konfidenzintervalle und Hypothesentests

Anwendungsbeispiel

Selbstkontrollfragen

Theorem (Dualität von Konfidenzintervallen und Hypothesentest)

$\upsilon = \upsilon_1, ..., \upsilon_n \sim p_\theta$ sei eine Stichprobe mit Ergebnisraum $\mathcal{Y}$ und Parameterraum Θ. Weiterhin sei $[G_u(\upsilon), G_o(\upsilon)]$ ein δ-Konfidenzintervall für θ. Dann ist der Hypothesentest

$$\phi_\theta : \mathcal{Y} \to \{0,1\}, y \mapsto \phi(y) := \begin{cases} 0, & [G_u(y), G_o(y)] \ni \theta_0 \\ 1, & [G_u(y), G_o(y)] \not\ni \theta_0 \end{cases} \tag{56}$$

ein Test vom Signifikanzlevel $\alpha_0 = 1 - \delta$ für die Hypothesen

$$\Theta_0 := \{\theta_0\} \text{ und } \Theta_1 := \Theta \setminus \{\theta_0\}. \tag{57}$$

Beweis

Aufgrund der einfachen Nullhypothese und somit $\alpha_0 = \alpha$ folgt

$$\alpha_0 = \alpha = \mathbb{P}_{\theta_0}(\phi(\upsilon) = 1) = \mathbb{P}_{\theta_0}([G_u(y), G_o(y)] \not\ni \theta) = 1 - \mathbb{P}_{\theta_0}([G_u(y), G_o(y)] \ni \theta) = 1 - \delta. \tag{58}$$

Bemerkung

- Mit δ-Konfidenzintervallen kann man also Hypothesentests mit Signifikanzlevel $\alpha_0 = 1 - \delta$ konstruieren.

Beispiel (Konstruktion eines Hypothesentests aus einem Konfidenzintervall)

Wir haben bereits gesehen, dass für eine Stichprobe $\upsilon = \upsilon_1, ..., \upsilon_n \sim N(\mu, \sigma^2)$, $\delta \in]0, 1[$ und

$$t_\delta := \Psi^{-1}\left(\frac{1+\delta}{2}; n-1\right) \tag{59}$$

ein δ-Konfidenzintervall durch

$$\kappa := \left[\bar{\upsilon} - \frac{S}{\sqrt{n}} t_\delta, \bar{\upsilon} + \frac{S}{\sqrt{n}} t_\delta\right]. \tag{60}$$

definiert ist. Mit der Dualität von Konfidenzintervallen und Hypothesentests können wir also folgenden Test für die Hypothesen $\Theta_0 = \{\mu_0\}$ und $\Theta_1 = \mathbb{R} \setminus \mu_0$ definieren:

$$\phi : \mathcal{Y} \to \{0, 1\}, y \mapsto \phi(y) := \begin{cases} 0, & \left[\bar{\upsilon} - \frac{S}{\sqrt{n}} t_\delta, \bar{\upsilon} + \frac{S}{\sqrt{n}} t_\delta\right] \ni \mu_0 \\ \\ 1, & \left[\bar{\upsilon} - \frac{S}{\sqrt{n}} t_\delta, \bar{\upsilon} + \frac{S}{\sqrt{n}} t_\delta\right] \not\ni \mu_0 \end{cases} \tag{61}$$

Dann gilt

$$\begin{aligned} \mathbb{P}_{\mu_0}(\phi(\upsilon) = 1) &= 1 - \mathbb{P}_{\mu_0}(\phi(\upsilon) = 0) \\ &= 1 - \mathbb{P}_{\mu_0}\left(\left[\bar{\upsilon} - \frac{S}{\sqrt{n}} t_\delta, \bar{\upsilon} + \frac{S}{\sqrt{n}} t_\delta\right] \ni \mu_0\right) \\ &= 1 - \delta. \end{aligned} \tag{62}$$

und wir haben gezeigt, dass ϕ ein Test vom Signifikanzlevel $\alpha_0 = 1 - \delta$ ist.

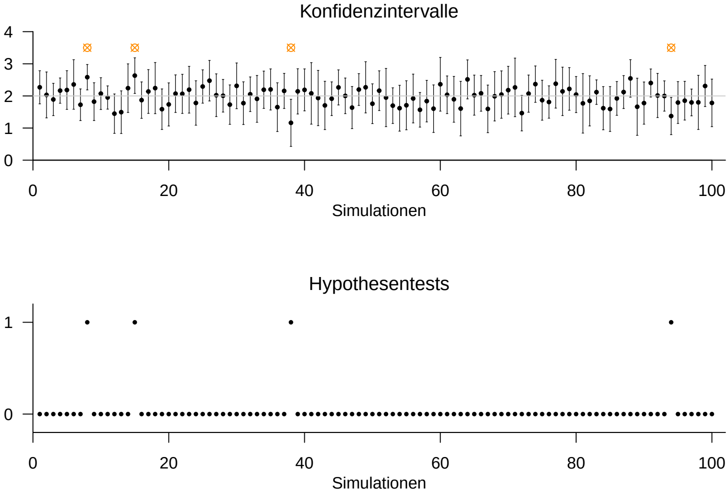
Simulation der Dualität von Konfidenzintervallen und Hypothesentests

```
# Modellformulierung
n       = 12                                            # Stichprobengröße
mu      = 2                                             # wahrer, aber unbekannter, Erwartungswertparameter
sigsqr  = 1                                             # wahrer, aber unbekannter, Varianzparameter

# Konfidenzintervallparameter und Testparameter
delta   = 0.95                                          # Konfidenzbedingung
t_delta = qt((1+delta)/2, n-1)                          # \Psi^{-1}((\delta + 1)/2, n-1)
mu_0    = mu                                            # Nullhypothesenparameter

# Simulationen
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
  # Stichprobenrealisation und Konfidezintervallevaluation
  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 KI Grenze
  kappa[i,2] = y_bar[i] + (s[i]/sqrt(n))*t_delta        # obere KI Grenze

  # Überdeckungs- und Testevaluation
  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

# Ausgabe
cat(   "Geschätztes Konfidenzniveau =", mean(kfn),
     "\nGeschätzter Testumfang      =", mean(phi))
```

```
> Geschätztes Konfidenzniveau = 0.96
> Geschätzter Testumfang      = 0.04
```

Konfidenzintervalle und Hypothesentests

Simulation der Dualität von Konfidenzintervallen und Hypothesentests



Grundlegende Definitionen

Einstichproben-T-Test

p-Werte

Konfidenzintervalle und Hypothesentests

Anwendungsbeispiel

Selbstkontrollfragen

Anwendungsbeispiel

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Pre-BDI


BDI-II Fragebogen

Name | Alter | Geschlecht m / w | Datum

Anleitung: Dieser Fragebogen enthält 21 Gruppen von Aussagen. Bitte lesen Sie jede dieser Gruppen von Aussagen sorgfältig durch und suchen Sie sich dann in jeder Gruppe **eine Aussage** heraus, die am besten beschreibt, wie Sie sich **in den letzten zwei Wochen, einschließlich heute, gefühlt haben**. Kreuzen Sie die Zahl neben der Aussage an, die Sie sich herausgesucht haben (0, 1, 2 oder 3). Falls in einer Gruppe mehrere Aussagen gleichermaßen auf Sie zutreffen, kreuzen Sie die Aussage mit der höheren Zahl an. Achten Sie bitte darauf, dass Sie in jeder Gruppe nicht mehr als eine Aussage ankreuzen, das gilt auch für Gruppe 16 (Veränderungen der Schlafgewohnheiten) oder Gruppe 18 (Veränderungen des Appetits).

1.) Traurigkeit
0 Ich bin nicht traurig.
1 Ich bin oft traurig.
2 Ich bin ständig traurig.
3 Ich bin so traurig oder unglücklich, dass ich es nicht aushalte.

2.) Pessimismus
0 Ich sehe nicht mutlos in die Zukunft.
1 Ich sehe mutloser in die Zukunft als sonst.
2 Ich bin mutlos und erwarte nicht, dass meine Situation besser wird.
3 Ich glaube, dass meine Zukunft hoffnungslos ist und nur noch schlechter wird.

3.) Versagensgefühle
0 Ich fühle mich nicht als Versager.
1 Ich habe häufiger Versagensgefühle.
2 Wenn ich zurückblicke, sehe ich eine Menge Fehlschläge.
3 Ich habe das Gefühl, als Mensch ein völliger Versager zu sein.

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.

5.) Schuldgefühle
0 Ich habe keine besonderen Schuldgefühle.
1 Ich habe oft Schuldgefühle wegen Dingen, die ich getan habe oder hätte tun sollen.
2 Ich habe die meiste Zeit Schuldgefühle.
3 Ich habe ständig Schuldgefühle.

6.) Bestrafungsgefühle
0 Ich habe nicht das Gefühl, für etwas bestraft zu sein.
1 Ich habe das Gefühl, vielleicht bestraft zu werden.
2 Ich erwarte, bestraft zu werden.
3 Ich habe das Gefühl, bestraft zu sein.

7.) Selbstablehnung
0 Ich halte von mir genauso viel wie immer.
1 Ich habe Vertrauen in mich verloren.
2 Ich bin von mir enttäuscht.
3 Ich lehne mich völlig ab.

8.) Selbstvorwürfe
0 Ich kritisiere oder tadle mich nicht mehr als sonst.
1 Ich bin mir gegenüber kritischer als sonst.
2 Ich kritisiere mich für all meine Mängel.
3 Ich gebe mir die Schuld für alles Schlimme, was passiert.

9.) Selbstmordgedanken
0 Ich denke nicht daran, mir etwas anzutun.
1 Ich denke manchmal an Selbstmord, aber ich würde es nicht tun.
2 Ich möchte mich am liebsten umbringen.
3 Ich würde mich umbringen, wenn ich die Gelegenheit dazu hätte.

10.) Weinen
0 Ich weine nicht öfter als früher.
1 Ich weine jetzt mehr als früher.
2 Ich weine beim geringsten Anlass.
3 Ich möchte gern weinen, aber ich kann nicht.

PEARSON PsychCorp

Summe Seite 1:

Post-BDI

⇒ Pre-Post BDI Score Reduktion

Anwendungsbeispiel

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Wir legen für die BDI Score Reduktion υ_i der $i$ten von n Patient:innen das Modell

$$\upsilon_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, ..., n \tag{63}$$

zugrunde.

Wir erklären die BDI Reduktion υ_i der $i$ten Patient:in also mithilfe einer über die Gruppe von Patient:innen identischen BDI Score Reduktion μ als Effekt der Therapieintervention und einer Patient:innen-spezifischen normalverteilten BDI Score Reduktionsabweichung ε_i, die sich aus sehr vielen additiven Prozessen zusammensetzt, für die wir also eine Normalverteilungsannahme treffen, und deren Varianz wir mit σ^2 parameterisieren.

Vor dem Hintergrund dieses Modells evaluieren wir die drei Standardprobleme der Frequentistischen Inferenz für den Effekt der Therapieintervention μ:

(1) Was ist unserer Schätzung für den Effekt μ der Therapie auf die BDI Score Reduktion?
(2) Welches 95%-Konfidenzintervall ist mit dieser Schätzung von μ assoziiert?
(3) Entscheiden wir uns sinnvoller Weise für die Nullhypothese $\mu = 0$?

Anwendungsbeispiel

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

```
fname = file.path(getwd(), "12_Hypothesentests.csv")
D     = read.table(fname, sep = ",", header = T)
```

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

Anwendungsbeispiel

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

```
# Einlesen und Auswahl der Daten
fname     = file.path(getwd(), "12_Hypothesentests.csv")  # Dateiname
D         = read.table(fname, sep = ",", header = T)      # Dataframe
y         = D$BDI.Reduktion                               # Datenrealisation
n         = length(y)                                     # Anzahl Datenpunkte

# Parameterschätzung
y_bar      = mean(y)                                      # Stichprobenmittel
mu_hat     = y_bar                                        # Unverzerrte Maximum-Likelihood Schätzung

# Konfidenzintervallevaluation
delta    = 0.95                                           # Konfidenzlevel
t_delta  = qt((1+delta)/2,n-1)                            # \Psi^-1((\delta + 1)/2, n-1)
G_u      = y_bar - (sd(y)/sqrt(n))*t_delta                # untere KI Grenze
G_o      = y_bar + (sd(y)/sqrt(n))*t_delta                # obere  KI Grenze

# Testevaluation
mu_0      = 0                                             # H_0 Hypothesenparameter, hier \mu = \mu_0
alpha_0   = 0.05                                          # Signifikanzlevel
k_alpha_0 = qt(1-alpha_0/2,n-1)                           # kritischer Wert
Tee       = sqrt(n)*((y_bar - mu_0)/sd(y))                # T-Teststatistik
if(abs(Tee) > k_alpha_0){phi = 1} else {phi = 0}          # Test 1_{|t| >= k_alpha_0}

# p-Wert Evaluation
p         = 2*(1 - pt(Tee,n-1))                           # p-Wert
```

Anwendungsbeispiel

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

```
# Ausgabe
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)
```

```
> Parameterschätzwert    = 3.17
> 95%-Konfidenzintervall = 0.807 5.53
> Signifikanzlevel       = 0.05
> Kritischer Wert        = 2.2
> Teststatistik          = 2.95
> Testwert               = 1
> p-Wert                 = 0.0131
```

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Frequentistische Inferenz mit R's `t.test()` Funktion

```
t.test(y)          # Anwendung der in Einheiten (1) bis (12) entwickelten Theorie
```

```
>
>   One Sample t-test
>
> data:  y
> t = 3, df = 11, p-value = 0.01
> alternative hypothesis: true mean is not equal to 0
> 95 percent confidence interval:
>  0.807 5.526
> sample estimates:
> mean of x
>      3.17
```

Anwendungsbeispiel

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

(1) Was ist unserer Schätzung für den wahren, aber unbekannten, Effekt μ der Therapie?

- Unsere Schätzung für den Therapieeffekt ist eine BDI Reduktion von $\hat{\mu} = 3.17$.

(2) Welches 95%-Konfidenzintervall ist mit dieser Schätzung von μ assoziiert?

- Das 95%-Konfidenzintervall für den Therapieeffekt ist eine BDI Reduktion $[0.81, 5.53]$.

(3) Entscheiden wir uns sinnvoller Weise für die Nullhypothese $\mu = 0$?

- Bei einem Signifikanzlevel von $\alpha_0 = 0.05$ lehnen wir die Nullhypothese $\mu = 0$ ab ($p = 0.01$).

Aus datenwisschenschaftlicher Sicht sind Ergebnis und Unsicherheit dieses Szenarios im Frequentistischen Sinne nun quantifiziert. Die Interpretation des Ergebnisses inklusive des Mehrwerts der Therapie bei einer geschätzten BDI Score Reduktion von ≈ 3 obliegt der Klinischen Psychologie.

Grundlegende Definitionen

Einstichproben-T-Test

p-Werte

Konfidenzintervalle und Hypothesentests

Anwendungsbeispiel

Selbstkontrollfragen

Selbstkontrollfragen

1. Erläutern Sie die grundlegende Logik statistischer Hypothesentests.
2. Geben Sie die Definition statistischer Hypothesen und eines Testszenarios wieder.
3. Definieren Sie die Begriffe der einfachen und zusammengesetzten Hypothesen.
4. Definieren Sie die Begriffe der einseitigen und zweiseitigen Hypothesen.
5. Definieren Sie den Begriff des Tests.
6. Definieren Sie den Begriff des Standardtests.
7. Definieren Sie den Begriff des kritischen Bereichs eines Tests.
8. Definieren Sie den Begriff des Ablehungsbereichs eines Tests.
9. Definieren Sie den Begriff des kritischen Wert-basierten Tests.
10. Definieren Sie richtige Testentscheidungen, Typ I Fehler und Typ II Fehler.
11. Definieren Sie die Testgütefunktion.
12. Erläutern Sie die Bedeutung der Testgütefunktion im Rahmen der Konstruktion statistischer Tests.
13. Definieren Sie die Begriffe des Signifikanniveaus und des Level-α_0-Tests.
14. Definieren Sie den Begriff des Testumfangs.
15. Erläutern Sie zentrale Schritte zur Konstruktion eines Hypothesentests.

Selbstkontrollfragen

16. Formulieren Sie das statistische Modell eines Einstichproben-T-Tests.
17. Formulieren Sie die einfache Nullhypothese und zusammengesetzte Alternativhypothese dieses Tests.
18. Definieren Sie den zweiseitigen Einstichproben-T-Test (ZETT).
19. Skizzieren Sie qualitativ die Testgütefunktionen eines ZETTs für verschiedene kritische Werte.
20. Wie muss der kritische Wert eines ZETTs definiert sein, damit der Test ein Level-α_0-Test ist?
21. Skizzieren Sie qualitativ die Bestimmung des kritischen Wertes k_{α_0} bei einem zws Einstichproben-T-Test.
22. Erläutern Sie das praktische Vorgehen zur Durchführung eines ZETTs.
23. Von welchen Werten hängt die Powerfunktion eines ZETTs ab?
24. Skizzieren Sie qualitativ die Powerfunktion des ZETTs bei fester Stichprobengröße.
25. Skizzieren Sie qualitativ die Powerfunktion des ZETTs bei festem Erwartungswertparameter.
26. Erläutern Sie das favorisierte praktische Vorgehen zur Durchführung einer Poweranalyse.
27. Erläutern Sie die Motivation zur Auswertung von p-Werten.
28. Definieren Sie den Begriff des p-Werts.
29. Geben Sie das Theorem zur Dualität von Konfidenzintervallen und Hypothesentests wieder.
30. Erläutern Sie die Dualität von Konfidenzintervallen und Hypothesentests.

References

Horvath, Lilla, Stanley Colcombe, Michael Milham, Shruti Ray, Philipp Schwartenbeck, and Dirk Ostwald. 2021. "Human Belief State-Based Exploration and Exploitation in an Information-Selective Symmetric Reversal Bandit Task." *Computational Brain & Behavior*, August. https://doi.org/10.1007/s42113-021-00112-3.

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.